NBER WORKING PAPER SERIES

THE IMPACT OF EARLY CHILDHOOD PARENTING INTERVENTIONS ON CHILD LEARNING: A SYSTEMATIC REVIEW AND META-ANALYSIS

Kimberly Dadisman
Andre Nickow
Philip Oreopoulos

Working Paper 32959
http://www.nber.org/papers/w32959

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2024

We are greatful to Lisa Turley Smith and Victoria Moura for research assistance, as well as funding from the Social Sciences and Humanities Research Council of Canada and the Jameel Poverty Action Lab North America. We also thank seminar participants at University College Dublin, the University of Toronto, and the London School of Economics. Any errors or omissions are our own responsibility. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Kimberly Dadisman, Andre Nickow, and Philip Oreopoulos. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Impact of Early Childhood Parenting Interventions on Child Learning: A Systematic Review and Meta-Analysis
Kimberly Dadisman, Andre Nickow, and Philip Oreopoulos
NBER Working Paper No. 32959
September 2024
JEL No. I2, J13

ABSTRACT

Parenting is widely considered to be among the most important influences on early childhood (EC) development. But to what extent and under what circumstances can EC parenting programs improve child learning outcomes? While substantial progress has been made toward addressing these questions in recent years, there have been few attempts to systematically synthesize the evidence thus far with a view toward scaling and policy implications. This paper works toward filling this gap through a systematic review including both a quantitative meta-analysis and a detailed narrative analysis of randomized evaluations that test the impacts of EC parenting programs on learning outcomes. We find that these programs generate substantial effects across a wide range of contexts, and that the largest impacts are associated with programs that are conducted in low- or middle-income countries and that use curricula focusing on cognitive stimulation. Group parenting programs tend to yield effect sizes that are, on average, comparable to home visiting programs, typically at substantially lower costs. Qualitative analysis of evaluations of scaled interventions reveals that administrative implementation barriers rather than program ineffectiveness likely represent the primary impediment to stronger impact. We conclude by reflecting on implications for theory, policy, and priorities for future research.

Kimberly Dadisman
Massachusetts Institute of Technology
kdadisman@povertyactionlab.org

Andre Nickow
Global Poverty Research Lab
Northwestern University
601 University Place
Chicago, IL 60208
a-nickow@kellogg.northwestern.edu

Philip Oreopoulos
Department of Economics
University of Toronto
150 St. George Street Toronto,
ON M5S 3G7
and NBER
philip.oreopoulos@utoronto.ca

Introduction

Researchers from the social, behavioral, and medical sciences agree that early childhood (EC) parenting ranks among the most powerful and lasting influences on children's skill development. Hundreds of studies over the past several decades have demonstrated the potential utility of targeted social programs in encouraging supportive parenting practices and thereby improving a range of child outcomes. However, the circumstances under which these programs are effective and particularly the question of whether and how such programs can be effectively disseminated and scaled in particular settings remain an open question. The drivers and effects of parenting investments emerge from complex interactions between institutions, economies, and cultural codes, within arenas from families and informal care networks to schools and health facilities. In this context, how and to what extent can early childhood parenting programs effectively contribute to skills development at scale?

To address this question, we conducted a systematic review of randomized controlled trial (RCT) impact evaluations that evaluate early childhood parenting programs across the world. We define EC parenting programs as interventions that aim to impact the parenting behaviors of individuals caring for children ages 0-5. We use the word *parent* to refer to a primary caregiver, without biological or legal assumptions. Parenting programs range in content from learning activities to abuse prevention to encouraging healthy emotional development. The programs are typically delivered via home visiting or clinic-based sessions.

Home visiting programs with parenting components have been prominent within the United States and other high-income countries (HICs) since early evidence of the Nurse-Family Partnership's (NFP) success circulated in the 1980s, and in lower- and middle-income countries (LMICs) since the success of the Jamaica Longitudinal Study's home visiting program rose to prominence during the 1990s. Experimental research on parenting programs has especially flowered in many countries across the world over the past two decades. Recent reviews demonstrate consistently substantial impacts across a wide range of outcome categories and contexts (Aboud & Yousafzai, 2015; Attanasio et al. 2022; García & Heckman 2023; Jeong et al. 2021A).

While findings have been promising, the vast majority of RCT evaluations of parenting programs conducted to date and covered in existing reviews have tested interventions implemented at small scale that avoid potential challenges associated with scaleup in real-life policy settings. For instance, costs, administrative logistics, labor supply, and access to beneficiaries can shape program impacts in ways that would not likely be captured in smaller-scale studies (Araujo et al., 2021B; Britto et al., 2017; List et al., 2021). In the present article, we leverage the experimental research to generate insights for social science and policy debates by complementing our quantitative meta-analysis with an analytical review centered on scaling and implementation.

We find that, while parenting programs have been effective at improving learning outcomes in a wide range of contexts, results are more varied in the larger-scale settings that best approximate policy conditions. Parenting programs exhibit the potential for transformative impact, but whether this potential is realized depends on a range of characteristics of the program and the context in which it is implemented. Impacts tend to be stronger in LMICs than HICs. Group parenting programs can be as effective as home visiting programs despite typically lower costs, but patterns of population density, social infrastructure, and transportation and time costs must be factored in when selecting a delivery model for a specific context. Implementation issues seem to have been the primary barrier to effectiveness for several prominent programs that did not show statistically significant impacts. Finding a supportive niche within government and civic institutions allowing for sufficient administrative capacity and related resources appears key to program success at large scales.

In the following section, we outline some of the key motivations for and existing evidence on parenting programs, and then position our review in the context of several related efforts. The subsequent section explains our systematic review methodology. We then go on to describe our study sample and highlight broad patterns in program models and results following a narrative approach, before moving on to our meta-analytic results followed by a more in-depth thematic analysis of delivery mode, interventionist characteristics, administrative capacity and integration, and long-term results. We additionally compare our findings to those of other recent meta-analyses.

Early Childhood Parenting Programs and Learning Outcomes

An enormous volume of scholarly research covering a variety of methodologies and disciplines has established the potentially transformative impact of EC parenting programs. These programs are primarily motivated by the understanding that the early years of a child's life are crucial for their cognitive, social, and emotional development (Heckman, 2006; Suskind, 2022) and are theoretically grounded in developmental psychology and neuroscience, which emphasize the impact of early experiences on brain development and lifelong learning (Shonkoff & Phillips, 2000; Bronfenbrenner, 1977). The goal of EC parenting programs is to equip caregivers with the knowledge and skills necessary to foster a nurturing and stimulating environment for their children based on the understanding that caregivers play a pivotal role in their children's early development (Attanasio, Cattan, & Meghir, 2022; Suskind, 2015). Programs that focus on parenting have shown results that are comparable to substantially more expensive, multi-component EC programs such that parenting programs may represent “alternatives to more expensive center-based programs at a fraction of their cost (5–10%) and, in many cases, with relatively low-skill requirements for the home visitors” (García & Heckman, 2023, p. 354).

While many reviews have covered research on parenting programs, most cover parenting programs as part of the broader early childhood development policy landscape and/or focus on narrower subsets of the parenting research (e.g., Aboud & Yousafzai 2015; Attanasio, Cattan, & Meghir, 2022; García & Heckman, 2023). Several recent meta-analyses define their scope to focus on parenting programs and—like this one—limit their study samples to experiments (Jeong et al. 2021A; Zhang et al. 2021). Their findings further reinforce the effectiveness of these programs and the pooled impact estimates presented provide benchmarks that are useful for gaining a broad sense of the range of feasible program impacts.

However, these meta-analyses rely almost exclusively on pooled impacts from diverse studies with differing contexts, limiting their ability to inform education policy. The estimates reflect “a hodgepodge of programs” that are “assessed using diverse measures on diverse populations”; drawing inference from pooled impacts in this context leads to “comparisons of incomparables” (García & Heckman, 2023, p. 350-351). Additional analytical steps are necessary for honing empirically-grounded hypotheses about the situations under which different intervention characteristics are likely to be most effective. For instance, comparing pooled results of programs delivered through group parenting sessions versus one-on-one home visiting is a worthwhile task, but it is likely that each has benefits and drawbacks tied to context, and that a clearer understanding of these contextual drivers and underlying mechanisms will ultimately be more useful for policymaking and social science scholarship than simply comparing undifferentiated averages.

García and Heckman take an alternative approach at the opposite end of the spectrum from meta-analyses focused on pooled impacts. They present close analyses of two “iconic” multifaceted programs that include center-based activities—the Perry Preschool and Abecedarian Projects—as well as several parent-focused home visiting programs. The goal is to identify the mechanisms driving impact so that they can be isolated and adapted to different contexts, with a focus on a wide range of outcomes measured longitudinally over the life course. The review presents a powerful critique of conventional meta-analysis and produces a wealth of compelling and policy-relevant insights, with a primary conclusion being the centrality of parenting and home life in driving EC program impact even for programs with a center-based component.

While these insights centrally inform our approach, the specific approach to synthesis used in García and Heckman's (2023) review necessitates reliance on a relatively narrow range of studies to avoid the "hodgepodge" pitfall. Additionally, the iconic programs that they discuss are highly unusual in the richness and duration of the studies that have evaluated them. Advancing the literature further would seem to require more inclusivity to avoid ignoring large swathes of potentially relevant literature, not to mention starkly limiting the range of feasible research questions and leaving open the extent to which the processes that occurred in iconic case studies replicate in other contexts—particularly at large scales. For instance, the iconic and transformative Jamaica Reach program showed strong success in a range of iterations across diverse countries, but virtually none of the adaptations in other countries showed effect sizes (ES's) approaching the 0.88 standard deviation (SD) impact of the original study.

We thus feel that it remains important to consider impacts and dynamics associated with a wider range of studies. Pooled impacts—even of diverse programs across diverse contexts—can be valuable as long as these estimates are unpacked and understood as comparative benchmarks rather than predictions for particular programs. Furthermore, relying exclusively on a narrow range of small-scale studies could mean missing important administrative and logistical considerations associated with scale. Our review integrates a conventional meta-analysis approach covering a wide range of studies including at multiple scales—albeit, limited and made manageable by our decision to focus exclusively on directly-measured standardized learning outcomes—while also considering context and scale through attention to mechanisms and patterns of variation in impact. We thus see our review as an important complement both to those that focus overwhelmingly on conventional meta-analysis (e.g., Jeong et al. 2021A; Zhang et al. 2021) on one hand and mechanism-oriented approaches (e.g., García & Heckman, 2023) on the other.

Another challenge to more fine-grained, policy-relevant analyses of the early childhood parenting program literature is the wide range of outcomes that have been reported in the studies and included within the aforementioned meta-analyses among others. This breadth is understandable—a robust theoretical and empirical literature (e.g., Cunha, Heckman, & Schennach, 2010) demonstrates the synergistic effects by which some types of skills in early childhood may help to support the development of others later in life. However, when combined with the size of the literature, taking a wide range of outcomes into account simultaneously tends to make context- and mechanism-oriented analysis unwieldy. While impacts on diverse measures like parenting practices, child learning, and child socio-emotional outcomes interrelate with one another, each of these merits close attention on its own terms.

A new literature is beginning to emerge looking at impacts on more specific sets of outcomes, for example child behavior and conduct problems (Furlong et al., 2012; Leijten et al., 2019) and abusive parenting practices (Levey et al. 2017). Surprisingly, no review to date has focused specifically on learning outcomes, one of the key intervening variables by which early childhood development programs are expected to yield lasting results throughout the life course. The present review contributes to the emerging sector-specific evaluation synthesis literature on EC parenting programs by filling this gap. Moreover, we focus only on studies that measure learning outcomes—including cognitive, academic, and linguistic skills—from measures that involve directly testing children. Since parenting programs are directed at parents, they may alter parental interpretations of their children's skills and thereby bias results of parent-reported measures. And, in contrast to most other reviews on this topic, we only use independent standardized measures since researcher-designed measures can lead to bias.

Another limitation is that the literature to date has mostly been split between HICs and LMICs. This is understandable given vast differences in the education and social policy settings. However, rising within-country diversity in recent years, along with the growth of middle-income countries like China and Mexico that blur the lines between developed and developing countries means that a more comprehensive framework for understanding how the dynamics of parenting program impact will become increasingly necessary in the coming years. Only one other project to date (Jeong et al. 2021A: 2021B) has presented a meta-analysis of RCT evaluations of parenting interventions in both HICs and LMICs, and as mentioned this study focuses almost entirely on meta-analytic averages rather than policy-oriented analysis. An extension explores the extent to which "implementation characteristics" moderate outcomes, but of the nine moderators they examine, only dosage overlaps with ours and the remaining categories are mostly abstract and difficult to interpret from a policy perspective, e.g., whether the program has a clearly-specified theory of change, and the number of behavior change techniques used (Ahun et al. 2024, p. 102). This contrasts with our focus on concrete typologies of delivery mode and curriculum type that emerged inductively from the literature. Furthermore, this endeavor and the others cited above include only programs targeting babies aged 0-2 or 0-3, whereas the present paper covers programs targeting the 0-5 age range since the latter represents a comprehensive window into the pre-primary learning environment.

In summary, this review relies on a context- and mechanism-oriented qualitative literature review to complement the quantitative meta-analysis with the intention of facilitating inferences for theory and policy. To sharpen the analysis, we narrow our focus to directly-measured child learning outcomes. We include LMICs and HICs within the same framework to facilitate comparison within and across these two categories. Finally, we pay special attention to issues of scaleup as perhaps the central challenge currently faced by policymakers and practitioners.

Meta-Analysis Methodology

To assemble our meta-analysis sample of parenting program RCTs, we searched a range of databases using the following sets of search terms for the years 1980-2023:

- "early childhood" AND "random*" AND "parent*"
- "early childhood" AND "random*" AND "caregiver"
- "home visit*" AND "random*"

Databases searched included Academic Search Complete, APA PsycArticles, APA PsycInfo, EconLil, Education Abstracts (H.W. Wilson), ERIC, JSTOR, SCOPUS, Web of Science, and Proquest Dissertations and Theses. We also searched for working papers at NBER and SSRN. To fill in gaps, we searched for relevant reviews at the websites of Campbell Collaboration and What Works Clearinghouse, searched gray literature at the websites of professional evaluation companies, and employed backward and forward bibliographic searching for each article in our sample and other key publications. Once potentially eligible papers had been assembled, we screened them using the following inclusion criteria:

- Published 1980-2023
- Sample size of 50 or larger
- Baseline age of 0-5 years old for children in the sample

- Outcomes from independent learning tests, i.e., not tests designed by the researcher for the purpose of the evaluation since they can inflate results (Slavin & Madden, 2011)
- Outcomes from tests administered directly to children
- Endlines administered a year or less after treatment (longer-term follow-ups are included only in the qualitative analysis of the literature
- Intent-to-treat estimate (other estimate types are noted in the narrative discussion where relevant)
- At least one experimental arm where a parenting program is the only substantial intervention (so that the independent effects of the parenting program can be discerned), and at least one control group receiving no parenting programs (to avoid studies that simply compare different types of parenting programs against one another)
- Provides sufficient data to calculate effect sizes (ES) using the procedures explained below

Once potentially eligible papers had been identified, we consolidated them to the level of the "study", the key unit of analysis. A study refers to the enactment of a research design with a specific sample, and in many cases multiple papers report on the same study. Studies were coded on key variables of interest, and key statistics were gathered. The first author conducted screening and coding, with spot checks from research staff. All ambiguities were discussed with the full research team. To ensure comparable ES's across studies, we calculated them ourselves rather than relying on author reports (Pigott & Polanin, 2020). Specifically, we calculated Hedges' *g* using adjusted means where reported and unadjusted means otherwise, along with unadjusted standard deviations (Dietrichson et al., 2017, p. 255).

We then ran a series of analyses relying on a robust variance estimation (RVE) metaregression framework. RVE adjusts for the number of reported estimates by study to avoid overweighting studies that report a higher number of estimates. To calculate pooled ES's, we used Stata's "robumeta" package (Hedberg et al. 2011). We present precision intervals along with the tau-squared statistic to examine heterogeneity. Precision intervals are calculated as $B \pm t_{df}\sqrt{V_m + T^2}$, where B is the pooled ES estimate, t_{df} is the t-score given degrees of freedom, and T^2 is tau-squared. The prediction interval is " an index of dispersion (based on the standard deviation) that tells us how widely the effects vary across populations" in contrast to the confidence interval, which is an "index of precision (based on the standard error) that tells us how precisely we have estimated the mean effect size" (Borenstein et al., 2017, p. 5). We also report tau-squared and Q as additional indicators of heterogeneity (p. 6)..

While our main estimates are pooled ES's for subsets of studies as defined by program and study characteristics, we additionally run a series of multivariate metaregressions aimed at testing robustness and, where possible, beginning to untangle the overlapping effects of various characteristics. Finally, we run an adjusted version of Egger's regressions to test the extent of selection bias. Traditional Egger's regression add standard error as a covariate and are thus distorted by "inflated Type 1 error due to the correlation between the effect size estimate and its standard error (Pustejovsky & Rodgers, 2019, p. 57)," we use a transformed standard error (p. 59) which avoids this shortcoming. Specifically, the standard error is replaced with W_i, defined as:

$$W_i = (\hat{\Delta})_i / \sigma_i^2$$

where $(\hat{\Delta})_i$ is the difference between treatment and control group means in unstandardized units and σ_i^2 is the variance of the outcome (i.e., the sum of treatment and control group variances). We run the regressions both with W_i as the only independent variable (Table 5A) and with an added dummy variable indicating the smallest sample size, between 50 and 200 (Table 5B).

Conceptual Framework of EC Parenting Program Efficacy

This section outlines a conceptual framework to highlight the key dimensions of interest explored in the analysis below. EC parenting programs are defined here as interventions intended to educate and/or encourage primary caregivers to engage in practices that are expected to optimize child development. Our framework consists of six primary dimensions that may be expected to shape program impact, and that are generally plausible to code across a wide range of articles: program content, delivery mode, child age at baseline, session frequency, program duration, and HIC vs. LMIC status.

Of these, the latter four are generally straightforward. Age is likely to matter since different parenting behaviors may yield different influences even across small differences in age. Higher levels of dosage—operationalized here as session frequency and program duration—would seem more likely to influence parenting behavior. HICs may be expected to have a wider range of resources than LMICs which could mean that parenting programs in the latter are more impactful, although on the other hand lower-resourced delivery systems in LMICs may constrain implementation relative to HICs. Our typologies of program content and delivery mode emerged from this review, and require additional explanation. Here we discuss each further in turn.

While EC parenting programs may engage a wide array of potential content, the programs evaluated in our study sample tend to fall into one or more of the following categories, which are not mutually exclusive. The first two categories encourage parents to directly teach their children skills that will support their learning. Education and neuroscience research have demonstrated the importance of parents and caregivers' capacities to drive foundational brain development for young children, particularly those born into poverty, and has emphasized the potential impact of caregivers' engagement in child learning on reducing the achievement gap and preventing early cognitive disparities (Suskind, 2015; 2022). We categorize programs intended to increase parent engagement around learning cognitive and related skills as *cognitive stimulation*. Programs emphasizing *language* are closely related, but we categorize these separately since language development may in many ways be distinct from other cognitive skills, and programs exclusively focused on language development have emerged as a relatively distinct area of study and practice (Suskind, 2015).

Third, many parenting programs focus on encouraging parenting behaviors expected to improve attachment security in the attachment theory literature, which posits that a strong emotional and physical attachment to at least one primary caregiver is critical to personal development. Attachment theory emphasizes the role of consistent, responsive parenting in shaping a child's ability to form secure and healthy relationships (Bowlby, 1988). While the primary goals of attachment-oriented programs may typically revolve around non-cognitive skills rather than learning outcomes, both theory and empirical evidence highlight potential "cross-productivity" dynamics as non-cognitive skills associated with secure attachment may provide a foundation for academic skills improvement (Cunha et al. 2010).

The fourth content category we identified—which we refer to as *lifestyle* programs—encompasses interventions that encourage parents to create a healthy environment for children, e.g., by reducing physical, emotional, or substance abuse, educating around nutrition, and connecting parents to government or community resources. These programs may draw inspiration from a range of theoretical traditions. For instance, social learning theory suggests that children learn behaviors, values, and attitudes through observation and imitation of their parents. Therefore, some early childhood parenting programs focus on enhancing caregiver practices to model positive behaviors and attitudes (Bandura, 1977).

Interventions in this category may engage broader systems issues. Bronfenbrenner's ecological systems theory suggests that an individual's development is influenced by a series of interconnected environmental systems, ranging from the immediate surroundings (e.g., family) to broad societal structures (e.g., culture) (Bronfenbrenner, 1977).

Once a program's content has been established, a mode of delivery must be selected to communicate this content to parents. We identified four program delivery categories inductively from our study sample. As is the case for the program content categories, delivery categories are not mutually exclusive since programs may involve multiple components. While program costs may vary as a result of many contextual factors, we expect costs to differ systematically by delivery mode. The delivery modes are listed here in order of likely costliness from most to least expensive.

First, in *home visiting* programs, interventionists travel in person to parents' homes in order to deliver the content one-on-one. This delivery mode has the benefit of minimizing costs for parents and allowing interventionists to observe and work with parents in the home environment, but it may be relatively expensive given large staffing requirements along with time and travel costs. Second, the delivery mode that we refer to as *individual-clinic* is similar in that interventionists engage with parents one-on-one in person, but in this latter category parents travel to meet the interventionist rather than the other way around.

Third, *group* parenting programs impart content to multiple parents at a time, typically in a clinic or other central location that parents travel to. These programs still require substantial time from interventionists, but costs are likely lower than the first two delivery modes since multiple parents receive the program at the same time. The group structure may reduce impacts if the quality of interventionists' attention to parents reduces relative to one-on-one engagement, or it may augment impact if the group structure is more conducive to encouraging behavior change. Fourth, some programs are delivered via information and communication technology (ICT). Like the above two categories, parents typically receive the content individually, but ICT delivery lacks in-person engagement with parents (although ICT programs can hypothetically include in-person activities separate from the ICT engagement). While ICT programs require a minimum level of technological infrastructure, this delivery mode may be inexpensive at scale since the programs are typically automated and do not necessarily require employing interventionists once the program has been created.

Program delivery can be a highly complex operation, and involves numerous dimensions beyond the broad classifications of delivery mode described here, for instance interventionist training and supervision. However, these nuances are difficult to code across our study sample. In light of our focus on program scalability, we devote substantial attention to other delivery-related characteristics in the broader analysis, but do not include them in our quantitative meta-analysis.

Charting the Experimental Research Landscape for EC Parenting Programs

This section provides an overview of the experimental EC parenting literature. We integrate a narrative discussion of the development of the literature since 1980 with a description of our study sample, highlighting examples of prominent programs. The sample consists of 90 studies that took place in 29 countries. Table 1 shows the number of studies and average sample sizes across key program characteristics for the full study sample, and also broken down by HIC versus LMIC status. The studies were split relatively evenly across HICs (47) and LMICs (43), although the latter had much larger average sample sizes. The United States accounts for a large share of the HIC research. Among LMICs, there are emerging concentrations of studies in Bangladesh, China, and India. A few southern and eastern African countries are represented, but most of the continent remains open. Over the past decade or so, Latin America has been home to some of the largest-scale and most policy-relevant evaluations. While there were several studies in HICs (two in the US and one in Bermuda) during the 1980s, research meeting our inclusion criteria began in LMICs in the 1990s with the Jamaica Reach Programme and only broadened geographically in the 2000s. Experimental research in both sets of countries has picked up further since the 2010s.

In terms of program content, cognitive stimulation is included in the largest number of programs (54), although these are disproportionately concentrated in LMICs whereas HICs are split between the other three content types. Turning to delivery mode, home visiting covers a large majority (62) followed by group (29), both distributed relatively evenly between HICs and LMICs. Only five studies relied on ICT for delivery.

Of the programs evaluated in these studies, the Nurse-Family Partnership (NFP) was among the earliest and remains among the most widely known. NFP was developed by pediatrician David Olds and colleagues during the 1970s. The program is targeted toward disadvantaged mothers and calls for home visits by registered nurses from the third trimester of pregnancy up to roughly the child's second birthday, with weekly or biweekly visits depending on the developmental stage and the family's needs. During visits, the nurse visitors guide mothers and other family members in behaviors and activities conducive to physical health for the mother and the baby, encourage the building of supportive family and friendship networks, and connect the families to a variety of services (Olds et al. 2003, p. 281). We therefore categorize NFP as a lifestyle innovation according to the program content typology in our conceptual framework. NFP has been evaluated in three major RCTs that were eligible for our study sample: the first in Elmira, New York beginning in 1978 (Olds et al. 1986), the second in Memphis, Tennessee beginning in 1990 (Heckman et al. 2017; Kitzman et al. 1997), and the third in Denver, Colorado (Olds et al. 2004A; 2004B; 2014) beginning in 1994.

NFP has also been adapted into or inspired several programs around the world, two of which were evaluated in studies that appear in our sample: Pro Kind in Germany (Kliem & Sandner 2021) and Building Blocks in the United Kingdom (Robling et al. 2016). Pro Kind was evaluated in three German states with a sample of 755 participants. Most elements of the program model closely followed NFP using translated guiding materials, with the only notable difference being that home visitors were midwives and/or social workers rather than registered nurses given Germany's different health/occupational system. Visits were called for at weekly, biweekly, or monthly intervals depending on the time, from pregnancy to around the child's second birthday, with a maximum of 52 visits (Kliem & Sandner 2021, p. 2). Similarly, the Building Blocks program was adapted for teenage mothers 19 years of age and under in the United Kingdom and evaluated with a sample of 1,645 mothers in Wales (Robling et al. 2016).

The landscape of parenting programs within LMICs for the past two decades has been heavily shaped by the rise and spread of the Reach Up Early Childhood Parenting Program model (henceforth referred to as Reach Up). This model was pioneered among low-income households in Kingston during the 1980s through a series of piloting efforts that culminated in the highly influential Jamaica Longitudinal Study (Grantham-McGregor et al. 1981). The model has officially been adapted and implemented in numerous countries, and programs in many more countries have been influenced more broadly by its strategy and curriculum. Because the program content consists primarily of teaching parents activities aimed at encouraging infant and toddler learning, we categorize it as a cognitive stimulation program.

Pooled Impacts across Program and Study Characteristics

We next discuss the central meta-analytic results: pooled impact estimates across program characteristics, presented in Table 2. Results are shown for the full sample and separately by LMIC vs. HIC status since this distinction fits naturally with the existing literature and because this divide turns out to be among the most influential in shaping ES variation. Coefficients represent the pooled ES's of the programs belonging to the indicated categories, in standard deviation units using Hedge's *g*. The *n* cells represent the number of estimates belonging to the category in question, while *k* represents the number of studies. Summing across the HIC and LMIC subsets, there are a total of 225 estimates in the sample. Standard errors are replaced with "N/A" when there are fewer than four degrees of freedom, the minimum needed to calculate a reliable standard error (Tanner-Smith & Tipton, 2016), and full estimates are replaced with "N/A" when there are fewer than two degrees of freedom. τ^2 and Q provide measures of heterogeneity, with the former measuring the distribution of true effects and the latter of observed effects.

The overall ES pooling across all studies in our sample is 0.218 SDs. While both LMICs and HICs show highly statistically significant positive impacts, the former—at 0.303 SD—is nearly three times as large as the latter, at 0.115 SD. LMICs show substantially more heterogeneity than do HICS. These estimates are intended to be interpreted as benchmarks for contextualized comparisons rather than predicted impacts. The pooled ES's bring together results from numerous, diverse evaluations that vary widely in terms of both program and study characteristics. A key task of policy-relevant research is identifying which approaches tend to be most successful in which contexts. The remainder of this paper disaggregates the results to move toward that goal. In the next section we integrate a discussion of our results for curriculum type with a narrative review of particular studies and research programs since curriculum type in practice turns out to be most central to program identity in our sample, and then discuss variation in pooled impacts across other program characteristics.

Curriculum Type

The Curriculum Type panel of Table 2 indicates that cognitive stimulation programs show substantially higher child learning impacts than do the other categories in the full sample, followed by attachment, language, and then lifestyle. However, because the overwhelming majority of curricula in LMICs were categorized as cognitive stimulation, the most interpretable variation in impact across curriculum type occurs within HICs. Within HICs, attachment, cognitive stimulation, and language programs show similar pooled ES's, while the lifestyle program pooled ES is still significant but roughly half the magnitude of the others. There are too few LMIC studies with curricula other than stimulation to reliably calculate standard errors, but language and attachment programs have two studies each with large pooled ES's showing substantial promise. Comparing HIC against LMIC cognitive stimulation programs reduces but does not eliminate the differential between the two country categories. We next summarize the research on cognitive stimulation programs in LMICs and then HICs.

Given the prominence of Reach Up and models based on or inspired by it, the success of this model accounts for much of the strong cognitive stimulation program impact in LMICs. Of the 38 studies in our sample that evaluate cognitive stimulation programs in LMICs, 15 evaluated Reach Up programs or programs that explicitly drew inspiration from the Reach Up curriculum. We thus review the origins of the experimental Reach Up literature in the Caribbean before outlining the program's diffusion and research on other LMIC cognitive stimulation programs.

Reach Up was first evaluated in a study meeting our criteria via the Jamaica Longitudinal Study which launched in 1987 with a sample of 129 stunted children 9-24 months of age. Participants were cross-randomized in a factorial design with two treatments, a cognitive stimulation program and a nutrition program, yielding four experimental groups.

The cognitive stimulation treatment consisted of weekly hour-long visits from “community health aides” for two years (Grantham-McGregor et al. 1991). Immediately after the intervention, the home visiting treatment yielded a 0.88 SD impact on the children’s development quotients scores as measured by the Griffiths Mental Development Scales. Over the next several years, cognitive impacts faded until they were non-significant at ages 7-8, but then began to grow again, remaining into adulthood along with strong effects on income and other important outcomes. Meanwhile, the nutrition treatment began with a smaller impact, and this faded out entirely over the years (Grantham-McGregor et al. 1991; 1997; Grantham-McGregor & Smith, 2016; Walker et al., 2000; 2005; 2006). At 22 years old, individuals assigned to the stimulation treatment had IQs that were 0.6 SD higher than those assigned to control among other positive impacts on educational achievements (Walker et al., 2011) as well as earnings that were 25% higher (Gertler et al. 2014). The most recent analysis released when participants were 31 finds that income effects had grown even further, with treatment individuals reporting earnings 37% higher than the control group as well as hourly wages that were 43% higher (Gertler et al., 2021).

Variations on the model used in the Jamaica Longitudinal Study have shown significant positive effects in at least three other studies carried out within the Caribbean, albeit on shorter time horizons. Following the first major longitudinal evaluation of the Jamaica Home Visit intervention, a new iteration of the program was tested beginning in 1999-2000 with a sample of 140 infants born in the main maternity hospital in Kingston. This version included weekly visits by paraprofessionals for the first eight weeks using a newly-developed curriculum on mother-child interaction, followed by delivery of the cognitive stimulation curriculum from the previous longitudinal study for months 7-24. The authors found positive effects at a midline seven months into the program (Gardner et al. 2003), and again at 15 and 24 months (Walker et al. 2004). However, “benefits were smaller than those found in previous Jamaican studies" potentially because "the length of the visits was reduced from 1 h to 30 min" or because of differences in the sample (Walker et al. 2004, p. 1420). Significant effects persisted at age six, with effect sizes of 0.38 SD, 0.40 SD, and 0.53 SD on performance IQ, total behavior difficulties, and visual-spatial memory respectively (Walker et al. 2010).

A third study was conducted beginning with a sample of "139 undernourished children aged 9 to 30 months" across 18 clinics. The treatment consisted of one year of weekly 30-minute home visits by paraprofessional "community health aides". Whereas the previous studies had utilized community health workers hired from the government health service specifically for the program, this study used local community health workers (although they were supervised by the researchers, which could in part account for the high effect size according to Grantham-McGregor & Smith, 2016). They found an impact on the children's Griffiths development quotient of 0.94 SDs (Grantham-McGregor & Smith, 2016, p. 7). A fourth study carried out by several members of the same research team evaluated a separate program for a similar age range, this one on a larger scale with a sample of 501 mother-infant pairs in Jamaica, Antigua, and St. Lucia. This iteration of the program integrated delivery with clinic-based primary healthcare however, rather than home visiting. They find a 0.3 SD impact on cognitive development (Chang et al. 2015). Results were similar for an adapted, lower-cost/lighter-touch version of the original home visiting program. However, these interventions did not show significant longterm effects at six years of age (Smith et al. 2021).

Following the Reach Up model's early successes in the Caribbean, efforts began to experiment with this and alternative cognitive stimulation models in several other regions. In Bangladesh, one evaluation of a government-delivered program explicitly adapted from Reach Up found an impact of 1.1 SDs on children's cognition in a sample of around 718 children across 90 clinics. The treatment consisted of "play sessions" of 40 to 60 minutes held at government health clinics fortnightly for a year, administered in groups of two mothers with their respective infants at a time (Hamadani et al. 2019, p. 369). Three previous trials using similar treatment interventions had helped to motivate and develop this effort, with Hamadani et al. (2006) and Nahar et al. (2012) having found significant impacts on cognition of 0.33 SD and 0.37 SD in samples of 214 and 507 respectively, while Tofail et al. 2013 found a positive significant impact for nonanemic children, although no significant impact for the anemic subsample. The Hamadani et al. (2006) intervention was delivered through a year of home visiting and group meetings, while the Nahar et al. (2012) intervention involved 6 months of play sessions at clinics and Tofail et al. (2013) consisted of weekly home visits.

In addition, two programs with curricula similar to that of Reach Up—both run in collaboration with Save the Children—have also been studied with RCTs in Bangladesh. The first of these studies included a sample of over 2,500 across three districts, one of the largest sample sizes in our review. The intervention was implemented through the governmental National Nutrition Service using government health workers and existing clinical facilities. The study finds significant positive impacts on cognitive development "with effect sizes ranging from 0.08 to 0.14 for the Bayley outcomes" (Chinen & Bos 2016, p. 13). Although these effect sizes are modest compared to evaluations of comparable models, this program was implemented at a much larger scale and at a lower cost than the vast majority of these studies (p. 13). Second, Aboud et al. (2013) evaluated a 10-month program with a sample of 463 children ages four to 14 months in a western Bangladeshi subdistrict. They found a significant positive impact on cognitive development and parenting practices one month after the end of the program.

Four stimulation studies in our sample took place in India, two on a relatively large scale. The largest of these, Grantham-McGregor et al. 2020 find, in a sample of 1,449 in rural areas of the eastern state of Odisha, impacts of 0.324 - 0.281 SD on cognition and 0.239 - 0.302 SD on language. Meanwhile, Rao et al. (2020; see also Vazir et al. 2013) find an impact of around a third of an SD for a stimulation intervention in a sample of 521 in the southern Indian state of Telengana. Additionally, Nair et al. (2009) and Andrew et al. (2020) find impacts of similar magnitudes with samples of 800 (not counting attrition) and 421 in the states of Kerala and Odisha, respectively.

As evidence of the potential benefits of early childhood cognitive stimulation programs proliferated, several efforts arose to adapt or design stimulation programs in Latin America. Variations of the Reach Up model have been tested on a large scale in coordination with governments in three South American countries: one home visiting (Attanasio et al. 2014; 2020) and one group parenting (Attanasio et al. 2018) program in Colombia, and one home visiting program each in Brazil (Brentani et al. 2021), Chile (Carneiro et al. 2024), and Peru (Araujo et al. 2021A). The Chile and Peru studies, however, was not included within our meta-analysis sample since they did not include directly-measured learning outcomes at endline. The Brazil and Colombia programs were implemented in coordination with national conditional cash transfer (CCT) programs, as was a third study in Mexico using a curriculum that was independently developed but comparable to Reach Up (Fernald et al. 2017). With some caveats, all showed substantial promise.

In another adaptation of the Reach Up model Brentani et al. (2021) evaluated a home visiting program in Sao Paulo, Brazil, with a sample of 826 9-15 month-old infants born at a single large public hospital. Participants in the treatment group received biweekly home visits. Although ITT impacts were not statistically significant at midline (one year into the program), a per-protocol analysis shows a statistically significant ES of 0.22 SD.

Finally, *Educación Inicial*—implemented in Mexico in integration with a national CCT—used an independently-developed curriculum broadly comparable to the Reach Up model delivered within groups of around 20 families. Sessions were intended to be held weekly for around two hours, with "26 sessions for mothers, fathers and caregivers, 5 sessions for fathers, 18 sessions focusing on children, 8 sessions for pregnant women, and 5 concluding sessions at the end of the annual cycle" (Fernald et al. 2017, p. 224). The study included a final sample of 1,113 children across 204 communities. There were two treatment groups—one in which the parenting program was promoted via a larger cash transfer system, and one where it was not. Only the former showed significant ES's, which ranged from 0.26 – 0.29 SD. Results were driven primarily by impact within indigenous communities.

Three large-scale RCTs have been conducted on EC stimulation programs in Africa. Of these, the most recent large-scale evaluation tested a stimulation program with two experimental arms, one delivered via groups and one via groups combined with home visiting on a sample of 1,152 in Kenya. The found an impact of 0.52 SD and 0.42 respectively on cognitive and receptive language scores for the group-only arm while the mixed-delivery arm showed a 0.34 SD impact on cognitive scores but no significant impact on language (Luoto et al. 2021). Muhoozi et al. find an impact of around 0.59 SD in a group parenting program in Uganda with a sample of 511 that teaches parents about nutrition and sanitation practices in addition to in-depth instruction on cognitive stimulation. Rockers et al. (2016) study a program in Zambia that includes group trainings about stimulation and nutrition alongside home visits focused on health services (e.g., medical screenings and immunization encouragement). They find no impact on learning outcomes, although a positive impact on child anthropometrics is apparent in line with the program's lifestyle focus on health and nutrition.. Finally a stimulation program delivered via home visiting for young children with

HIV in South Africa, with a focus on teaching parents learning activities that can be incorporated into daily activities, found an impact of 0.27 SD (Potterton et al. 2009).

Additionally, results of three large-scale studies have been released since the late 2010s evaluating EC stimulation programs in China. Heckman et al. 2020 evaluate a program in China called China REACH (adapted from the Reach Up model) with a sample of over 1,500 in a county within a "poor region of western China" consisting of "more than 200 tasks related to language and cognitive skill development and…about 20 tasks targeting gross motor skills development" (Heckman et al. 2020, p. 5). The intervention was given for an hour weekly, and had yielded ES's around a full standard deviation after a year. A separate set of overlapping research teams evaluated three other program iterations in rural China based on Reach Up (albeit, more loosely than the program Heckman et al. 2020 evaluated)—two home visiting programs (Sylvia et al. 2018; Luo et al. 2019) and an individual center-based program (Sylvia et al. 2020; 2022). The former yielded an ES of 0.24 SD (Luo et al. 2019), while the latter did not show statistically significant impacts (Sylvia et al. 2022).

The cognitive stimulation programs in HICs (with a pooled ES of 0.18 SD) tended to be more diverse in content than those in LMICs. The pooled ES masks considerable diversity, with around seven studies showing no significant impact and seven showing substantial impacts of 0.18 SD and up and the rest showing smaller impacts. Only seven of the 16 studies in this category are coded solely as stimulation (Bierman et al. 2015; Doss et al. 2022; Drotar et al. 2008; Holtzman et al. 2023; Imhof et al. 2023; Knight et al. 2019; Scarr & McCartney 1988)—the others blend in substantial elements from other curriculum types. Four studies evaluate programs intended to complement preschool programs, including Head Start in the US (Bierman et al. 2015; Knight et al. 2019) and preschools in China (Leung, Tsang, & Heung, 2015; Leung, Tsang, & Kwan, 2017).

Attachment programs in HICs show a pooled impact only slightly lower than stimulation programs, but this average covers another divergence. Eight of the 16 attachment studies (Doyle, 2020A; Guttentag, 2014; Hilado, 2018; Landry et al. 2017; 2021; Leung, Tsang, & Kwan, 2017; Mendelsohn, 2007; Singla, Kumbakumba, & Aboud, 2015) included cognitive stimulation as well as attachment-oriented content while the remaining eight studies relied on attachment alone within our classification system. Programs coded as combining the two curriculum elements showed a pooled impact higher than that of the overall pooled ES for cognitive stimulation (0.24 SD for combined stimulation-attachment programs versus 0.18 SD for all programs with a stimulation component), while those with stimulation alone show a non-significant 0.05 SD pooled impact. This lower performance of attachment programs with no cognitive stimulation component is not necessarily surprising since the primary goals of attachment-oriented programs tend to be more social-emotional in nature. One attachment study (Lunkenheimer et al., 2008) even describes the school readiness outcomes presented in that article as “collateral benefits” in the title. It is certainly possible that social-behavioral benefits could help to improve academic outcomes along trajectories that are too longterm or subtle to be picked up in evaluations not specifically designed and powered to capture them.

Only three studies report on attachment programs in LMICs (Francis & Baker-Henningham 2021; Murray et al., 2015; Singla et al. 2015). Of these Francis & Baker-Henningham (2021) find no impact on learning outcomes, although there is a significant reduction in harsh punishment, the program’s main focus, along with other parenting behaviors. Murray et al. (2015) find a significant effect on learning outcomes for children coded as low-risk but not high-risk. The Singla et al. (2015) evaluation showed especially impressive results with a 0.34 SD impact on cognitive scores and a 0.27 SD impact on receptive language, although this program also included components on cognitive stimulation as well as lifestyle, including nutrition and maternal mental health.

14 studies evaluated language programs. These programs tended to cluster into two broad categories: offering guidance to parents on activities to improve language development (McGillion et al., 2017; Teepe et al., 2019; Wake et al., 2011; York, Loeb, & Doss, 2019), versus specific focus on encouraging shared reading (Chacko et al., 2018; Goldfeld et al., 2011; Piccolo et al. 2023; Weisleder at. 2018). On average, language programs show an impact only marginally lower than cognitive stimulation programs within HICs. Only two studies on language program took place in an LMIC, both in Brazil and both focused on shared reading: Weisleder et al. (2018), which showed strong impact on language outcomes and Piccolo et al. (2022), which did not.

Finally, we consider the effects of lifestyle programs, which showed the smallest ES (albeit still positive and significant) of all curriculum categories at 0.13 SD in the full sample. Although these ES's may appear modest, it bears remembering that improving children's cognitive and language development may not have been among the main objectives for many of these programs (as discussed above with regard to attachment programs). The five LMIC studies (Jeong et al. 2023; Muhoozi et al. 2017; Rockers et al. 2016; Singla et al. 2015; Tomlinson et al. 2016) show a large coefficient of 0.25 SD, albeit not statistically significant because of the large standard error and small study sample size. However, all of these studies additionally include stimulation components with the exception of Tomlinson et al. (2016). This latter study finds a small but statistically significant effect size of around 0.10 SD on vocabulary which the authors describe as "notable, as cognitive stimulation is not a specific focus" (p. 944).

Lifestyle programs in HICs, in contrast, show a modest but statistically significant pooled impact of 0.08 SD. For lifestyle programs in HICs, the example of NFP and its offshoot programs provides an instructive example of how these programs can be generally effective on important outcomes but still yield relatively low short-term child cognitive and language score impacts. Given these programs' prominence, we review the findings in detail in the present section, beginning with the three RCTs of NFP iterations: Elmira, New York, beginning in 1978 (Olds et al. 1986), Memphis, Tennessee beginning in 1990 (Heckman et al. 2017; Kitzman et al. 1997), and Denver, Colorado (Olds et al. 2004A; 2004B; 2014) beginning in 1994.

The Elmira study did not show statistically significant effects on learning outcomes in the full sample at the program's end, although there were substantial effects on parenting behavior and reduction in emergency room and physician visits, two of the main intended outcomes (Olds et al. 1986). Learning outcomes were also non-significant at the end of the Memphis intervention (Kitzman et al. 1997), but statistically significant impacts were registered during a followup four years later when the children were around six years old, and boys continued to show statistically significant treatment effects at age 12 (Heckman et al. 2017). At age 18, although no statistically significant effects remained for the full sample, children whose mothers had exhibited "low psychological resources" showed significant treatment effects of 0.24 SD and 0.38 SD on receptive language and math achievement, respectively.

The Denver study showed a significant impact when participants were around four years old of 0.18 SD and 0.17 SD on mental processing composite and reading achievement scores, respectively (Olds et al. 2004A). By age six, there was no significant impact for the full sample, but statistically significant impacts of 0.31 SD and 0.47 SD on total language and executive function composite scores, respectively, remained for the nursing treatment. The paraprofessional group saw no significant effects in the full sample and a 0.29 SD impact on executive function in the low resource sample (Olds et al. 2004B). At age 9, there was a marginally significant effect of 0.30 SD on a visual attention/task switching test for participants whose mothers exhibited "low psychological resources), but no other effects remained (Olds et al. 2014).

Results for NFP's offshoots—Pro Kind in Germany and Building Blocks in Wales—were generally similar. Despite showing substantial benefits on a range of outcomes, Pro Kind's impacts on cognitive and learning outcomes were relatively minor. Although there were significant impacts shown within the full sample during midline tests, by endline the cognitive impact was non-significant in the full sample, although there was a statistically significant impact on the cognitive index for girls of around 0.20 SD. Similarly, effects on language were non-significant with the exception of a few sub-tests for girls (Sandner & Jungmann 2017). The same situation persisted a year after the program at age three, although with small positive impacts reported for "high risk" children as well as for girls (Jungmann et al. 2015). The most recent results come from a follow-up conducted when the children were around seven years old (with 533 of the original 755 participants); although a range of statistically significant impacts remained, there were none for learning outcomes (Kliem & Sandner 2021).

In contrast to Pro Kind, the Building Blocks evaluation found no substantial effects on the study's primary outcomes, which did not include learning. However, while learning outcomes were pre-registered as secondary, the posttest found small positive impacts on a language test administered to children, in addition to mother-reported cognitive and language outcomes (Robling et al. 2016). At age seven, impacts school-reported learning outcomes were consistently positive, albeit with a mix of significant and non-significant coefficients. Among the most impressive outcomes were a roughly 25 percent increase in likelihood of achieving a "good level of development" on five government-specified subjects, as well as a positive impact on passing reading and writing thresholds, whereas math and science coefficients were non-significant (Robling et al. 2021).

Delivery Mode

As shown in the "Delivery Mode" panel of Table 2, home visiting and group delivery—the two most common delivery modes—show nearly identical pooled ES's at 0.21 SD and 0.23 SD respectively. Home visiting shows a greater pooled impact than group delivery in LMICs and vice versa in HICs in terms of impact magnitude. Since home visiting is typically a more expensive delivery mode than group sessions, the lack of evident advantage shown for home visiting is noteworthy. We discuss this further below with reference to policy implications and compare a handful of instances in which the two delivery modes are compared against one another in the same study using the same program content.

Turning back to Table 2, the individual-clinic delivery mode yields the highest point estimate, albeit with a smaller study sample and larger standard error than home visiting and group delivery. The pooled ES for the five LMIC studies (Chinen & Bos 2016; Hamadani et al. 2019; Nair et al. 2009; Potterton et al. 2009 Sylvia et al. 2022) is quite large at 0.40 SD, just over twice the ES for the HIC programs that used this delivery mode. Results were promising in almost all of the studies, with Hamadani et al. (2019) showing the largest ES's in our sample at over a full SD for both cognitive and language outcomes. Three of the four individual-clinic studies in HICs showed significant impacts on at least one eligible outcome (Kim et al. 2021; Mendelson et al. 2007; Roberts et al. 2014), while Rauh et al. (1988) showed no impact on learning outcomes.

ICT represents the only delivery mode with a negligible pooled ES. However, with only five studies there are too few degrees of freedom even to reliably calculate a standard error using RVE much less represent a fair test of the delivery mode's potential.

Experimental research on ICT-delivered parenting programs is still in its infancy, with all studies in our sample published since 2019, and all conducted in developed countries. Given the small study sample size, evidence on ICT-delivered parenting programs is best considered using individual studies rather than pooled effect sizes. Two of the studies showed statistically significant positive effects of 0.11 SD (York, Loeb, & Doss, 2019) and 0.21 SD (Chamberlain et al., 2021), according to our ES calculations. Both consisted of thrice weekly text messages providing tips on literacy development for preschool-aged children. Doss et al. (2022) evaluate two versions of a similar program that focuses on math, one consisting entirely of math and the other additionally including language and social-emotional learning. Despite the lack of overall treatment effects, the authors find promising results for girls and children at the lower end of the distribution. Interestingly, the two ICT-delivered programs involving higher dosages and likely higher program costs and time costs for parents—one consisting of 20 program hours alternating between online modules and home visits (Landry et al., 2021) and the other text messages that linked to activity guides and videos (Doss et al., 2022) did not show significant impacts. A fifth study on an early math intervention showed a statistically impact on one math domain but not others (Holtzman, Quick, & Keuter, 2023).

Timing and Dosage

Turning to the variables in the timing and dosage panel of Table 2, coefficients for all child age ranges at baseline—for the full sample as well as separated by LMIC vs. HIC status—are positive and statistically significant, with the exception of studies with the 0-1 baseline age range in HICs. The 1-2 ranges show the highest magnitude. The “Session Frequency” panel indicates that more sessions per week are associated with larger effect sizes for LMICs as well as for HICs. Many programs have only a fraction of a session per week, in some cases being composed only of a handful of sessions spread out over months or even years.

Pooled ES's for all program duration lengths are also positive and statistically significant, with the exception of that of HIC programs lasting longer than two years, which barely misses statistical significance. Generally, programs of shorter duration seem to be associated with larger ES's, with a steadier gradient within LMICs. While counterintuitive on first glance, this finding may at least in part reflect less frequent sessions in longer programs. More broadly, the smaller program effects at endline for education interventions with longer durations has been previously noted in the literature (see Nickow, Oreopoulos, & Quan, 2024, p. 94).

Robustness Checks and Exploratory Analyses

In the present section, we attempt to further unpack the above findings in order to assess the extent to which they may be influenced by study characteristics and selection bias. We first explore variation in pooled ES's by key study characteristics, and then add study characteristics into multivariate regression models alongside key program characteristics. Finally, we assess the extent of selection bias.

Effect Sizes by Study Characteristics

Pooled ES's broken down by study characteristics are presented in Table 3. First, we consider the relationship between sample size and ES since the stronger the correlation between the two, the more concerns there are to be addressed in program scaleup. The results in Table 3 are thus reassuring, at least for LMICs. Although studies with the smallest sample size range—50-200—show the largest pooled impact, studies with sample sizes above 400 show a comparable pooled impact and do not appear to continue to decline further with sample size. Pooled ES's for HIC programs above the smallest sample size category are, however, non-significant. Some decline in impact magnitude is expected with larger samples given a variety of supply- and demand-side implementation issues that come with scaling, some of which are discussed below. Table 3 additionally shows that impact magnitude is similar across cognitive and language outcomes. While pooled impact was lower for the eight studies from the 1980s, each of the remaining decades has a pooled impact exceeding 0.20 SD.

Multivariate Regressions

Examining variation in pooled ES's across categories rather than multivariate meta-regression affords the strongest approach to describing variation in program effectiveness for this study given limited statistical power and variables that are potentially ambiguous to code. However, at an exploratory level multivariate meta-regression framework allows us some additional leverage in disentangling different program and study elements from one another and charting the robustness of results. We next analyse a series of multivariate regressions aimed at further disentangling the relationships between program characteristics, study characteristics, and program impact to the extent possible given limited degrees of freedom. Results are shown in Tables 4A-4D. Each table begins with a model that includes two key study characteristics—a dummy variable indicating LMIC status, and log sample size. Table 4A adds content and delivery, 4B age and dosage, 4C content, age, and dosage, and 4D all sets of variables.

LMIC status is statistically significant in nearly all specifications, with the exception of the specifications that respectively include all covariates (column 7) in Tables 4A, 4C, and 4D, as well as column 3 in Table 4C which includes curriculum and child age. Statistically significant coefficients range from around 0.10-0.20 SD. The sample size coefficients are negative and significant for many of the models, but the relationship is far less robust than LMIC status, losing significance in most of the specifications that include covariates beyond LMIC status.

Turning to program content, lifestyle program status shows statistically significant negative coefficients across all specifications for which it is included, while language and attachment programs show negative coefficients that are only occasionally signification (stimulation, the highest scoring curriculum classification is the reference category). Delivery mode coefficients are generally non-significant except for some negative coefficients on ICT (with home visiting as the reference category). Categories representing greater program frequency show significant positive coefficients in some but not all models (with the lowest frequency, 0.25 days per week or less, as the reference category). On the other hand, program duration dummies are non-significant in all specifications. Of the age dummies (with prenatal as the reference category), those with children aged 1-2 at baseline show significant positive coefficients in around half of the specifications in which they're included.

Accounting for Selection Bias

Tables 5A and 5B show the main pooled ES's (leftmost columns) alongside Egger's regressions that add a covariate representing a transformed standard error, labeled W_i. The middle columns show the coefficient for W_i. Systematic patterns of statistically significant tests for the W_i coefficient could constitute evidence of selection bias. However, Egger's regressions cannot distinguish between real differences varying across sample sizes (which we treat as a substantive issue relating to scaling), and statistical artifacts that would occur if, for example, studies with significant findings are more likely to be published (i.e., "selective reporting"). Additionally, substantial between-study heterogeneity can inflate the likelihood of significant coefficients in Egger's (Rodgers & Pustjovsky, 2021, p. 146). We therefore present regressions both with W_i as the only independent variable (Table 5A) and with the addition of a dummy signifying membership in the smallest sample size (50-200) since small-sample studies are most susceptible to selective reporting.

The results overall are reassuring. As shown in the first row of Table 5A, the coefficient for W_i is not statistically significant in the full study sample even with no control variables to account for heterogeneity. Several subsets of studies generate statistically significant coefficients: the coefficients for home visiting, group, and language programs are significant at p=0.10, the coefficient for HICs is significant at p=0.05, and the coefficient for a baseline age of two or older is significant at p=0.01. Some significant coefficients would be expected by random chance given the large number of statistical tests conducted. The results could indicate that pooled ES's for HICs and studies with a baseline age of two or older should be interpreted with more caution. However, adding a dummy for small sample size (50-200) completely removes significance from all estimates. Heterogeneity attributable to real differences resulting from sample size is directly explored as a key part of this review and does not constitute a validity threat.

Discussion

Limitations

Our analysis faces several important limitations. For one, a key reason behind EC programs' potential effectiveness lies in the fact that they occur early on in the life course and thus have the potential for transformative effects in the longterm. However, there were too few estimates for longterm outcomes to include those in our quantitative analysis. Instead, we consider them in the qualitative review.

Further, the wide range of studies included in our review means that attention to—and particularly quantitative coding of—subtle differences in program content and implementation is not possible. In addition to the inherent complexity and diversity of these program characteristics, descriptions across studies were uneven. While we are confident that the categories we used to structure our analyses accurately capture meaningful and empirically-grounded distinctions in the intervention technologies employed, the categories are relatively coarse. Relatedly, categories are not equally well represented because of the uneven nature of the experimental literature to date, and pooled impact estimates associated with particular categories of course only represent the programs evaluated in studies that were eligible for inclusion in our sample.

Additionally, we are not able to comment substantially on intra-household dynamics, including gender and household dynamics that may play a central role in shaping the extent of program efficacy. The majority of studies in our sample focused on mothers with relatively little discussion of alternative possibilities, although there have also been studies focusing on both parents or fathers in particular. The nature of the available research unfortunately afforded little opportunity for delving into the ways in which these dimensions may contribute to shaping the effectiveness of the parenting programs. Finally, because of our decision to focus on the effects of parenting programs in particular, we are not able to comment on programs that combine parenting with center-based programs that directly serve children.

Implications for Scale-Up

The analysis so far has demonstrated that early childhood parenting programs can generate substantial impact in a wide range of contexts. However, a more nuanced approach is required to draw out clearer implications for social science and policy. Taking programs to scale requires close consideration of a complex range of factors for the vast majority of social policies (Al-Ubaydli, List, & Suskind, 2019). This is especially the case for EC programs given the EC sector's fragmented institutional domain and other considerations (Gupta et al., 2021). Alongside the quantitative meta-analysis, we closely read each study and considered how subtler trends in findings within and across studies might shed additional light on what works most effectively in which contexts. Specifically, we focused on the findings from the more recent, larger-scale studies, paying special attention to comparisons of multiple treatment groups within the same or similar studies and other contextualized dynamics.

A central topic implicit in much of this discussion that bears direct acknowledgement is cost-effectiveness, given its centrality to scaleup and policy considerations. Unfortunately, only a small handful of studies within our sample reported cost-effectiveness data, too few to be meaningfully included within the meta-analysis. And, although cost-effectiveness is critical, specific costs are likely to vary substantially across time, location, and program context, and it may be equally valuable to consider the likely relative costs of broad structural parameters, rather than the detailed costs of a specific instantiation of a particular program. These considerations underly much of the below discussion, and we cite specific cost-effectiveness findings where possible.

Optimizing Program Delivery

All else equal, group parenting programs may save costs relative to home visiting programs by avoiding the time and transportation expenses of sending staff to individual participants' homes (although this could mean simply passing those costs onto parents). Additionally, providing the program to multiple participants at one time could be a more efficient use of interventionists' time. Groups may reduce the interventionists' attention to individual parents, but conversely the group setting could strengthen impact if parents benefit from interacting with one another in this context. Given the meta-analytic findings that the group parenting programs in our sample were at least as effective on average as home visiting programs, group parenting programs may in the majority of contexts prove more cost-effective than home visiting. Individual clinic-based programs, particularly if they can occur when the parents are going to a clinic anyway such as for well-child visits (e.g., Chinen & Bos, 2016) may lie somewhere in the middle in that they save transportation costs (or pass them onto parents) but demand more of the interventionists' time per parent.

Only two studies in our sample, both recent large-scale evaluations, directly compare group and home visiting treatment arms in a single study: Grantham-McGregor et al.'s (2020) evalution of a Reach Up iteration implemented in Odisha, India and Luoto et al.'s (2021) evaluation of a cognitive stimulation program in Kenya. Grantham-McGregor et al. (2020) test two treatment arms with the same curriculum, but one is delivered via home visiting as the program was originally designed for, and the other is administered via group parenting sessions. The treatment arms showed statistically indistinguishable effects, despite the group sessions requiring a fraction of the cost. The authors also report a cost analysis, which shows that the home visiting arm cost about 3.5 times more than the group arm, $135 versus $38 per child. The group delivery arm was thus much more cost-effective.

Similarly, Luoto et al. (2021) compare a group treatment arm against a mixed group-home visiting treatment arm in western Kenya, finding stronger impacts for the group-only model. Given that costs were similar ($140 per child for group-only and $145 per child for mixed), the group-only model was also more cost-effective, with a cost-analysis calculating an incremental cost-effectiveness ration of 0.37 SD cognition improvement per USD$100, a benefit-cost ratio of 15.5, and a return on investment of 127% (García, Saya, & Luoto, 2021). No treatment arm consists solely of home visits because "individual home visits in dispersed rural settings such as those found in Kenya would be prohibitively time-consuming and expensive for scaling" (Luoto et al. 2021, p. e310)." This highlights the point that the geography of settlements and transportation infrastructure can critically shape the relative cost-effectiveness of home visiting models versus those held at a central clinic.

Two separate but overlapping research teams evaluated EC parenting programs in Colombia around the same time, also using an adapted version of the Reach Up curriculum, one using home visiting and the other using group delivery (Attanasio et al. 2014; 2018). Effects were similar; the main impact point estimate is marginally larger for home visiting, but as the authors point out, the group program was run almost entirely by the government while the researchers were more directly involved in the home visiting program. The home visiting program also lasted longer and had a pure control group while the control group in the group-based program received a different group intervention (unrelated to parenting). These studies thus also provide some support for the cost-effectiveness of group delivery.

However, another comparison of two closely connected studies highlights the importance of caution in generalizing this finding without attending to local context. Sylvia et al. (2020; 2022) conducted and compared two studies in western China with similar populations and intervention content, but one with home visiting and the other via individual center-based sessions. In this case, the authors find that the home visiting yielded significant impacts, while the center-based program did not. This divergence seems to have arisen from the fact that the positive impacts of the home visiting were driven by effects for the more disadvantaged participants, and these same participants were more likely to drop out of the clinic-based sessions: "the greater compliance of more vulnerable children in the home-visiting program may have led to larger average impacts than in the center-based parenting program" (Sylvia et al. 2020, p. 4-5). This finding highlights the fact that home visiting programs may be worth the added investment to the extent that doing so increases equitable access. Unless care is taken to avoid this, some population segments may be excluded by a group or individual center-based programs even in cases where the latter are more cost-effective on average.

Overall, this discussion shows that group parenting programs are likely to be more cost-effective than home visiting programs, but which one is more cost-effective in a particular situation ultimately depends on context-specific factors relating to population density, transportation structures, and parent propensities. Even where group programs do exhibit strong and significant impacts, it is important to ensure that the most disadvantaged households are not being excluded. Relatedly, at least some of the cost-effectiveness advantage of a group parenting program over home visiting may arise from simply shifting transportation costs onto participants and thus off of the implementers' books. This point raises the issue of considering participant as well as implementer costs more broadly. Home visiting may thus best be considered as an option where transportation is expected to be burdensome to parents or otherwise prohibitive of takeup. Research on individual-clinic delivery models is much thinner, but some promising evidence has already emerged for integrating EC parenting sessions into medical appointments that parents would be taking their children to anyway, and using them to supplement home visiting (Chinen & Bos, 2016).

Interventionist Characteristics

The relative cost-effectiveness of programs that leverage interventionists with different characteristics represents another key area for scaleup consideration that overlaps substantially with delivery mode. Interventionists with professional certifications like nurses may have acquired skills that strengthens the effectiveness of their delivery, but they may also require greater compensation and be more difficult to recruit at scale depending on local labor markets, and they may have greater difficulty in relating to the target population.

Specific labor market and socio-cultural structures associated with a given intervention area should thus be considered in deciding on interventionists.

Only one study in our sample—Olds et al., (2002)—directly compare two types of interventionists within the same experiment with other factors remaining the same. This study was primarily oriented around health and behavioral outcomes that are outside the scope of the present review, but the arm with nurses as interventionists seems to have had a more robust impact overall. However, the study evaluated the Nurse-Family Partnership, an HIC lifestyle program that was designed around nurses. Reviewing evidence from a wide range of Reach programs, Attanasio, Cattan, & Meghir (2022) conclude that local rootedness and program-specific training override a need for professionally-certified interventionists: "…emphasis on local women with no specific qualifications, but a strong training on the curriculum delivery, clearly solves any problem of scarcity of human resources and has the potential to make the program culturally more acceptable within their communities" (p. 875).Still, strong training on program procedures seems essential. Araujo et al. (2021B) cite the two-week preservice and one week in-service training in Colombia as one among several potential reasons it may this iteration of Reach Up showed stronger impacts than a larger-scale version in Peru (where training and supervision standards were harder to maintain given the larger scale), which included only four days of pre-service training and no in-service practice (p. 21). Supervision was also much stronger in Colombia than in Peru. More broadly, Araujo et al. (2021B) conclude from this comparison that training/mentoring is one of two key "non-negotiable components" that must be carefully maintained during scaleup, alongside program content which, when sufficiently-well structured, can work alongside training and supervision to make up for a lack of education or experience.

While the above models all depend heavily on interventionists' time—which likely take up a substantial portion of costs—programs delivered via ICT offer the possibility of disseminating parenting advice while bypassing the need for interventionists altogether (other than content designers and program administrators). Text messages and other ICT platforms have also been shown to be effective at precipitating behavioral change in a variety of other contexts. These programs might especially offer added value given that changes in parenting behavior are often constrained by behavioral frictions—e.g., limited bandwidth—in addition to or instead of informational frictions (Gennetian, 2021). Because digital messaging is so inexpensive and digital platforms are so central to daily life for some populations, ICT offers the opportunity to provide real-time reminders and nudges at relatively low cost.

The evaluation research for ICT-delivered EC parenting program is in its infancy, but three of the ICT interventions in our sample show promising results (Chamberlain et al., 2021; Holtzman, Quick, & Keuter, 2023; York, Loeb, & Doss, 2019). The remaining two (Doss et al., 2022; Landry et al., 2021) did not show significant impacts. However, as noted above, all content was contained within the text messages for the programs that showed impact, whereas the two that did not required parents to access online learning modules and videos. In light of the above distinction between behavioral and informational frictions, this disjuncture may represent early evidence that the comparative advantage of the ICT mode reside at the behavioral end, helping to provide reminders and small bits of information but less effective at communicating more information-intensive material.

Administrative Capacity and Integration

A close reading of the larger-scale studies reveals that difficulties associated with mobilizing administrative capacity represent a key binding constraint to cost-effectiveness in many contexts. Administrative structures—including organizational machinery, technical infrastructure, and the skills and human capital of administrators—may prove at least as influential in shaping impact variation as substantive program features like curriculum, or the nature of local needs. This point is highlighted by insights from a recent evaluation of a large-scale government-run initiative in Brazil using Reach Up curriculum. While the study was an RCT with parents randomly assigned to treatment or control groups, parents in the treatment group received two variations of the treatment, depending on their location of residence. One of the variations was implemented by an existing government agency, and the other by a newly-developed cadre. Only the latter generated statistically significant impacts, and even then only in a per-protocol analysis (no statistically significant effects were found for intent-to-treat). The authors found that the employees of the existing agency had too little disposable time to implement the parenting program (Brentani et al., 2021).

One potential way to maximize administrative capacity at scale is to integrate parenting programs into existing social support programs–provided staff time and other resources are sufficient. Two studies highlight successful examples integration with conditional cash transfer programs: one of the two Colombia studies discussed above (Attanasio et al., 2014), and a program in Mexico called *Educación Inicial*. This latter program was developed by the Mexican government (with curriculum loosely inspired by the Reach Up model), and implemented in conjunction with the national cash transfer program, Prospera – evaluated by Fernald et al (2017). There were two treatment arms, one in which the parenting program was promoted by the cash transfer program, and the other where it was not; only the treatment arm with promotion attained enough dosage to show statistically significant impacts. This finding highlights the important role of administrative capacity not only in delivering the program but also in ensuring takeup. This is particularly critical given that the families who could most benefit from EC parenting programs tend to be the most disadvantaged, and these same families may be least likely to find and join these programs without additional outreach steps (García & Heckman, 2023).

Where early childhood parenting programs are effectively delivered, results can be very large. To take a positive case, programs in Bangladesh have shown some of the greatest impacts; while it is not possible to prove why that is exactly (and it probably depends on many factors), the success of these programs seem explicable at least in part because Bangladesh contains a dense population of poor households willing to take up EC development inputs while also having a public health infrastructure with some relatively robust service delivery characteristics (at least when NGO support is taken into account) that is able to effectively deliver programs. In contrast, an illustrative (albeit anecdotal) counterexample is Cambodia, which likely has at least as great a need for development inputs as Bangladesh. In a recent evaluation of an EC parenting program here, the authors report that there was very little implementation (Bouguen et al. 2013). In cases where governments lack administrative delivery capacity or where civil society plays a strong role in education or health delivery, NGOs and nonprofits can help to fill the gaps. Included in our sample are, for example, highly successful programs led by Save the Children in Bangladesh (Chinen & Bos 2016) and China (Luo et al. 2019) and Pratham in India (Grantham-McGregor et al. 2020).

Exploring Longterm Trajectories

Among the potential advantages motivating EC programs is that they can lead to skill improvements which can then help to provide the basis for more efficient skill development over the course of the life, creating more opportunity for transformative impacts. What does the literature say about such a possibility? Unfortunately, longterm effects are difficult and cost-intensive to collect, and too few of the studies in our sample examine outcomes more than a year or two after the end of an intervention, precluding rigorous meta-analysis. Jeong, Pitchik, and Fink (2021B) present meta-analytic results for longer-term followups in LMICs, occurring more than a year after program inclusion, generally finding reduction in effects over time. But their sample consists only of seven studies, most of which as the authors acknowledge were small in sample and not adequately powered to capture longerm results. Their main conclusion is thus that the analysis is inconclusive and more long-term studies are needed.

While we felt that calculating meta-analytic results for long-term followups would be unhelpful at best and misleading at worst, there are enough longterm studies in our sample to warrant a close qualitative reading and narrative review to distil insights. We identified 14 studies from our sample that report outcomes more than a year after the intervention. The results indicate substantial promise—albeit with the nonlinearities, variation in fadeout, and occasional lack of longterm statistical impact which likely result from the complex interaction between skill and institutional factors in shaping the course of human development as well as the methodological difficulties inherent in longterm studies.

Perhaps the most dramatic longterm success thus far in our sample is the Jamaica Home Visiting program, which continued to show dramatic effects on income three decades after the program when the participants were 31 years old (Gertler et al., 2021). Another of the Jamaica studies collected results when children were six and found persisting impacts on cognitive outcomes as high as a half standard deviation. However, a variation of the program carried out in three Caribbean countries using clinic-based delivery rather than home visiting showed no significant persisting effects at age six (Smith et al., 2021). And the otherwise promising variation of Reach carried out in Colombia (Attanasio et al., 2014) did not show continuing significant impacts two years after the program ended (Andrew et al. 2018). Observable impacts on parental behavior had also declined by this point, so the authors suggest that the decline may have resulted from a change in parental beliefs around too narrow an age range alongside the general tendency of impacts to decline as a result of limits to the self-productivity of most skills (Attanasio, Cattan, & Meghir, 2022, p. 882).

Following the original Jamaica Home Visiting study, three cognitive and/or attachment oriented home visiting programs in HICs find the most promising longterm results in our sample. A recent followup to the evaluation of Preparing for Life—a home visiting program with cognitive and attachment-oriented elements—shows undiminished impacts on cognitive and school achievement scores at age nine, albeit with some fadeout in non-cognitive scores (Doyle 2020B). The Los Angeles portion of the Legacy for Children attachment home visiting evaluation—the only one of the two sites that saw significant impact at endline—found continuing impacts of four IQ point, roughly a third of an SD six years after the program had ended, although no significant impacts on language (Perou et al., 2019). The REDI cognitive stimulation home visiting program—designed as a Head Start supplement—continued to show learning outcomes three years after the intervention when participants were in second grade, with respective impacts of 0.30 SD and 0.50 SD on English and math (Bierman et al., 2015). Brennan et al. (2013) find using SEM analysis that the Family Check-Up attachment-oriented home visiting program was associated with higher academic achievement at ages 5 and 7.5.

On the other hand, Black et al. (2007) find no impact of a cognitive stimulation and lifestyle home visiting program on learning outcome when participants had reached age 8, but this is not surprising since the original evaluation only showed statistically significant impacts for a subset of the sample (younger children), and the study was conducted with a sample from a relatively specialized population, children with non-organic failure to thrive.

NFP and its offshoots also show robust evidence of longterm effects, albeit less consistently than some of the above studies for learning outcomes. Historically, the NFP evaluations showed lasting substantial impacts on an array of important outcomes sufficient to motivate an influx of interest and development in home visiting programs and related research. Its longterm impacts on learning outcomes, while more mixed, still shows signs of continuing impact. As discussed above, both the Memphis (Heckman et al., 2017) and Denver NFP (Olds et al., 2014) evaluations involved longterm followups through ages 18 and 9, respectively. It is especially noteworthy that the Memphis study's original endline showed no significant impacts on learning outcomes, but significant impacts showed up in the full sample at age six (four years after treatment) and continued for some subgroups at ages 12 and 18 (Heckman et al., 2017). Turning to the NFP offshoot programs, Building Blocks continued to have a significant impact on several learning outcomes when participants had turned seven (Robling et al., 2021). Pro Kind continued to show impacts on several outcomes at, but not on any learning outcomes (Kliem & Sandner, 2021).

Carneiro et al. (2024) present an evaluation of a low-cost, eight-week parenting program delivered in Chile at scale using existing government resources. Their study is not included in our meta-analysis since eligible outcomes within a year of the intervention's end are not reported, but remains highly relevant given the large scale and incorporation with government. They find an impact of 0.1 SD on child vocabulary three years after the program ends. While modest, this shows strong success given the short duration and low costs of the program and long time period following the intervention.

Several premises emerge from the above discussion. First, the number of longterm significant impacts is encouraging, particularly in light of methodological difficulties associated with longterm followup and also given limitations in the availability of measures to capture the full range of skills taught. Second, impacts often seem to unfold in a nonlinear trajectory, in some cases disappearing in one followup wave only to reappear years later. Third, longterm significant impacts often show up for some variables and not others in ways that may not necessarily have been predicted in advance. These latter two points make sense when viewed in light of the literature in economics and human development showing that various types of cognitive and non-cognitive skills interact with one another through complex patterns of self-productivity and cross-productivity over the life course (Attanasio, Cattan, & Meghir, 2022, p. 861; Cunha, Heckman, & Schennach, 2010).

Findings in Context

Up to this point, we have primarily considered impact magnitude by comparing findings for studies and study clusters in our sample relative to one another. How do our findings compare to those of other reviews of parenting programs and alternative EC-oriented social programs? In this section, we first position our findings within the context of other parenting program reviews, and then within the broader field of EC development.

In the review that overlaps most extensively with ours in terms of study coverage, Jeong et al. (2021A) present meta-analytic pooled impacts for 102 RCTs (published 1974-2020) of parenting programs in HICs and LMICs involving children who are up to three years old at baseline. This review complements ours methodologically in that they focus almost entirely on presenting meta-analytic results on a range of outcomes, while ours focuses exclusively on one set of outcomes (i.e., learning as measured from direct tests of children) and accompanies the meta-analytic results with a detailed narrative analysis. Jeong et al. (2021A) report pooled impacts of 0.32 SD on children's cognitive development and 0.28 SD on children's language development, similar to our findings (0.266 SD for cognitive outcomes and 0.241 SD for linguistic, when restricting our sample to a child age of 36 months or less at baseline). Aside from the differing study samples, their impact estimates may have been inflated relative to ours because of their inclusion of cognitive scores derived from parent reports, like the Ages and Stages Questionnaire. Additionally, Jeong et al. (2021A) find significant impacts on three other child development outcomes: motor (0.24 SD), socioemotional (0.19 SD) and attachment (0.29 SD) as well as reductions in behavior problems (-0.13 SD). Impacts on parenting knowledge, practice, and parent-child interactions are even stronger, ranging from 0.33-0.56 SD), although the pooled impact on parents' depression was not significant. Like our review, they find stronger pooled impacts for LMICs than HICs. Their delivery mode breakdown differs from ours, but their findings like ours show similar or even slightly lower impacts for home visiting than alternative modes.

Zhang et al.'s (2021) review also focuses on meta-analytic results for parenting program RCTs, but they limit their sample to developing country programs with an average baseline child age of two and under. They estimate pooled impacts of 0.44 SD and 0.33 SD on cognitive and language development, respectively. The corresponding estimates we calculate when limiting our study sample to LMIC programs with children 24 months and under at baseline are quite similar: 0.38 SD and 0.29 SD. Although their estimated impact on social-emotional development is not statistically significant, this may in part result from a relatively small study sample given the substantial positive coefficient (0.17 SD). Additionally, they find a stronger advantage of group sessions (0.53 SD for cognitive and 0.38 for linguistic) relative to home visiting (0.28 SD for cognitive and 0.17 SD for language) than our study or Jeong et al.'s (2021A). For LMIC programs in our sample with children 24 months and under at baseline, we estimate impacts of 0.369 SD and 0.275 SD for group program effects on cognitive and language outcomes, respectively, versus 0.370 and 0.241 for home visiting programs.

Filene et al. (2013) present a meta-analysis of 51 evaluations of home visiting programs in the United States between 1979 and 2010 for children ages three and under at baseline. They find an impact of 0.25 SD on child cognitive outcomes for the 24 studies in their sample falling into that category. On the other hand, we find a smaller—albeit still statistically significant—impact of 0.082 SD on cognitive outcomes for HIC home visiting programs for children ages three and under at baseline. The difference might be attributable to inclusion of parent-reported and/or researcher-designed measures or differences in estimation technique given that meta-analysis methodology has changed substantially since that study's publication.

Conclusion

In this paper, we have presented results from a systematic review and meta-analysis of EC parenting programs aimed at generating evidence for policy and scaleup. Specifically, we focused on learning outcomes, including cognitive, linguistic, and achievement, measured directly from children. We found that the evaluated programs have proven effective on average across a broad range of circumstances, including location, delivery model, and curriculum. Effect sizes have been substantially larger in LMICs than in HICs. While home visiting and group parenting—the two most prominent delivery models—have been comparably effective, curricula based on cognitive stimulation tend to outperform other categories in HICs (studies in our sample from LMICs focus almost exclusively on programs using cognitive stimulation). Language and attachment-based programs in HICs also show promise for improving learning outcomes, but lifestyle programs, although potentially effective in reaching their primary goals, have not proved as effective at improving learning outcomes, at least in the short term.

This endeavor highlights several critical areas for future research. First, there is a strong need for meta-analyses with comparable designs to ours that look at different sets of outcomes. As discussed above, meta-analyses incorporating multiple outcome types—including parental behavior and children's social emotional skills—for the full range of EC parenting programs tend to get bogged down in too many moving parts. Yet the interactions across such domains is central to child development, as robust bodies of research on cross-productivity have shown. Systematically charting the dynamics of parenting programs with regard to separate domains in the short term represents an important intermediate step, upon which future analyses build that examine the interactions across domains.

Second, while home visiting and group delivery models have both proven effective, more studies are needed of clinic-based one-on-one and ICT models, as well as studies that blend and/or compare these across different randomized treatment arms. At first glance, one-on-one clinic based programs would not appear to hold much promise over home visiting programs, since they would seem to simply shift transportation costs onto parents while lacking potential benefits that may come from having a group. Nonetheless, this model can work quite efficiently with strategic planning, for instance holding parenting sessions at a clinic that a parent is taking their child to anyway for well-child visits. Although there are too few studies in our sample to calculate a reliable standard error, the pooled effect size from these studies is large. Turning to ICT models, these can cut costs drastically, and this relative advantage is likely to improve as mobile phones and internet become increasingly widespread and widely used for an ever-growing range of tasks. Given substantial success in fields like agricultural extension and adult education (Aker et al.) it is surprising that virtually no ICT-based EC parenting programs in LMICs and only a handful in HICs have been randomly evaluated (at least that report learning outcomes).

Third, notwithstanding a few cohesive strands like the evaluations of NFP and Reach Up, the experimental literature on EC parenting programs to date has been somewhat disjointed. The focus has been on testing whether a particular model works in the study context, rather than building up an empirically-grounded theory of how the effects of different program characteristics vary across contexts. More research is needed that systematically compares program variations within a particular population, or compares a common program model across different populations. While existing studies occasionally report heterogeneity, intentional research design considerations will be needed to systematically compare results across countries, regions, or even different social groups living in the same place. These additional layers will be needed to gain an understanding of the scope conditions in patterns identified. The potentially transformative promise of EC parenting programs has already been well-established; the present frontier of research, theory, and policy analysis is the elaboration of insights on how best to adapt these programs for scaleup.

References

*Abessa, Teklu Gemechu, Berhanu Nigussie Worku, Mekitie Wondafrash, Tsinuel Girma, Johan Valy, Johan Lemmens, Liesbeth Bruckers, Patrick Kolsteren, and Marita Granitzer. 2019. “Effect of Play-Based Family-Centered Psychomotor/Psychosocial Stimulation on the Development of Severely Acutely Malnourished Children under Six in a Low-Income Setting: A Randomized Controlled Trial.” *BMC Pediatrics* 19 (1): 336. https://doi.org/10.1186/s12887-019-1696-z.

*Aboud, Frances E., and Sadika Akhter. 2011. “A Cluster-Randomized Evaluation of a Responsive Stimulation and Feeding Intervention in Bangladesh.” *Pediatrics* 127 (5): e1191-1197. https://doi.org/10.1542/peds.2010-2160

*Aboud, Frances E., Daisy R. Singla, Md Imam Nahil, and Ivelina Borisova. 2013. “Effectiveness of a Parenting Program in Bangladesh to Address Early Childhood Health, Growth and Development.” *Social Science & Medicine (1982)* 97 (November):250–58. https://doi.org/10.1016/j.socscimed.2013.06.020.

Aboud, Frances E., and Aisha K. Yousafzai. 2015. “Global Health and Development in Early Childhood.” *Annual Review of Psychology* 66 (January):433–57. https://doi.org/10.1146/annurev-psych-010814-015128.

Ahun, Marilyn N., Nazia Binte Ali, Elizabeth Hentschel, Joshua Jeong, Emily Franchett, and Aisha K. Yousafzai. 2024. “A Meta-Analytic Review of the Implementation Characteristics in Parenting Interventions to Promote Early Child Development.” *Annals of the New York Academy of Sciences* 1533 (1): 99–144. https://doi.org/10.1111/nyas.15110.

Aker, Jenny C., and Christopher Ksoll. 2019. “Call Me Educated: Evidence from a Mobile Phone Experiment in Niger☆.” *Economics of Education Review* 72 (October):239–57. https://doi.org/10.1016/j.econedurev.2019.05.001.

Al-Ubaydli, Omar, John A. List, and Dana Suskind. 2019. “The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments.” Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w25848.

*Andrew, Alison, Orazio Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina. 2020. “Effects of a Scalable Home-Visiting Intervention on Child Development in Slums of Urban India: Evidence from a Randomised Controlled Trial.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 61 (6): 644–52. https://doi.org/10.1111/jcpp.13171.

Andrew, Alison, Orazio Attanasio, Emla Fitzsimons, Sally Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2018. "Impacts 2 Years after a Scalable Early Childhood Development Intervention to Increase Psychosocial Stimulation in the Home: A Follow-up of a Cluster Randomised Controlled Trial in Colombia." *PLOS Medicine* 15 (4): e1002556. https://doi.org/10.1371/journal.pmed.1002556.

*Ara, Gulshan, Mansura Khanam, Nowshin Papri, Baitun Nahar, Iqbal Kabir, Kazi Istiaque Sanin, Sihan Sadat Khan, Md Shafiqul Alam Sarker, and Michael J Dibley. 2019. "Peer Counseling Promotes Appropriate Infant Feeding Practices and Improves Infant Growth and Development in an Urban Slum in Bangladesh: A Community-Based Cluster Randomized Controlled Trial." *Current Developments in Nutrition* 3 (7): nzz072. https://doi.org/10.1093/cdn/nzz072.

Araujo, M. Caridad, Marta Dormal, Sally Grantham-McGregor, Fabiola Lazarte, Marta Rubio-Codina, and Norbert Schady. 2021A. "Home Visiting at Scale and Child Development." *Journal of Public Economics Plus* 2 (January):100003. https://doi.org/10.1016/j.pubecp.2021.100003.

Araujo, María Caridad, Marta Rubio-Codina, and Norbert Schady. 2021B. "70 to 700 to 70,000: Lessons from the Jamaica Experiment." *IDB Publications*, April. https://doi.org/10.18235/0003210.

*Astuto, Jennifer, and Larue Allen. 2018. "Learning More about Home Visitation: RCT Evaluation of the Parent Child+ Program for Latino Spanish Speaking Children of Immigrants."

*Attanasio, Orazio, Helen Baker-Henningham, Raquel Bernal, Costas Meghir, Diana Pineda, and Marta Rubio-Codina. 2018. "Early Stimulation and Nutrition: The Impacts of a Scalable Intervention." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w25059.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." https://doi.org/10.3386/w25059.

Attanasio, Orazio, Sarah Cattan, and Costas Meghir. 2022. "Early Childhood Development, Human Capital, and Poverty." *Annual Review of Economics* 14 (Volume 14, 2022): 853–92. https://doi.org/10.1146/annurev-economics-092821-053234.

*Attanasio, Orazio, Camila Fernández, Emla Fitzsimons, Sally Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2014. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial." *BMJ: British Medical Journal* 349. https://www.jstor.org/stable/26517341.

Bandura, Albert. 1977. *Social Learning Theory.* Social Learning Theory. Oxford, England: Prentice-Hall.

*Bierman, Karen L., Janet A. Welsh, Brenda S. Heinrichs, Robert L. Nix, and Erin T. Mathis. 2015. "Helping Head Start Parents Promote Their Children's Kindergarten Adjustment: The Research-Based Developmentally Informed Parent Program." *Child Development* 86 (6): 1877–91. https://doi.org/10.1111/cdev.12448.

*Black, Maureen, Howard Dubowitz, Jacqueline Hutcheson, Julie Berenson-Howard, and Raymond Starr. 1995. "A Randomized Clinical Trial of Home Intervention for Children with Failure to Thrive." *Pediatrics* 95 (6): 807–14.

Black, Maureen, Howard Dubowitz, Ambika Krishnakumar, and Raymond Starr. 2007. "Early Intervention and Recovery among Children with Failure to Thrive: Follow-up at Age 8." *Pediatrics* 120 (1): 59–69. https://doi.org/10.1542/peds.2006-1657.

Borenstein, Michael, Julian P. T. Higgins, Larry V. Hedges, and Hannah R. Rothstein. 2017. "Basics of Meta-Analysis: I2 Is Not an Absolute Measure of Heterogeneity." *Research Synthesis Methods* 8 (1): 5–18. https://doi.org/10.1002/jrsm.1230.

*Bouguen, Adrien, Deon Filmer, Karen Macours, and Sophie Naudeau. 2013. "Impact Evaluation of Three Types of Early Childhood Development Interventions in Cambodia." *Policy Research Working Paper Series*, Policy Research Working Paper Series, , July. https://ideas.repec.org//p/wbk/wbrwps/6540.html.

Bowlby, John. 1988. *A Secure Base: Parent-Child Attachment and Healthy Human Development*. Basic Books.

Brennan, Lauretta M., Elizabeth C. Shelleby, Daniel S. Shaw, Frances Gardner, Thomas J. Dishion, and Melvin Wilson. 2013. "Indirect Effects of the Family Check-Up on School-Age Academic Achievement Through Improvements in Parenting in Early Childhood." *Journal of Educational Psychology* 105 (3). https://doi.org/10.1037/a0032096.

*Brentani, Alexandra, Susan Walker, Susan Chang-Lopez, Sandra Grisi, Christine Powell, and Günther Fink. 2021. "A Home Visit-Based Early Childhood Stimulation Programme in Brazil-a Randomized Controlled Trial." *Health Policy and Planning* 36 (3): 288–97. https://doi.org/10.1093/heapol/czaa195.

Britto, Pia, L. Angelicaa Ponguta, Chin Reyes, and Romilla Karnati. 2015. "A Systematic Review of Parenting Programmes for Young Children in Low- and Middle-Income Countries." UNICEF.

Britto, Pia R., Stephen J. Lye, Kerrie Proulx, Aisha K. Yousafzai, Stephen G. Matthews, Tyler Vaivada, Rafael Perez-Escamilla, et al. 2017. "Nurturing Care: Promoting Early Childhood Development." *Lancet (London, England)* 389 (10064): 91–102. https://doi.org/10.1016/S0140-6736(16)31390-3.

Bronfenbrenner, Urie. 1977. “Toward an Experimental Ecology of Human Development.” *American Psychologist* 32 (7): 513–31. https://doi.org/10.1037/0003-066X.32.7.513.

*Caldera, Debra, Lori Burrell, Kira Rodriguez, Sarah Shea Crowne, Charles Rohde, and Anne Duggan. 2007. “Impact of a Statewide Home Visiting Program on Parenting and on Child Health and Development.” *Child Abuse & Neglect* 31 (8): 829–52. https://doi.org/10.1016/j.chiabu.2007.02.008.

Carneiro, Pedro, Emanuela Galasso, Italo Lopez Garcia, Paula Bedregal, and Miguel Cordero. 2024. “Impacts of a Large-Scale Parenting Program: Experimental Evidence from Chile.” *Journal of Political Economy* 132 (4): 1113–61. https://doi.org/10.1086/727288.

*Chacko, Anil, Gregory A. Fabiano, Greta L. Doctoroff, and Beverly Fortson. 2018. “Engaging Fathers in Effective Parenting for Preschool Children Using Shared Book Reading: A Randomized Controlled Trial.” *Journal of Clinical Child & Adolescent Psychology* 47 (1): 79–93. https://doi.org/10.1080/15374416.2016.1266648.

*Chamberlain, Lisa J., Janine Bruce, Monica De La Cruz, Lynne Huffman, Jecca Rhea Steinberg, Ren Bruguera, Jaime W. Peterson, et al. 2021. “A Text-Based Intervention to Promote Literacy: An RCT.” *Pediatrics* 148 (4): e2020049648. https://doi.org/10.1542/peds.2020-049648.

*Chang, Susan M., Sally M. Grantham-McGregor, Christine A. Powell, Marcos Vera-Hernández, Florencia Lopez-Boo, Helen Baker-Henningham, and Susan P. Walker. 2015. “Integrating a Parenting Intervention With Routine Primary Health Care: A Cluster Randomized Trial.” *Pediatrics* 136 (2): 272–80. https://doi.org/10.1542/peds.2015-0119.

*Chinen, Marjorie, and Johannes Bos. 2016. “Impact Evaluation of the Save the Children Early Childhood Stimulation Program in Bangladesh: Final Report.”

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. “Estimating the Technology of Cognitive and Noncognitive Skill Formation.” *Econometrica* 78 (3): 883–931.

Dietrichson, Jens, Martin Bøg, Trine Filges, and Anne-Marie Klint Jørgensen. 2017. “Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis.” *Review of Educational Research* 87 (2): 243–82. https://doi.org/10.3102/0034654316687036.

*Doss, Christopher, Hans Fricke, Susanna Loeb, and Justin B. Doromal. 2022. “Engaging Girls in Math: The Unequal Effects of Text Messaging to Help Parents Support Early Math Development.” *Economics of Education Review* 88 (June):102262. https://doi.org/10.1016/j.econedurev.2022.102262.

*Doyle, Orla. 2020A. "Can Early Intervention Have a Sustained Effect on Human Capital?" Working Papers 202001. Geary Institute, University College Dublin. https://EconPapers.repec.org/RePEc:ucd:wpaper:202001.

———. 2020B. "The First 2,000 Days and Child Skills." *Journal of Political Economy* 128 (6): 2067–2122. https://doi.org/10.1086/705707.

*Drotar, D., J. Robinson, L. Jeavons, and H. Lester Kirchner. 2008. "A Randomized, Controlled Evaluation of Early Intervention: The Born to Learn Curriculum." *Child: Care, Health and Development* 35 (5): 643–49. https://doi.org/10.1111/j.1365-2214.2008.00915.x.

*Duggan, Anne K., Elizabeth C. McFarlane, Amy M. Windham, C. A. Rohde, David S. Salkever, Loretta Fuddy, Leon A. Rosenberg, Sharon B. Buchbinder, and Calvin C. J. Sia. 1999. "Evaluation of Hawaii's Healthy Start Program." *The Future of Children* 9 (1): 66–90; discussion 177-178.

Eric C. Hedberg. 2011. "ROBUMETA: Stata Module to Perform Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates." Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s457219.html.

*Fernald, Lia C. H., Rose M. C. Kagawa, Heather A. Knauer, Lourdes Schnaas, Armando Garcia Guerra, and Lynnette M. Neufeld. 2017. "Promoting Child Development through Group-Based Parent Support within a Cash Transfer Program: Experimental Effects on Children's Outcomes." *Developmental Psychology* 53 (2): 222–36. https://doi.org/10.1037/dev0000185.

Filene, Jill H., Jennifer W. Kaminski, Linda Anne Valle, and Patrice Cachat. 2013. "Components Associated with Home Visiting Program Outcomes: A Meta-Analysis." *Pediatrics* 132 Suppl 2 (0 2): S100-109. https://doi.org/10.1542/peds.2013-1021H.

*Francis, Taja, and Helen Baker-Henningham. 2021. "The Irie Homes Toolbox: A Cluster Randomized Controlled Trial of an Early Childhood Parenting Program to Prevent Violence against Children in Jamaica." *Children and Youth Services Review* 126 (July):106060. https://doi.org/10.1016/j.childyouth.2021.106060.

Furlong, Mairead, Sinead McGilloway, Tracey Bywater, Judy Hutchings, Susan M. Smith, and Michael Donnelly. 2012. "Behavioural and Cognitive-Behavioural Group-Based Parenting Programmes for Early-Onset Conduct Problems in Children Aged 3 to 12 Years." *Campbell Systematic Reviews* 8 (1): 1–239. https://doi.org/10.4073/csr.2012.12.

Garcia, Italo Lopez, Uzaib Y. Saya, and Jill E. Luoto. 2021. "Cost-Effectiveness and Economic Returns of Group-Based Parenting Interventions to Promote Early

Childhood Development: Results from a Randomized Controlled Trial in Rural Kenya." *PLOS Medicine* 18 (9): e1003746. https://doi.org/10.1371/journal.pmed.1003746.
García, Jorge Luis, and James J. Heckman. 2023. "Parenting Promotes Social Mobility Within and Across Generations." *Annual Review of Economics* 15 (Volume 15, 2023): 349–88. https://doi.org/10.1146/annurev-economics-021423-031905.
*Gardner, Julie M. Meeks, Christine A. Powell, Helen Baker-Henningham, Susan P. Walker, Tim J. Cole, and Sally M. Grantham-McGregor. 2005. "Zinc Supplementation and Psychosocial Stimulation: Effects on the Development of Undernourished Jamaican Children." The American Journal of Clinical Nutrition 82 (2): 399–405. https://doi.org/10.1093/ajcn.82.2.399.
*Gardner, Julie Meeks, Susan P. Walker, Christine A. Powell, and Sally Grantham-McGregor. 2003. "A Randomized Controlled Trial of a Home-Visiting Intervention on Cognition and Behavior in Term Low Birth Weight Infants." The Journal of Pediatrics 143 (5): 634–39. https://doi.org/10.1067/S0022-3476(03)00455-4.
Gennetian, Lisa A. 2021. "How a Behavioral Economic Framework Can Support Scaling of Early Childhood Interventions." In *The Scale-Up Effect in Early Childhood and Public Policy*. Taylor & Francis. https://doi.org/10.4324/9780367822972-4.
Gertler, Paul, James J. Heckman, Rodrigo Pinto, Susan M. Chang, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright. 2021. "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w29292.
Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeerch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science (New York, N.Y.)* 344 (6187): 998–1001. https://doi.org/10.1126/science.1251178.
*Goldfeld, Sharon, Natasha Napiza, Jon Quach, Sheena Reilly, Obioha Ukoumunne, and Melissa Wake. 2011. "Outcomes of a Universal Shared Reading Intervention by 2 Years of Age: The Let's Read Trial." *Pediatrics* 127 (February):445–53. https://doi.org/10.1542/peds.2009-3043.
*Grantham-McGregor, Sally., Christine Powell, Susan Walker, and John Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *Lancet (London, England)* 338 (8758): 1–5. https://doi.org/10.1016/0140-6736(91)90001-6.
*Grantham-McGregor, Sally, Akanksha Adya, Orazio Attanasio, Britta Augsburg, Jere Behrman, Bet Caeyers, Monimalika Day, et al. 2020. "Group Sessions or Home

Visits for Early Childhood Development in India: A Cluster RCT." *Pediatrics* 146 (6): e2020002725. https://doi.org/10.1542/peds.2020-002725.
Grantham-McGregor, Sally, and Joanne Smith. 2016. "Extending The Jamaican Early Childhood Development Intervention." *Journal of Applied Research on Children: Informing Policy for Children at Risk* 7 (2). https://doi.org/10.58464/2155-5834.1306.
Grantham-McGregor, Sally, Susan Walker, Susan Chang, and Christine Powell. 1997. "Effects of Early Childhood Supplementation with and without Stimulation on Later Development in Stunted Jamaican Children." *The American Journal of Clinical Nutrition* 66 (2): 247–53. https://doi.org/10.1093/ajcn/66.2.247.
Gupta, Snigdha, Lauren H. Supplee, Dana Suskind, and John A. List. 2021. "Failed to Scale: Embracing the Challenge of Scaling in Early Childhood." In *The Scale-Up Effect in Early Childhood and Public Policy*. Routledge.
*Guttentag, Cathy L., Susan H. Landry, Jeffrey M. Williams, Kathleen M. Baggett, Christine W. Noria, John G. Borkowski, Paul R. Swank, et al. 2014. "'My Baby & Me': Effects of an Early, Comprehensive Parenting Intervention on At-Risk Mothers and Their Children." *Developmental Psychology* 50 (5): 1482–96. https://doi.org/10.1037/a0035682.
*Hamadani, Jena D., Syed N. Huda, Fahmida Khatun, and Sally M. Grantham-McGregor. 2006. "Psychosocial Stimulation Improves the Development of Undernourished Children in Rural Bangladesh." *The Journal of Nutrition* 136 (10): 2645–52. https://doi.org/10.1093/jn/136.10.2645.
*Hamadani, Jena D, Syeda F Mehrin, Fahmida Tofail, Mohammad I Hasan, Syed N Huda, Helen Baker-Henningham, Deborah Ridout, and Sally Grantham-McGregor. 2019. "Integrating an Early Childhood Development Programme into Bangladeshi Primary Health-Care Services: An Open-Label, Cluster-Randomised Controlled Trial." *The Lancet Global Health* 7 (3): e366–75. https://doi.org/10.1016/S2214-109X(18)30535-7.
Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900–1902. https://doi.org/10.1126/science.1128898.
Heckman, James J., Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda. 2017. "An Analysis of the Memphis Nurse-Family Partnership Program." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w23610.
*Heckman, James J., Bei Liu, Mai Lu, and Jin Zhou. 2020. "The Impacts of a Prototypical Home Visiting Program on Child Skills." Working Paper. Working

Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w27356.

*Heinicke, Christopher M., N.r. Fineman, G. Ruth, S.l. Recchia, D. Guthrie, and C. Rodning. 1999. "Relationship-Based Intervention with at-Risk Mothers: Outcome in the First Year of Life." *Infant Mental Health Journal* 20 (4): 349–74. https://doi.org/10.1002/(SICI)1097-0355(199924)20:4<349::AID-IMHJ1>3.0.CO;2-X.

*Helmizar, Helmizar, Fasli Jalal, Nur Indrawati Lipoeto, and Endang L. Achadi. 2017. "Local Food Supplementation and Psychosocial Stimulation Improve Linear Growth and Cognitive Development among Indonesian Infants Aged 6 to 9 Months." *Asia Pacific Journal of Clinical Nutrition* 26 (1): 97–103. https://doi.org/10.6133/apjcn.102015.10.

*Hilado, Aimee, Christine Leow, and Yinmei Yang. 2018. "Baby TALK- RefugeeOne Study: A Randomized Controlled Trial Examining the Impact of Home Visiting Services with Refugees and Immigrants." https://babytalk.org/family-engagement-model/research/baby-talk-refugeeone-randomized-controlled-trial-a-study-examining-the-impact-of-home-visitation-among-refugee-and-immigrant-families/.

*Holtzman, Deborah J., Heather E. Quick, and Shannon Keuter. 2023. "Math for 2s and 3s: The Impact of Parent-Child Math Activities on Parents' Beliefs and Behaviors and Young Children's Math Skill Development." *Early Childhood Research Quarterly* 62 (January):163–74. https://doi.org/10.1016/j.ecresq.2022.07.015.

*Huebner, Colleen E. 2000. "Promoting Toddlers' Language Development Through Community-Based Intervention." *Journal of Applied Developmental Psychology* 21 (5): 513–35. https://doi.org/10.1016/S0193-3973(00)00052-6.

*Hutchings, J., N. Griffith, T. Bywater, and M. E. Williams. 2017. "Evaluating the Incredible Years Toddler Parenting Programme with Parents of Toddlers in Disadvantaged (Flying Start) Areas of Wales." *Child: Care, Health and Development* 43 (1): 104–13. https://doi.org/10.1111/cch.12415.

*Imhof, Andrea, Sihong Liu, Lisa Schlueter, Tiffany Phu, Sarah Watamura, and Philip Fisher. 2023. "Improving Children's Expressive Language and Auditory Comprehension Through Responsive Caregiving: Evidence from a Randomized Controlled Trial of a Strength-Based Video-Coaching Intervention." *Prevention Science* 24 (1): 84–93. https://doi.org/10.1007/s11121-022-01455-4.

*Jeong, Joshua, Marilyn N. Ahun, Nilupa S. Gunaratna, Ramya Ambikapathi, Frank Mapendo, Lauren Galvin, Mary Pat Kieffer, et al. 2023. "Effects of Engaging Fathers and Bundling Parenting and Nutrition Interventions on Early Child Development and Maternal and Paternal Parenting in Mara, Tanzania: A Factorial

Cluster-Randomized Controlled Trial." *Journal of Child Psychology and Psychiatry* 65 (5): 694–709. https://doi.org/10.1111/jcpp.13897.

Jeong, Joshua, Emily E. Franchett, Clariana V. Ramos de Oliveira, Karima Rehmani, and Aisha K. Yousafzai. 2021A. "Parenting Interventions to Promote Early Child Development in the First Three Years of Life: A Global Systematic Review and Meta-Analysis." *PLOS Medicine* 18 (5): e1003602. https://doi.org/10.1371/journal.pmed.1003602.

Jeong, Joshua, Helen O. Pitchik, and Günther Fink. 2021B. "Short-Term, Medium-Term and Long-Term Effects of Early Parenting Interventions in Low- and Middle-Income Countries: A Systematic Review." *BMJ Global Health* 6 (3): e004067. https://doi.org/10.1136/bmjgh-2020-004067.

*Jin, Xingming, Yajuan Sun, Fan Jiang, Jun Ma, Christopher Morgan, and Xiaoming Shen. 2007. "'Care for Development' Intervention in Rural China: A Prospective Follow-up Study." *Journal of Developmental and Behavioral Pediatrics: JDBP* 28 (3): 213–18. https://doi.org/10.1097/dbp.0b013e31802d410b.

*Jungmann, Tanja, Tilman Brand, Verena Dähne, Peggy Herrmann, Hüsamettin Günay, Malte Sandner, and Susan Sierau. 2015. "Comprehensive Evaluation of the Pro Kind Home Visiting Program: A Summary of Results." *Mental Health & Prevention* 3 (3): 89–97. https://doi.org/10.1016/j.mhp.2015.06.001.

*Kaaresen, Per Ivar, John A. Rønning, Jorunn Tunby, Solveig Marianne Nordhov, Stein Erik Ulvund, and Lauritz B. Dahl. 2008. "A Randomized Controlled Trial of an Early Intervention Program in Low Birth Weight Children: Outcome at 2 Years." *Early Human Development* 84 (3): 201–9. https://doi.org/10.1016/j.earlhumdev.2007.07.003.

*Katz, Kathy S., Marian H. Jarrett, Ayman A. E. El-Mohandes, Susan Schneider, Doris McNeely-Johnson, and Michele Kiely. 2011. "Effectiveness of a Combined Home Visiting and Group Intervention for Low Income African American Mothers: The Pride in Parenting Program." *Maternal and Child Health Journal* 15 (1): 75–84. https://doi.org/10.1007/s10995-011-0858-x.

*Kemp, Lynn, Elizabeth Harris, Catherine McMahon, Stephen Matthey, Graham Vimpani, Teresa Anderson, Virginia Schmied, Henna Aslam, and Siggi Zapart. 2011. "Child and Family Outcomes of a Long-Term Nurse Home Visitation Programme: A Randomised Controlled Trial." *Archives of Disease in Childhood* 96 (6): 533–40. https://doi.org/10.1136/adc.2010.196279.

*Kim, YaeBin, and Dave Riley. 2021. "Accelerating Early Language and Literacy Skills Through a Preschool-Home Partnership Using Dialogic Reading: A Randomized

Trial." *Child & Youth Care Forum* 50 (5): 901–24. https://doi.org/10.1007/s10566-021-09598-1.

*Kitzman, Harriet, David Olds, Charles Henderson, C. Hanks, Robert Cole, Robert Tatelbaum, Kenneth McConnochie, et al. 1997. "Effect of Prenatal and Infancy Home Visitation by Nurses on Pregnancy Outcomes, Childhood Injuries, and Repeated Childbearing. A Randomized Controlled Trial." *JAMA* 278 (8): 644–52.

Kliem, Sören, and Malte Sandner. 2021. "Prenatal and Infancy Home Visiting in Germany: 7-Year Outcomes of a Randomized Trial." *Pediatrics* 148 (2): e2020049610. https://doi.org/10.1542/peds.2020-049610.

*Knight, David S., Susan Landry, Tricia A. Zucker, Emily C. Merz, Cathy L. Guttentag, and Heather B. Taylor. 2019. "Cost-Effectiveness of Early Childhood Interventions to Enhance Preschool: Evidence from a Randomized Experiment in Head Start Centers Enrolling Historically Underserved Populations." *Journal of Policy Analysis and Management* 38 (4): 891–917. https://doi.org/10.1002/pam.22145.

*Kynø, Nina M., Ingrid Helen Ravn, Rolf Lindemann, Morten W. Fagerland, Nina Aarhus Smeby, and Anne Mari Torgersen. 2012. "Effect of an Early Intervention Programme on Development of Moderate and Late Preterm Infants at 36 Months: A Randomized Controlled Study." *Infant Behavior & Development* 35 (4): 916–26. https://doi.org/10.1016/j.infbeh.2012.09.004.

*Landry, Susan H., Tricia A. Zucker, Janelle J. Montroy, Hsien-Yuan Hsu, Mike A. Assel, Cheryl Varghese, April Crawford, and Edward G. Feil. 2021. "Replication of Combined School Readiness Interventions for Teachers and Parents of Head Start Pre-Kindergarteners Using Remote Delivery." *Early Childhood Research Quarterly* 56 (July):149–66. https://doi.org/10.1016/j.ecresq.2021.03.007.

*Landry, Susan H., Tricia A. Zucker, Jeffrey M. Williams, Emily C. Merz, Cathy L. Guttentag, and Heather B. Taylor. 2017. "Improving School Readiness of High-Risk Preschoolers: Combining High Quality Instructional Strategies with Responsive Training for Teachers and Parents." *Early Childhood Research Quarterly* 40 (July):38–51. https://doi.org/10.1016/j.ecresq.2016.12.001.

Leijten, Patty, Frances Gardner, G. J. Melendez-Torres, Jolien van Aar, Judy Hutchings, Susanne Schulz, Wendy Knerr, and Geertjan Overbeek. 2019. "Meta-Analyses: Key Parenting Program Components for Disruptive Child Behavior." *Journal of the American Academy of Child and Adolescent Psychiatry* 58 (2): 180–90. https://doi.org/10.1016/j.jaac.2018.07.900.

*Leung, Cynthia, Sandra Tsang, and Kitty Heung. 2015. "The Effectiveness of Healthy Start Home Visit Program: Cluster Randomized Controlled Trial." *Research on Social Work Practice* 25 (3): 322–33. https://doi.org/10.1177/1049731514533390.

*Leung, Cynthia, Sandra Tsang, and H. W. Kwan. 2017. "Efficacy of a Universal Parent Training Program (HOPE-20): Cluster Randomized Controlled Trial." *Research on Social Work Practice* 27 (5): 523–37. https://doi.org/10.1177/1049731515593810.

Levey, Elizabeth J, Bizu Gelaye, Paul Bain, Marta B Rondon, Christina PC Borba, David C Henderson, and Michelle A Williams. 2017. "A Systematic Review of Randomized Controlled Trials of Interventions Designed to Decrease Child Abuse in High-Risk Families." *Child Abuse & Neglect* 65:48–57.

List, John, Dana Suskind, and Lauren Supplee. 2021. *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. Routledge.

*Lunkenheimer, Erika S., Thomas J. Dishion, Daniel S. Shaw, Arin M. Connell, Frances Gardner, Melvin N. Wilson, and Emily M. Skuban. 2008. "Collateral Benefits of the Family Check-Up on Early Childhood School Readiness: Indirect Effects of Parents' Positive Behavior Support." *Developmental Psychology* 44 (6): 1737–52. https://doi.org/10.1037/a0013858.

Luo, Renfu, Dorien Emmers, Nele Warrinnier, Scott Rozelle, and Sean Sylvia. 2019. "Using Community Health Workers to Deliver a Scalable Integrated Parenting Program in Rural China: A Cluster-Randomized Controlled Trial." *Social Science & Medicine* 239 (October):112545. https://doi.org/10.1016/j.socscimed.2019.112545.

*Luoto, Jill E, Italo Lopez Garcia, Frances E Aboud, Daisy R Singla, Lia C H Fernald, Helen O Pitchik, Uzaib Y Saya, Ronald Otieno, and Edith Alu. 2021. "Group-Based Parenting Interventions to Promote Child Development in Rural Kenya: A Multi-Arm, Cluster-Randomised Community Effectiveness Trial." *The Lancet Global Health* 9 (3): e309–19. https://doi.org/10.1016/S2214-109X(20)30469-1.

*McGillion, Michelle, Julian M. Pine, Jane S. Herbert, and Danielle Matthews. 2017. "A Randomised Controlled Trial to Test the Effect of Promoting Caregiver Contingent Talk on Language Development in Infants from Diverse Socioeconomic Status Backgrounds." *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 58 (10): 1122–31. https://doi.org/10.1111/jcpp.12725.

*Mendelsohn, Alan L., Purnima T. Valdez, Virginia Flynn, Gilbert M. Foley, Samantha B. Berkule, Suzy Tomopoulos, Arthur H. Fierman, Wendy Tineo, and Benard P. Dreyer. 2007. "Use of Videotaped Interactions During Pediatric Well-Child Care: Impact at 33 Months on Parenting and on Child Development." *Journal of Developmental and Behavioral Pediatrics : JDBP* 28 (3): 206–12. https://doi.org/10.1097/DBP.0b013e3180324d87.

*Muhoozi, Grace K. M., Prudence Atukunda, Lien M. Diep, Robert Mwadime, Archileo N. Kaaya, Anne B. Skaare, Tiril Willumsen, Ane C. Westerberg, and Per O. Iversen.

2017. “Nutrition, Hygiene, and Stimulation Education to Improve Growth, Cognitive, Language, and Motor Development among Infants in Uganda: A Cluster-Randomized Trial.” *Maternal & Child Nutrition* 14 (2): e12527. https://doi.org/10.1111/mcn.12527.

*Murray, Lynne, Peter Cooper, Adriane Arteche, Alan Stein, and Mark Tomlinson. 2015. “Randomized Controlled Trial of a Home-Visiting Intervention on Infant Cognitive Development in Peri-Urban South Africa.” *Developmental Medicine & Child Neurology* 58 (3): 270–76. https://doi.org/10.1111/dmcn.12873.

*Murray, Lynne, Susie Jennings, Hayley Perry, Martin Andrews, Katherine De Wilde, Amber Newell, Alicia Mortimer, et al. 2023. “Effects of Training Parents in Dialogic Book-sharing: The Early-Years Provision in Children’s Centers (EPICC) Study.” *Early Childhood Research Quarterly* 62 (January):1–16. https://doi.org/10.1016/j.ecresq.2022.07.008.

*Nahar, Baitun, Iqbal Hossain, J. D. Hamadani, Tan Ahmed, S. N. Huda, Sally Grantham-McGregor, and Lars-Åke Persson. 2012. “Effects of a Community-Based Approach of Food and Psychosocial Stimulation on Growth and Development of Severely Malnourished Children in Bangladesh: A Randomised Trial.” *European Journal of Clinical Nutrition* 66 (6): 701–9. https://doi.org/10.1038/ejcn.2012.13.

*Nair, M. K. C., Elsie Philip, L. Jeyaseelan, Babu George, Suja Mathews, and K. Padma. 2009. “Effect of Child Development Centre Model Early Stimulation among at Risk Babies--a Randomized Controlled Trial.” *Indian Pediatrics* 46 Suppl (January):s20-26.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2024. “The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence.” *American Educational Research Journal* 61 (1): 74–107. https://doi.org/10.3102/00028312231208687.

*Norr, Kathleen F., Kathleen S. Crittenden, Evelyn L. Lehrer, Olga Reyes, Cynthia B. Boyd, Karla W. Nacion, and Kaoru Watanabe. 2003. “Maternal and Infant Outcomes at One Year for a Nurse-Health Advocate Home Visiting Program Serving African Americans and Mexican Americans.” *Public Health Nursing (Boston, Mass.)* 20 (3): 190–203. https://doi.org/10.1046/j.0737-1209.2003.20306.x.

*Olds, David, Charles Henderson, Robert Chamberlin, and Robert Tatelbaum. 1986. “Preventing Child Abuse and Neglect: A Randomized Trial of Nurse Home Visitation.” *Pediatrics* 78 (1): 65–78.

Olds, David L., Peggy L. Hill, Ruth O’Brien, David Racine, and Pat Moritz. 2003. “Taking Preventive Intervention to Scale: The Nurse-Family Partnership.” *Cognitive* *and Behavioral Practice* 10 (4): 278–90. https://doi.org/10.1016/S1077-7229(03)80046-9.

Olds, David L., John R. Holmberg, Nancy Donelan-McCall, Dennis W. Luckey, Michael D. Knudtson, and Joann Robinson. 2014. "Effects of Home Visits by Paraprofessionals and by Nurses on Children: Follow-up of a Randomized Trial at Ages 6 and 9 Years." *JAMA Pediatrics* 168 (2): 114–21. https://doi.org/10.1001/jamapediatrics.2013.3817.

Olds, David L., Harriet Kitzman, Robert Cole, JoAnn Robinson, Kimberly Sidora, Dennis W. Luckey, Charles R. Henderson, Carole Hanks, Jessica Bondy, and John Holmberg. 2004. "Effects of Nurse Home-Visiting on Maternal Life Course and Child Development: Age 6 Follow-up Results of a Randomized Trial." *Pediatrics* 114 (6): 1550–59. https://doi.org/10.1542/peds.2004-0962.

*Olds, David L., JoAnn Robinson, Ruth O'Brien, Dennis W. Luckey, Lisa M. Pettitt, Charles R. Henderson, Rosanna K. Ng, et al. 2002. "Home Visiting by Paraprofessionals and by Nurses: A Randomized, Controlled Trial." *Pediatrics* 110 (3): 486–96. https://doi.org/10.1542/peds.110.3.486.

Olds, David L., JoAnn Robinson, Lisa Pettitt, Dennis W. Luckey, John Holmberg, Rosanna K. Ng, Kathy Isacks, Karen Sheff, and Charles R. Henderson. 2004. "Effects of Home Visits by Paraprofessionals and by Nurses: Age 4 Follow-up Results of a Randomized Trial." *Pediatrics* 114 (6): 1560–68. https://doi.org/10.1542/peds.2004-0961.

*Perou, Ruth, Lara R. Robinson, Melissa L. Danielson, Angelika H. Claussen, Susanna N. Visser, Keith G. Scott, Leila Beckwith, Lynne Katz, and D. Camille Smith. 2019. "The Legacy for Children™ Randomized Control Trial: Effects on Cognition Through Third Grade for Young Children Experiencing Poverty." *Journal of Developmental and Behavioral Pediatrics: JDBP* 40 (4): 275–84. https://doi.org/10.1097/DBP.0000000000000656.

*Piccolo, Luciane R., João Batista Araujo Oliveira, Guilherme Hirata, Walfrido Duarte Neto, and Alan L. Mendelsohn. 2022. "Supporting Reading Aloud Beginning Prenatally and in Early Infancy: A Randomized Trial in Brazil." *Journal of Developmental and Behavioral Pediatrics: JDBP* 43 (9): e590–97. https://doi.org/10.1097/DBP.0000000000001118.

Pigott, Terri D., and Joshua R. Polanin. 2020. "Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review." *Review of Educational Research* 90 (1): 24–46. https://doi.org/10.3102/0034654319877153.

*Potterton, Joanne, Aimee Stewart, Peter Cooper, and Pieter Becker. 2009. "The Effect of a Basic Home Stimulation Programme on the Development of Young Children

Infected with HIV." *Developmental Medicine & Child Neurology* 52 (6): 547–51. https://doi.org/10.1111/j.1469-8749.2009.03534.x.

*Powell, Christine, Helen Baker-Henningham, Susan Walker, Jacqueline Gernay, and Sally Grantham-McGregor. 2004. "Feasibility of Integrating Early Stimulation into Primary Care for Undernourished Jamaican Children: Cluster Randomised Controlled Trial." *BMJ* 329 (7457): 89. https://doi.org/10.1136/bmj.38132.503472.7C.

Pustejovsky, James E., and Melissa A. Rodgers. 2019. "Testing for Funnel Plot Asymmetry of Standardized Mean Differences." *Research Synthesis Methods* 10 (1): 57–71. https://doi.org/10.1002/jrsm.1332.

Rao, Sylvia, Margaret E. Bentley, Nagalla Balakrishna, Paula Griffiths, Hilary Creed-Kanashiro, Shahnaz Vazir, and Susan L. Johnson. 2020. "A Complementary Feeding and Play Intervention Improves the Home Environment and Mental Development among Toddlers in Rural India." *Maternal & Child Nutrition* 16 (S3): e13066. https://doi.org/10.1111/mcn.13066.

*Rauh, Virginia A., Thomas M. Achenbach, Barry Nurcombe, Catherine T. Howell, and Douglas M. Teti. 1988. "Minimizing Adverse Effects of Low Birthweight: Four-Year Results of an Early Intervention Program." *Child Development* 59 (3): 544–53. https://doi.org/10.2307/1130556.

*Roberts, Megan Y., and Ann P. Kaiser. 2015. "Early Intervention for Toddlers with Language Delays: A Randomized Controlled Trial." *Pediatrics* 135 (4): 686–93. https://doi.org/10.1542/peds.2014-2134.

*Robling, Michael, Marie-Jet Bekkers, Kerry Bell, Christopher C Butler, Rebecca Cannings-John, Sue Channon, Belen Corbacho Martin, et al. 2016. "Effectiveness of a Nurse-Led Intensive Home-Visitation Programme for First-Time Teenage Mothers (Building Blocks): A Pragmatic Randomised Controlled Trial." *The Lancet* 387 (10014): 146–55. https://doi.org/10.1016/S0140-6736(15)00392-X.

Robling, Michael, Fiona Lugg-Widger, Rebecca Cannings-John, Julia Sanders, Lianna Angel, Sue Channon, Deborah Fitzsimmons, et al. 2021. "The Family Nurse Partnership to Reduce Maltreatment and Improve Child Health and Development in Young Children: The BB:2 6 Routine Data-Linkage Follow-up to Earlier RCT." *Public Health Research* 9 (2): 1–160. https://doi.org/10.3310/phr09020.

*Rockers, Peter C., Günther Fink, Arianna Zanolini, Bowen Banda, Godfrey Biemba, Cierra Sullivan, Simon Mutembo, Vichaels Silavwe, and Davidson H. Hamer. 2016. "Impact of a Community-Based Package of Interventions on Child Development in Zambia: A Cluster-Randomised Controlled Trial." *BMJ Global Health* 1 (3): e000104. https://doi.org/10.1136/bmjgh-2016-000104.

Rodgers, Melissa A., and James E. Pustejovsky. 2021. “Evaluating Meta-Analytic Methods to Detect Selective Reporting in the Presence of Dependent Effect Sizes.” *Psychological Methods* 26 (2): 141–60. https://doi.org/10.1037/met0000300.

Sandner, Malte, and Tanja Jungmann. 2017. “Gender-Specific Effects of Early Childhood Intervention: Evidence from a Randomized Controlled Trial.” *Labour Economics*, Field experiments in labor economics and social policies, 45 (April):59–78. https://doi.org/10.1016/j.labeco.2016.11.006.

*Scarr, Sandra, and Kathleen McCartney. 1988. “Far from Home: An Experimental Evaluation of the Mother-Child Home Program in Bermuda.” *Child Development* 59 (3): 531–43. https://doi.org/10.2307/1130555.

*Schaub, Simone, Erich Ramseier, Alex Neuhauser, Susan C. A. Burkhardt, and Andrea Lanfranchi. 2019. “Effects of Home-Based Early Intervention on Child Outcomes: A Randomized Controlled Trial of Parents as Teachers in Switzerland.” *Early Childhood Research Quarterly* 48 (July):173–85. https://doi.org/10.1016/j.ecresq.2019.03.007.

*Schwarz, Donald F., Ann L. O’Sullivan, Judith Guinn, Jennifer A. Mautone, Elyse C. Carlson, Huaqing Zhao, Xuemei Zhang, Tara L. Esposito, Megan Askew, and Jerilynn Radcliffe. 2012. “Promoting Early Intervention Referral Through a Randomized Controlled Home-Visiting Program.” *Journal of Early Intervention* 34 (1): 20–39. https://doi.org/10.1177/1053815112451849.

Shonkoff, Jack P., and Deborah A. Phillips, eds. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/9824.

*Singla, Daisy, Elias Kumbakumba, and Frances E. Aboud. 2015. “Effects of a Parenting Intervention to Address Maternal Psychological Wellbeing and Child Development and Growth in Rural Uganda: A Community-Based, Cluster-Randomised Trial.” *The Lancet Global Health* 3 (8): e458–69. https://doi.org/10.1016/S2214-109X(15)00099-6.

Slavin, Robert, and Nancy A. Madden. 2011. “Measures Inherent to Treatments in Program Effectiveness Reviews.” *Journal of Research on Educational Effectiveness* 4 (4): 370–80. https://doi.org/10.1080/19345747.2011.558986.

Smith, Joanne A., Susan M. Chang, Florencia Lopez Boo, Maria de la Paz Ferro, and Susan P. Walker. 2021. “Are Benefits From a Parenting Intervention Delivered Through the Health Services Sustainable? Follow-Up of a Randomized Evaluation in Jamaica.” *Academic Pediatrics* 21 (4): 638–45. https://doi.org/10.1016/j.acap.2021.01.003.

*Smith, Joanne A., Christine A. Powell, Susan M. Chang, Emily Ganga, Hillary Tanyanyiwa, and Susan P. Walker. 2023. "A Cluster Randomised Controlled Trial of an Early Childhood Parenting Programme Delivered through Early Childhood Education Centres in Rural Zimbabwe." *Child: Care, Health and Development* 50 (1): e13189. https://doi.org/10.1111/cch.13189.

Suskind, Dana. 2015. *Thirty Million Words: Building a Child's Brain*. Penguin Publishing Group.

———. 2022. *Parent Nation: Unlocking Every Child's Potential, Fulfilling Society's Promise*. Penguin Publishing Group.

*Sylvia, Sean, Renfu Luo, Jingdong Zhong, Sarah Dill, Alexis Medina, and Scott Rozelle. 2020. "Passive versus Active Service Delivery: Comparing the Effects of Two Parenting Interventions on Early Cognitive Development in Rural China."

*Sylvia, Sean, Renfu Luo, Jingdong Zhong, Sarah-Eve Dill, Alexis Medina, and Scott Rozelle. 2022. "Passive versus Active Service Delivery: Comparing the Effects of Two Parenting Interventions on Early Cognitive Development in Rural China." *World Development* 149 (January):105686. https://doi.org/10.1016/j.worlddev.2021.105686.

Sylvia, Sean, Nele Warrinnier, Luo Renfu, Ai Yue, Orazio P. Attanasio, Alexis Medina, and Scott Rozelle. 2018. "From Quantity to Quality: Delivering a Home-Based Parenting Intervention through China's Family Planning Cadres." Working Paper 402. LICOS Discussion Paper. https://www.econstor.eu/handle/10419/200486.

Tanner-Smith, Emily E., Elizabeth Tipton, and Joshua R. Polanin. 2016. "Handling Complex Meta-Analytic Data Structures Using Robust Variance Estimates: A Tutorial in R." *Journal of Developmental and Life-Course Criminology* 2 (1): 85–112. https://doi.org/10.1007/s40865-016-0026-5.

*Teepe, Rosa Catharina, Inge Molenaar, Ron Oostdam, Ruben Fukkink, and Ludo Verhoeven. 2019. "Helping Parents Enhance Vocabulary Development in Preschool Children: Effects of a Family Literacy Program." *Early Childhood Research Quarterly* 48 (July):226–36. https://doi.org/10.1016/j.ecresq.2019.03.001.

*Tofail, Fahmida, Jena D. Hamadani, Fardina Mehrin, Deborah A. Ridout, Syed N. Huda, and Sally M. Grantham-McGregor. 2013. "Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-Deficient Children." *The Journal of Nutrition* 143 (6): 885–93. https://doi.org/10.3945/jn.112.160473.

*Tomlinson, Mark, Mary Jane Rotheram-Borus, Ingrid M. le Roux, Maryann Youssef, Sandahl H. Nelson, Aaron Scheffler, Robert E. Weiss, Mary O'Connor, and Carol M. Worthman. 2016. "Thirty-Six-Month Outcomes of a Generalist Paraprofessional Perinatal Home Visiting Intervention in South Africa on Maternal Health and Child

Health and Development." *Prevention Science* 17 (8): 937–48. https://doi.org/10.1007/s11121-016-0676-x.

*Vazir, Shahnaz, Patrice Engle, Nagalla Balakrishna, Paula L. Griffiths, Susan L. Johnson, Hilary Creed-Kanashiro, Sylvia Fernandez Rao, Monal R. Shroff, and Margaret E. Bentley. 2013. "Cluster-Randomized Trial on Complementary and Responsive Feeding Education to Caregivers Found Improved Dietary Intake, Growth and Development among Rural Indian Toddlers." *Maternal & Child Nutrition* 9 (1): 99–117. https://doi.org/10.1111/j.1740-8709.2012.00413.x.

*Wake, Melissa, Sherryn Tobin, Luigi Girolametto, Obioha C Ukoumunne, Lisa Gold, Penny Levickis, Jane Sheehan, Sharon Goldfeld, and Sheena Reilly. 2011. "Outcomes of Population Based Language Promotion for Slow to Talk Toddlers at Ages 2 and 3 Years: Let's Learn Language Cluster Randomised Controlled Trial." *BMJ* 343. https://doi.org/10.1136/bmj.d4741.

*Walker, Susan, Susan Chang, Christine Powell, and Sally Grantham-McGregor. 2004. "Psychosocial Intervention Improves the Development of Term Low-Birth-Weight Infants." *The Journal of Nutrition* 134 (6): 1417–23. https://doi.org/10.1093/jn/134.6.1417.

———. 2005. "Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-Stunted Jamaican Children: Prospective Cohort Study." *Lancet (London, England)* 366 (9499): 1804–7. https://doi.org/10.1016/S0140-6736(05)67574-5.

Walker, Susan, Susan Chang, Christine Powell, Emily Simonoff, and Sally Grantham-McGregor. 2006. "Effects of Psychosocial Stimulation and Dietary Supplementation in Early Childhood on Psychosocial Functioning in Late Adolescence: Follow-up of Randomised Controlled Trial." *BMJ (Clinical Research Ed.)* 333 (7566): 472. https://doi.org/10.1136/bmj.38897.555208.2F.

Walker, Susan, Susan Chang, Marcos Vera-Hernández, and Sally Grantham-McGregor. 2011. "Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior." *Pediatrics* 127 (5): 849–57. https://doi.org/10.1542/peds.2010-2231.

Walker, Susan, Susan Chang, Novie Younger, and Sally Grantham-McGregor. 2010. "The Effect of Psychosocial Stimulation on Cognition and Behaviour at 6 Years in a Cohort of Term, Low-Birthweight Jamaican Children." *Developmental Medicine and Child Neurology* 52 (7): e148-154. https://doi.org/10.1111/j.1469-8749.2010.03637.x.

Walker, Susan, Sally Grantham-Mcgregor, Christine Powell, and Susan Chang. 2000. "Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at

Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychosocial Stimulation." *The Journal of Pediatrics* 137 (1): 36–41. https://doi.org/10.1067/mpd.2000.106227.

*Wallander, Jan L., Carla M. Bann, Fred J. Biasini, Shivaprasad S. Goudar, Omrana Pasha, Elwyn Chomba, Elizabeth McClure, and Waldemar A. Carlo. 2014. "Development of Children at Risk for Adverse Outcomes Participating in Early Intervention in Developing Countries: A Randomized Controlled Trial." *Journal of Child Psychology and Psychiatry* 55 (11): 1251–59. https://doi.org/10.1111/jcpp.12247.

*Weisleder, Adriana, Denise Mazzuchelli, Aline Sá Lopez, Walfrido Duarte Neto, Carolyn Brockmeyer Cates, Hosana Alves Gonçalves, Rochele Paz Fonseca, João Oliveira, and Alan L. Mendelsohn. 2018. "Reading Aloud and Child Development: A Cluster-Randomized Trial in Brazil." *Pediatrics* 141 (1): e20170723. https://doi.org/10.1542/peds.2017-0723.

*Worku, Berhanu Nigussie, Teklu Gemechu Abessa, Mekitie Wondafrash, Johan Lemmens, Jan Valy, Liesbeth Bruckers, Patrick Kolsteren, and Marita Granitzer. 2018. "Effects of Home-Based Play-Assisted Stimulation on Developmental Performances of Children Living in Extreme Poverty: A Randomized Single-Blind Controlled Trial." *BMC Pediatrics* 18 (1): 29. https://doi.org/10.1186/s12887-018-1023-0.

*York, Benjamin N., Susanna Loeb, and Christopher Doss. 2019. "One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers." *Journal of Human Resources* 54 (3): 537–66. https://doi.org/10.3368/jhr.54.3.0517-8756R.

Zhang, Linlin, Derrick Ssewanyana, Marie-Claude Martin, Stephen Lye, Greg Moran, Amina Abubakar, Kofi Marfo, Joyce Marangu, Kerrie Proulx, and Tina Malti. 2021. "Supporting Child Development Through Parenting Interventions in Low- to Middle-Income Countries: An Updated Systematic Review." *Frontiers in Public Health* 9 (July). https://doi.org/10.3389/fpubh.2021.671988.

Table 1: Study Counts

	All		Low- and Middle-Income Countries		High-Income Countries	
	Number of Studies	*Average Sample Size*	*Number of Studies*	*Average Sample Size*	*Number of Studies*	*Average Sample Size*
All	90	514.83	43	756.86	47	293.40
Curriculum Type						
Cognitive Stimulation	54	636.57	38	784.05	16	286.31
Language	14	266.07	2	384.50	12	246.33
Lifestyle	19	408.58	5	552.40	14	357.21
Attachment	17	261.94	3	345.00	14	244.14
Delivery Mode						
Home Visiting	63	425.19	31	571.65	32	283.31
Group	29	726.93	18	1000.56	11	279.18
Clinic	9	596.33	5	1011.00	4	78.00
ICT	5	592.60	0	0.00	5	592.60
Baseline Age						
Prenatal-Birth	26	341.81	6	481.00	20	300.05
0-1	17	1127.88	14	1298.36	3	332.33
1-2	26	413.77	18	533.11	8	145.25
2+	21	357.90	5	377.20	16	351.88
Session Frequency						
Quarter	17	412.35	7	518.86	10	337.80
Half	25	386.08	14	459.43	11	292.73
More than half	48	618.19	22	1021.86	26	276.62
Program Duration						
0-6 Months	27	218.67	10	235.50	17	208.76
7-12 Months	29	528.66	18	636.22	11	352.64
13-24 Months	22	511.82	11	623.91	11	399.73
24+ Months	12	1153.33	4	2968.75	8	245.63
Sample Size						
50-200	35	125.71	11	134.18	24	121.83
201-400	20	286.05	10	301.90	10	270.20
401-600	16	481.56	9	460.00	7	509.29
Over 600	19	1500.47	13	1839.23	6	766.50
Year of Study						
1980s	3	121.00	0	0.00	3	121.00
1990s	5	323.40	1	127.00	4	372.50
2000s	15	270.40	7	209.57	8	323.63
2010s	49	603.22	25	945.20	24	247.00
2020s	18	596.72	10	732.10	8	427.50

Notes: This table reports the frequency of studies alongside average sample sizes of children in the studies, divided after the first row into categories based on program and study characteristics. The frequencies and average sample sizes are reported for the full study sample in the leftmost columns, and then separately for low- and middle-income countries and high-income countries, respectively.

Table 2: Pooled Estimates

	All							LMICs							HICs						
	B	*SE*	*k*	*n*	*PI*	*Tau-2*	*Q*	*B*	*SE*	*k*	*n*	*PI*	*Tau-2*	*Q*	*B*	*SE*	*k*	*n*	*CI*	*Tau-2*	*Q*
All	0.218***	0.029	[90]	225		0.080	718.4	0.303	0.047	[43]	105	(-0.258, 0.863)	0.109	551.7	0.115	0.026	[47]	120	(-0.108, 0.337)	0.017	97.5
Curriculum Type																					
Cognitive Stimulation	0.291***	0.043	[54]	144	(-0.267, 0.849)	0.109	621.0	0.330	0.052	[38]	95	(-0.260, 0.920)	0.119	523.4	0.178	0.066	[16]	49	(-0.215, 0.571)	0.046	60.4
Language	0.150***	0.044	[14]	35	(-0.092, 0.391)	0.016	26.2	0.176	0.094	[2]	6	(0.000, 0.000)	0.027	3.5	0.144	0.052	[12]	29	(-0.114, 0.402)	0.018	21.7
Lifestyle	0.133**	0.051	[19]	39	(-0.200, 0.467)	0.034	73.5	0.252	0.154	[5]	12	(-0.589, 1.093)	0.104	54.5	0.083	0.031	[14]	27	(0.028, 0.138)	0.000	12.4
Attachment	0.152**	0.053	[17]	49	(-0.193, 0.498)	0.036	47.6	0.130	0.100	[3]	5	(-1.490, 1.750)	0.056	9.1	0.158	0.064	[14]	44	(-0.202, 0.519)	0.037	38.1
Delivery Mode																					
Home Visiting	0.209***	0.032	[63]	146	(-0.236, 0.655)	0.070	448.6	0.281	0.053	[31]	71	(-0.256, 0.819)	0.097	358.9	0.116	0.029	[32]	75	(-0.091, 0.323)	0.014	58.9
Group	0.229***	0.051	[29]	97	(-0.204, 0.662)	0.062	176.0	0.257	0.067	[18]	56	(-0.208, 0.722)	0.066	129.7	0.175	0.079	[11]	41	(-0.261, 0.612)	0.051	39.8
Clinic	0.317*	0.134	[9]	21	(-0.498, 1.132)	0.167	163.2	0.404	0.213	[5]	9	(-0.730, 1.538)	0.187	158.6	0.194	0.099	[4]	12	(-0.298, 0.687)	0.019	4.2
ICT	0.0	0.048	[5]	17	(-0.227, 0.274)	0.009	9.1	0.000	0.000	[0]	0	(0.000, 0.000)	0.000	0.0	0.023	0.048	[5]	17	(-0.227, 0.274)	0.009	9.1
Baseline Age																					
Prenatal Birth	0.130***	0.029	[26]	56	(-0.048, 0.308)	0.010	43.0	0.220	0.059	[6]	9	(-0.134, 0.574)	0.024	15.9	0.093	0.028	[20]	47	(-0.010, 0.197)	0.003	22.3
0-1	0.236**	0.074	[17]	54	(-0.209, 0.682)	0.059	133.8	0.283	0.084	[14]	44	(-0.193, 0.759)	0.064	124.0	0.007	0.053	[3]	10	(-0.328, 0.342)	0.000	2.0
1-2	0.366***	0.066	[26]	60	(-0.3060, 1.091)	0.176	355.2	0.383	0.082	[18]	42	(-0.404, 1.169)	0.196	330.4	0.322	0.107	[8]	18	(-0.201, 0.844)	0.061	20.6
2+	0.099**	0.036	[21]	55	(-0.140, 0.338)	0.018	50.8	0.138	0.077	[5]	10	(-0.254, 0.529)	0.022	11.5	0.088	0.041	[16]	45	(-0.161, 0.337)	0.018	37.9
Session Frequency																					
Quarter	0.109**	0.037	[17]	38	(-0.105, 0.323)	0.014	35.9	0.199	0.064	[7]	12	(-0.109, 0.507)	0.019	19.6	0.023	0.026	[10]	26	(-0.024, 0.071)	0.000	8.0
Half	0.244***	0.061	[25]	54	(-0.346, 0.834)	0.115	246.6	0.298	0.093	[14]	35	(-0.437, 1.033)	0.162	203.8	0.149	0.056	[11]	19	(-0.117, 0.415)	0.018	21.5
More than half	0.241***	0.042	[48]	133	(-0.261, 0.743)	0.088	415.1	0.334	0.069	[22]	58	(-0.275, 0.943)	0.120	313.8	0.140	0.041	[26]	75	(-0.138, 0.419)	0.025	64.2
Program Duration																					
0-6 Months	0.253***	0.056	[27]	68	(-0.240, 0.746)	0.080	123.9	0.355	0.084	[10]	19	(-0.174, 0.884)	0.074	36.3	0.187	0.070	[17]	49	(-0.252, 0.625)	0.058	61.4
7-12 Months	0.241***	0.058	[29]	69	(-0.248, 0.730)	0.079	275.2	0.324	0.084	[18]	46	(-0.252, 0.900)	0.102	243.5	0.074	0.038	[11]	23	(-0.047, 0.195)	0.003	12.2
13-24 Months	0.191***	0.055	[22]	44	(-0.419, 0.800)	0.122	289.3	0.272	0.100	[11]	27	(-0.557, 1.102)	0.195	237.3	0.083	0.035	[11]	17	(0.019, 0.147)	0.000	9.8
24+ Months	0.135***	0.039	[12]	44	(-0.067, 0.337)	0.011	20.9	0.170	0.048	[4]	13	(-0.129, 0.469)	0.008	6.5	0.107	0.063	[8]	31	(-0.182, 0.396)	0.018	13.6

Notes: B = pooled effect size for the subset of estimates in question; SE = standard error of the estimates in question; k = number of studies included in the estimate; n = number of effect sizes included in the estimate; PI = prediction intervals

Table 3: Pooled Estimates Study Characteristics

	All							LMICs							HICs						
	B	*SE*	*k*	*n*	*CI*	*Tau-2*	*Q*	*B*	*SE*	*k*	*n*	*CI*	*Tau-2*	*Q*	*B*	*SE*	*k*	*n*	*CI*	*Tau-2*	*Q*
All	0.218***	0.029	[90]	225	(-0.254, 0.689)	0.080	718.4	0.303***	0.047	[43]	105	(-0.258, 0.863)	0.109	551.7	0.115***	0.026	[47]	120	(-0.108, 0.337)	0.017	97.5
Outcome Measure																					
Cognitive	0.291***	0.043	[54]	144	(-0.267, 0.849)	0.109	621.0	0.330***	0.052	[38]	95	(-0.260, 0.920)	0.119	523.4	0.178***	0.066	[16]	49	(-0.215, 0.571)	0.046	60.4
Language	0.150***	0.044	[14]	35	(-0.092, 0.391)	0.016	26.2	0.176***	0.094	[2]	6	(0.000, 0.000)	0.027	3.5	0.144***	0.052	[12]	29	(-0.114, 0.402)	0.018	21.7
Sample Size																					
50-200	0.302***	0.039	[35]	67	(0.030, 0.574)	0.024	57.5	0.390***	0.069	[11]	16	(0.133, 0.648)	0.015	14.3	0.259***	0.048	[24]	51	(-0.032, 0.551)	0.027	40.3
201-400	0.140***	0.033	[20]	46	(-0.044, 0.323)	0.010	32.1	0.213***	0.040	[10]	22	(0.056, 0.370)	0.006	12.5	0.064***	0.040	[10]	24	(-0.082, 0.210)	0.005	11.8
401-600	0.186*	0.080	[16]	56	(-0.370, 0.743)	0.093	143.8	0.330*	0.127	[9]	26	(-0.423, 1.084)	0.142	104.8	0.000	0.014	[7]	30	(-0.089, 0.089)	0.002	7.2
More than 600	0.203**	0.078	[19]	56	(-0.397, 0.803)	0.113	466.6	0.282**	0.107	[13]	41	(-0.422, 0.987)	0.142	413.8	0.036**	0.036	[6]	15	(-0.123, 0.196)	0.004	9.0
Year																					
1980s-1990s	0.1	0.040	[8]	14	(-0.025, 0.143)	0.000	7.1	.	.	.	.	(- , -)	.	.	0.0	0.024	[7]	12	(-0.012, 0.084)	0.000	1.7
2000s	0.253***	0.058	[15]	28	(-0.090, 0.596)	0.034	40.7	0.414***	0.070	[7]	9	(0.141, 0.687)	0.013	9.4	0.112***	0.047	[8]	19	(-0.094, 0.318)	0.009	11.0
2010s	0.224***	0.041	[49]	129	(-0.231, 0.679)	0.072	357.1	0.285***	0.066	25	64	(-0.251, 0.820)	0.093	292.3	0.142***	0.041	[24]	65	(-0.128, 0.412)	0.023	52.9
2020s	0.205**	0.072	[18]	54	(-0.448, 0.858)	0.135	298.9	0.247**	0.099	10	30	(-0.551, 1.045)	0.174	223.9	0.112**	0.083	[8]	24	(-0.277, 0.501)	0.033	29.6
Publication Type																					
Published	0.208***	0.029	[84]	215	(-0.192, 0.609)	0.057	465.4	0.293***	0.046	40	99	(-0.188, 0.774)	0.079	327.4	0.110***	0.026	[44]	116	(-0.111, 0.331)	0.017	91.5
Evaluation	0.204*	0.096	[3]	6	(-0.750, 1.158)	0.014	4.0	.	.	.	.	(- , -)	.	.	0.333*	0.043	[2]	3	(- , -)	.	.
Working Paper	0.2	0.203	[5]	8	(-1.181, 1.650)	0.320	210.5	0.4	0.333	3	6	(-4.210, 4.960)	0.417	160.0	0.0	0.011	[2]	2	(- , -)	.	.

Notes: B = pooled effect size for the subset of estimates in question; SE = standard error of the pooled effect size; k = number of studies included in the estimate; n = number of effect sizes included in the estimate; PI = prediction intervals

Table 4A: Multivariate Regression

	Specification 1	Specification 2	Specification 3	Specification 4	Specification 5	Specification 6	Specification 7
	B	*B*	*B*	*B*	*B*	*B*	*B*
	(SE)	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*
LMIC Dummy	0.198***		0.118*		0.185***		0.092
	(0.049)		(0.060)		(0.050)		(0.062)
Ln Sample Size	-0.073*		-0.070*		-0.071*		-0.059
	(0.036)		(0.036)		(0.040)		(0.043)
Language		-0.144*	-0.095			-0.113	-0.088
		(0.070)	(0.073)			(0.076)	(0.082)
Attachment		-0.137*	-0.093			-0.131*	-0.101
		(0.075)	(0.076)			(0.075)	(0.074)
Lifestyle		-0.240***	-0.176*			-0.249***	-0.192**
		(0.058)	(0.070)			(0.058)	(0.074)
Group				0.017	-0.004	-0.018	-0.016
				(0.064)	(0.065)	(0.068)	(0.075)
Individual-Clinic				0.114	0.099	0.048	0.063
				(0.147)	(0.143)	(0.148)	(0.149)
ICT				-0.187**	-0.049	-0.225*	-0.129
				(0.066)	(0.078)	(0.089)	(0.110)
Constant	0.532**	0.311***	0.619**	0.213***	0.520*	0.323***	0.584**

Notes: Each column reports coefficients and standard errors from multivariate meta-regressions using robust variance estimations

Table 4B: Multivariate Regression

	Specification 1	Specification 2	Specification 3	Specification 4	Specification 5	Specification 6	Specification 7
	B	*B*	*B*	*B*	*B*	*B*	*B*
	(SE)	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*
LMIC Dummy	0.198***		0.134**		0.186***		0.132**
	(0.049)		(0.051)		(0.047)		(0.051)
Ln Sample Size	-0.073*		-0.059		-0.067*		-0.049
	(0.036)		(0.041)		(0.038)		(0.046)
Up to 1 Year		0.103	0.055			0.079	0.035
		(0.080)	(0.080)			(0.076)	(0.082)
Up to 2 Years		0.231**	0.167*			0.190*	0.137*
		(0.076)	(0.080)			(0.078)	(0.081)
Over 2 Years		-0.012	0.001			-0.063	-0.036
		(0.050)	(0.047)			(0.058)	(0.057)
Freq Half				0.125*	0.091	0.066	0.060
				(0.072)	(0.068)	(0.068)	(0.065)
Freq More				0.109*	0.096*	0.060	0.065
				(0.065)	(0.054)	(0.061)	(0.055)
7-12				0.011	0.008	-0.039	-0.022
				(0.089)	(0.084)	(0.081)	(0.087)
13-24				-0.049	-0.022	-0.085	-0.052
				(0.087)	(0.079)	(0.084)	(0.088)
Over 24				-0.079	-0.043	-0.097	-0.060
				(0.081)	(0.080)	(0.083)	(0.090)
Constant	0.532**	0.134***	0.428*	0.145*	0.439*	0.161*	0.371
	(0.203)	(0.029)	(0.230)	(0.083)	(0.212)	(0.087)	(0.239)

Notes: Each column reports coefficients and standard errors from multivariate meta-regressions using robust variance estimations

Table 4C: Multivariate Regression

	Specification 1	Specification 2	Specification 3	Specification 4	Specification 5	Specification 6	Specification 7
	B	*B*	*B*	*B*	*B*	*B*	*B*
	(SE)	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*
LMIC Dummy	0.198***		0.085		0.117*		0.087
	(0.049)		(0.056)		(0.062)		(0.057)
Ln Sample Size	-0.073*		-0.055		-0.069*		-0.047
	(0.036)		(0.042)		(0.038)		(0.047)
Language		-0.070	-0.054	-0.136*	-0.082	-0.068	-0.043
		(0.079)	(0.076)	(0.071)	(0.077)	(0.082)	(0.082)
Attachment		-0.110*	-0.086	-0.125*	-0.079	-0.111*	-0.079
		(0.066)	(0.067)	(0.073)	(0.077)	(0.065)	(0.069)
Lifestyle		-0.198**	-0.158*	-0.221***	-0.164*	-0.192**	-0.153*
		(0.064)	(0.072)	(0.062)	(0.077)	(0.067)	(0.078)
Up to 1 Year		-0.002	-0.001			-0.018	-0.016
		(0.079)	(0.088)			(0.080)	(0.093)
Up to 2 Years		0.117	0.100			0.086	0.075
		(0.075)	(0.081)			(0.079)	(0.084)
Over 2 Years		-0.088	-0.061			-0.130*	-0.093
		(0.065)	(0.067)			(0.070)	(0.077)
Freq Half				0.080	0.075	0.055	0.056
				(0.069)	(0.068)	(0.064)	(0.065)
Freq More				0.064	0.073	0.043	0.053
				(0.059)	(0.054)	(0.055)	(0.054)
7-12				-0.015	0.013	-0.050	-0.024
				(0.085)	(0.084)	(0.082)	(0.089)
13-24				-0.021	0.015	-0.068	-0.033
				(0.088)	(0.084)	(0.083)	(0.087)
Over 24				-0.066	-0.028	-0.105	-0.067
				(0.082)	(0.082)	(0.081)	(0.089)
Constant	0.532**	0.275***	0.529*	0.268**	0.544*	0.306***	0.477*

Notes: Each column reports coefficients and standard errors from multivariate meta-regressions using robust variance estimations

Table 4D: Multivariate Regression

	Specification 1	Specification 2	Specification 3	Specification 4	Specification 5	Specification 6	Specification 7
	B	*B*	*B*	*B*	*B*	*B*	*B*
	(SE)	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*	*(SE)*
LMIC Dummy	0.198***		0.163***		0.118*		0.069
	(0.049)		(0.048)		(0.052)		(0.060)
Ln Sample Size	-0.073*		-0.057		-0.048		-0.043
	(0.036)		(0.044)		(0.049)		(0.050)
Group		-0.004	-0.013	-0.016	-0.013	-0.009	-0.006
		(0.063)	(0.067)	(0.066)	(0.066)	(0.066)	(0.070)
Individual-Clinic		0.095	0.096	0.090	0.095	0.056	0.071
		(0.150)	(0.143)	(0.147)	(0.141)	(0.151)	(0.146)
ICT		-0.264**	-0.119	-0.162*	-0.078	-0.170	-0.108
		(0.082)	(0.098)	(0.097)	(0.103)	(0.106)	(0.113)
Freq Half		0.135*	0.105	0.085	0.075	0.071	0.070
		(0.077)	(0.074)	(0.072)	(0.070)	(0.071)	(0.071)
Freq More		0.149*	0.118*	0.092	0.084	0.075	0.075
		(0.071)	(0.062)	(0.068)	(0.062)	(0.064)	(0.061)
7-12		0.020	0.006	-0.027	-0.019	-0.032	-0.016
		(0.087)	(0.085)	(0.086)	(0.091)	(0.086)	(0.093)
13-24		-0.055	-0.028	-0.070	-0.038	-0.055	-0.023
		(0.088)	(0.080)	(0.087)	(0.086)	(0.085)	(0.086)
Over 24		-0.083	-0.049	-0.076	-0.046	-0.085	-0.054
		(0.083)	(0.085)	(0.088)	(0.091)	(0.086)	(0.091)
Up to 1 Year				0.094	0.057	-0.005	0.002
				(0.083)	(0.077)	(0.083)	(0.085)
Up to 2 Years				0.189*	0.145*	0.087	0.082
				(0.083)	(0.086)	(0.088)	(0.090)
Over 2 Years				-0.007	-0.003	-0.075	-0.055
				(0.069)	(0.062)	(0.080)	(0.079)
Language						-0.064	-0.047
						(0.082)	(0.088)
Attachment						-0.101	-0.078
						(0.064)	(0.067)
Lifestyle						-0.186**	-0.151*
						(0.069)	(0.082)
Constant	0.532**	0.130	0.385	0.120	0.337	0.262*	0.429
	(0.203)	(0.095)	(0.240)	(0.100)	(0.264)	(0.111)	(0.275)

Notes: Each column reports coefficients and standard errors from multivariate meta-regressions using robust variance estimations

Table 5A: Egger's Regressions, No Controls

	Pooled Effect Size		W		Sample Size	
	B	*(SE)*	*B*	*(SE)*	*n*	*[k]*
All	0.218***	(0.029)	1.049	(0.802)	90	[225]
Country Type						
LMIC	0.303***	(0.047)	1.385	(1.195)	43	[105]
HIC	0.115***	(0.026)	2.286**	(0.712)	47	[120]
Curriculum Type						
Cognitive Stimulation	0.291***	(0.043)	1.733	(1.165)	54	[144]
Language	0.150***	(0.044)	2.040*	(1.162)	14	[35]
Lifestyle	0.152**	(0.053)	1.645	(1.206)	17	[49]
Attachment	0.133**	(0.051)	-0.782	(1.043)	19	[39]
Delivery Mode						
Home Visiting	0.209***	(0.032)	1.699*	(0.939)	63	[146]
Group	0.229***	(0.051)	3.200*	(1.457)	29	[97]
Clinic	0.317*	(0.134)	-1.716	(2.222)	9	[21]
ICT	0.023	(0.048)	0.989	(5.137)	5	[17]
Baseline Age						
Prenatal-Birth	0.130***	(0.029)	0.286	(0.811)	26	[56]
0-1	0.236**	(0.074)	2.252	(2.257)	17	[54]
1-2	0.366***	(0.066)	-0.270	(1.898)	26	[60]
2+	0.099**	(0.036)	3.437***	(0.798)	21	[55]
Session Frequency						
Quarter	0.109**	(0.037)	0.828	(1.057)	17	[38]
Half	0.244***	(0.061)	-1.050	(1.695)	25	[54]
More than half	0.241***	(0.042)	1.750	(1.086)	48	[133]
Program Duration						
0-6 Months	0.253***	(0.056)	2.223	(1.714)	27	[68]
7-12 Months	0.241***	(0.058)	0.140	(1.385)	29	[69]
13-24 Months	0.191***	(0.055)	0.894	(2.370)	22	[44]
24+ Months	0.135***	(0.039)	0.974	(1.306)	12	[44]
Sample Size						
50-200	0.302***	(0.039)	-0.527	(1.364)	35	[67]
201-400	0.140***	(0.033)	0.625	(3.182)	20	[46]
401-600	0.186*	(0.080)	6.095	(7.415)	16	[56]
More than 600	0.203**	(0.078)	-6.272	(8.478)	19	[56]
Year						
1980-1999	0.059	(0.040)	1.405	(1.032)	8	[14]
2000-2009	0.253***	(0.058)	2.409	(1.661)	15	[28]
2010-2019	0.224***	(0.041)	1.102	(1.051)	49	[129]

Notes: The "Pooled Effect Size" columns respectively contain the pooled effect size coefficients and standard errors for the subset of studies in question (repeated from previous tables). The "W" column shows the coefficient for W in an Egger's regression (using robust variance estimation) with no controls. n = number of studies; k = number of estimates.

Table 5B: Egger's Regressions, Small Sample Control

	Pooled Effect Size		W		Sample Size	
	B	*(SE)*	*B*	*(SE)*	*n*	*[k]*
All	0.218***	(0.029)	-0.657	(1.321)	90	[225]
Country Type						
LMIC	0.218***	(0.029)	-0.657	(1.321)	90	[225]
HIC	0.291***	(0.043)	0.707	(1.848)	54	[144]
Curriculum Type						
Cognitive Stimulation	0.150***	(0.044)	-1.444	(1.872)	14	[35]
Language	0.152**	(0.053)	-2.569	(1.909)	17	[49]
Lifestyle	0.133**	(0.051)	-1.259	(1.327)	19	[39]
Attachment	0.209***	(0.032)	0.773	(1.557)	63	[146]
Delivery Mode						
Home Visiting	0.229***	(0.051)	1.167	(2.622)	29	[97]
Group	0.317*	(0.134)	3.044	(12.441)	9	[21]
Clinic	0.023	(0.048)	0.989	(5.137)	5	[17]
ICT	0.130***	(0.029)	-1.171	(1.044)	26	[56]
Baseline Age						
Prenatal-Birth	0.236**	(0.074)	4.675	(3.646)	17	[54]
0-1	0.366***	(0.066)	-3.408	(2.583)	26	[60]
1-2	0.099**	(0.036)	0.851	(1.415)	21	[55]
2+	0.109**	(0.037)	-1.267	(1.613)	17	[38]
Session Frequency						
Quarter	0.244***	(0.061)	-3.706	(2.353)	25	[54]
Half	0.241***	(0.042)	0.420	(1.716)	48	[133]
More than half	0.253***	(0.056)	-0.896	(2.087)	27	[68]
Program Duration						
0-6 Months	0.241***	(0.058)	0.608	(2.462)	29	[69]
7-12 Months	0.191***	(0.055)	-0.932	(4.968)	22	[44]
13-24 Months	0.135***	(0.039)	-1.756	(2.266)	12	[44]
24+ Months	0.302***	(0.039)	-0.527	(1.364)	35	[67]
Sample Size						
50-200	0.140***	(0.033)	0.625	(3.182)	20	[46]
201-400	0.186*	(0.080)	6.095	(7.415)	16	[56]
401-600	0.203**	(0.078)	-6.272	(8.478)	19	[56]
More than 600	0.059	(0.040)	0.009	(2.064)	8	[14]
Year						
1980-1999	0.253***	(0.058)	-1.235	(2.058)	15	[28]
2000-2009	0.224***	(0.041)	0.728	(1.632)	49	[129]
2010-2019	0.224***	(0.041)	0.728	(1.632)	49	[129]

Notes: The "Pooled Effect Size" columns respectively contain the pooled effect size coefficients and standard errors for the subset of studies in question (repeated from previous tables). The "W" column shows the coefficient for W in an Egger's regression (using robust variance estimation) with no controls. n = number of studies; k = number of estimates.