ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-964

The Challenges of Scaling up Effective Child-Rearing Practices Using Technology in Developing Settings: Experimental Evidence From India

Irma Arteaga
University of Missouri

Andreas de Barros
University of California, Irvine

Alejandro J. Ganimian
New York University

Home-visitation programs have improved child development in low- and middle-income countries, but they are costly to scale due to their reliance on trained workers. We evaluated an inexpensive and low-tech alternative with 2,433 caregivers of children aged 6 to 30 months served by 250 public childcare centers in Uttarakhand, India: automated phone calls offering parenting advice. The intervention was implemented largely as intended, with more than two-thirds of caregivers completing at least 10 calls. Yet, counter to expectations, it had negative but statistically insignificant effects on caregivers' knowledge and interactions with their children, reduced their self-efficacy (by 0.11 standard deviations), and increased their anxiety (by 0.10 standard deviations). Consistent with this pattern, it had precisely estimated null effects on children's development and language. An analysis of program materials suggests four reasons why the program may not have had the desired effects.

VERSION: May 2024

Suggested citation: Arteaga, Irma, Andreas de Barros, and Alejandro J. Ganimian. (2024). The Challenges of Scaling up Effective Child-Rearing Practices Using Technology in Developing Settings: Experimental Evidence From India. (EdWorkingPaper: 24-964). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/9fb8-g360

The Challenges of Scaling up Effective Child-Rearing Practices Using Technology in Developing Settings: Experimental Evidence From India*

Irma Arteaga†
University of Missouri

Andreas de Barros‡
University of California, Irvine

Alejandro J. Ganimian§
New York University

May 1, 2024

Abstract

Home-visitation programs have improved child development in low- and middle-income countries, but they are costly to scale due to their reliance on trained workers. We evaluated an inexpensive and low-tech alternative with 2,433 caregivers of children aged 6 to 30 months served by 250 public childcare centers in Uttarakhand, India: automated phone calls offering parenting advice. The intervention was implemented largely as intended, with more than two-thirds of caregivers completing at least 10 calls. Yet, counter to expectations, it had negative but statistically insignificant effects on caregivers' knowledge and interactions with their children, *reduced* their self-efficacy (by 0.11 standard deviations), and *increased* their anxiety (by 0.10 standard deviations). Consistent with this pattern, it had precisely estimated null effects on children's development and language. An analysis of program materials suggests four reasons why the program may not have had the desired effects.

Keywords: early childhood development; education technology; state capacity; India.

JEL codes: C93; I21; I25; I38; J13.

*We thank *Dost* and the Government of Uttarakhand for making this study possible. We also thank Sneha Sheth, Sindhuja Jeyabal, Prateek Agarwal, Neelansha Dwivedi, Aronee Ghosh, and Jhanvi Kotak for their support at various stages of the project. We thank Kenji Kitamura, Dana McCoy, Jonathan Seiden, Marcus Waldman, and their team at Harvard University for discussing the study's use of the CREDI instrument and making data and code available. We thank Sally Kendall for discussing the study's use of the TOPSE instrument. Arja Dayal and Rashi Maheshwari provided excellent research assistance. The study was registered with the AEA Trial Registry (RCT ID: AEARCTR-0010129). It was approved by the Internal Review Boards at New York University and the Institute for Financial Management and Research. We gratefully acknowledge funding from The Agency Fund and the Institute for Human Development and Social Change at New York University. The authors have no competing interests to declare. All views expressed are those of the authors and not of any of the institutions with which they are affiliated.

†Associate Professor, Truman School of Government and Public Affairs, University of Missouri. E-mail: arteagai@missouri.edu

‡Assistant Professor, University of California, Irvine. Corresponding author. E-mail: adb@uci.edu. Postal address: UC Irvine Education Building, Office 2042, Irvine, CA 92697, USA

§Assistant Professor of Applied Psychology and Economics, Steinhardt School of Culture, Education, and Human Development, New York University. E-mail: alejandro.ganimian@nyu.edu.

1 Introduction

Children's early interactions with their caregivers have lasting impacts on their life outcomes. During the first years of life, vital development occurs in multiple domains (CISCD, 2000). Specifically, brain development in some domains (e.g., seeing and hearing) starts earlier and sets the foundation for others (e.g., receptive language and cognitive functions), and disruptions can impact the brain's structure and function (Grantham-McGregor et al., 2007). Environmental factors, including maternal caregiving, can catalyze or delay this process, affecting children's cognitive and emotional development (Young, 2002; Landry et al., 2006), and in turn, their schooling and productivity as adults (Psacharopoulos and Patrinos, 2004).

Children in low- and middle-income countries (LMICs) are in particular need of interactions with their caregivers that are developmentally appropriate (i.e., match their emerging skills). They are disproportionately likely to face risk factors that may disrupt their development. According to a recent estimate (Lu et al., 2016), 249 million young children in these settings are exposed to two widely measured such factors—growth stunting and poverty—and are therefore at risk of not reaching their developmental potential (see also Walker et al., 2007). Interactions between caregivers and children can ameliorate the deleterious effects of these factors by promoting neurocognitive processing and brain functioning (Engle et al., 2007).

Interventions that encourage caregivers to provide psycho-social stimulation to their children have improved development, school performance, and labor-market outcomes in LMICs. Most famously, a program in Kingston, Jamaica, in which community health aides visited the mothers of 129 stunted children ages 9-24 months to facilitate weekly play sessions at home impacted development outcomes after two years (Grantham-McGregor et al., 1991). By age 17-18, those who had been randomly assigned to the intervention performed better on fluid intelligence and language development than their control peers (Walker et al., 2005). And 20 years after the program, the wages of its beneficiaries were 25% higher than those of the control group and on par with those of a non-stunted group (Gertler et al., 2014).

Replicating the success of this intervention at scale, however, has proven to be challenging. In recent years, many have sought to promote early stimulation through various modalities, including conditional cash transfers in Colombia (Attanasio et al., 2014), a health program in Sindh, Pakistan (Yousafzai et al., 2014), a program for pregnant and vulnerable women in Colombia (Attanasio et al., 2018), home visits in urban areas of Odisha, India (Andrew et al., 2020), and mother group sessions in rural Odisha (Grantham-McGregor et al., 2020). These delivery mechanisms have boosted children's cognitive, language, and motor skills, but they require staff to engage with caregivers, which is both time-consuming and costly.

We evaluated an inexpensive and low-tech alternative: automated phone calls with parenting advice for caregivers of children aged 6 to 30 months. We partnered with a nonprofit (*Dost*)

to randomize the rollout of this program in the Indian state of Uttarakhand for 2,433 caregivers served by 250 public childcare centers, which are locally known as *anganwadis*. These caregivers are supposed to receive regular home visits from workers at these centers. As Ganimian et al. (2023) have shown, however, these workers are severely overburdened. They are expected to complete myriad tasks across health, nutrition, and education, and as a result, some of those tasks occur less frequently or for less time than regulations stipulate. The calls sought to offer caregivers a consistent source of evidence-based child-rearing guidance and complement workers' in-person (but possibly less frequent) home visits. Given the scope of India's public childcare system, the Integrated Child Development Services (ICDS), which serves 46 million children ages 0 to 3 and 36 million children ages 3 to 6, this intervention had considerable potential to meaningfully impact child development.

The intervention was implemented largely as intended during the eight months of the evaluation: 69% of caregivers who signed up to receive the calls and were randomly assigned to receive them completed at least 10 of them, and 81% completed at least five. Yet, counter to expectations, it *reduced* caregivers' knowledge of child development (by 0.11 standard deviations [SDs], $q = 0.11$) and self-efficacy about child rearing (by 0.11 SDs, $q = 0.06$), it *increased* their anxiety (by 0.10 SDs, $q = 0.08$), and it had a null effect on their interactions with their children (-0.01 SDs, $q = 0.80$). Only the impacts on self-efficacy and anxiety were statistically significant once we account for multiple hypothesis testing, but the general pattern of results for caregivers is in the opposite direction than intended. Consistent with this pattern, it had precisely estimated null effects on child development (0.01 SDs, $q = 0.80$) and language and vocabulary skills (0.05 SDs, $q = 0.49$).

Our first and most direct contribution is to the scarce but growing research literature on the use of technology to scale up evidence-based guidance on child-rearing practices in LMICs. To our knowledge, there have only been three prior evaluations of similar interventions, which have yielded mixed results. Arteaga and Trias (2023) found that automated calls for 1,400 caregivers of 6-to-33-month-olds in rural Guatemala increased caregivers' interactions with their children (by 0.11 SDs) and improved children's vocabulary (by 0.22 SDs) after only two months, but had no impact on maternal anxiety or overall child development. Smith et al. (2023) found that a parent manual and play materials, bi-weekly calls from community health workers, and weekly text messages for 117 caregivers of 5-to-24-month olds in Jamaica increased interactions with children (by 0.34 SDs) and praise (a twofold increase), but had no impact on play materials in the home or use of interactive language. Lastly, Rafla et al. (2024) found that phone calls from community health volunteers to 1,158 caregivers of 6-to-36-month-olds in Jordan reduced caregiver depression (by 0.11 SDs) but had no effect on their anxiety or self-efficacy or on children's behavior and development. We compare our intervention to the most effective of the three and identify four potential

reasons for our disappointing results: the flexible pace at which caregivers completed the program, the lack of differentiated activities by children's age, an insufficient focus on child development, and the absence of caregiver-child activities or games linked to each call. We also explain why these hypotheses are consistent with the effects of the other studies, thereby providing a roadmap for advancing global evidence on this important question.

Our study also contributes to the broader literature on the use of technology—calls, text messages, mobile platforms—to complement public-service delivery of education services. The COVID-19 pandemic has revealed how interruptions to such services may adversely affect child development (Yoshikawa et al., 2020; Abufhele et al., 2021; McCoy et al., 2021; González et al., 2022). Yet, impact evaluations of efforts to mitigate the effects of such shocks in LMICs have mostly focused on older children (Angrist et al., 2022, 2023; Singh et al., 2024). Our study illustrates the challenges of efforts to build comparable supports for younger children, for whom interruptions in service delivery are likely even more consequential.

Lastly, our results add to existing evidence on how interventions that seek to encourage effective parenting practices for low-income families impact caregivers' mental health. Psychological theory indicates that economic hardship contributes to psychological distress, adversely affecting parenting quality and hence child outcomes (Masarik and Conger, 2017). Based on this premise, parenting programs seek to ameliorate the deleterious effects of poverty on children by breaking the link between caregiver distress and parenting quality. Yet, recent studies find some interventions *increase* caregiver stress (Magnuson et al., 2022). Our study suggests that one possible mechanism through which this may occur is by widening the gap between what caregivers believe they *should* do and what they *can* do.

2 Setting and intervention

India is home to one in five of the world's children (UNICEF, 2023). Unfortunately, however, a third of Indian children weigh less and are less tall than they should be for their age, according to global standards (MHFW, 2020). Early learning levels are also alarmingly low: in a recent representative survey of rural India, only one in five four-year-olds in the public pre-school system could count objects and one in three could compare two quantities; just one in two could describe a picture, and one in ten could understand a story (ASER, 2019).

2.1 The Integrated Child Development Services (ICDS)

India is also home to the world's largest public childcare system: the Integrated Child Development Services (ICDS). When it was created in 1975, it targeted disadvantaged areas,

but it subsequently expanded to the rest of the country. It now serves over 46 million children ages 0 to 3 and 36 million ages 3 to 6. It provides a range of services through *anganwadi* centers. Each center serves 400-800 people and is staffed by a worker, who is responsible for myriad tasks across health, nutrition, and education,[1] and a helper tasked with cleaning, cooking, and accompanying children to and from the centers (PEO, 2011). Several studies have found that children exposed to *anganwadi* centers had better nutrition and education outcomes (Hazarika and Viren, 2013; Nandi et al., 2020; Ravindran, 2020).

ICDS is coordinated by the central government and managed by the states. The central government stipulates that *anganwadi* workers should be women ages 18 to 44 from the local village who passed the grade 12 board exams (i.e., graduated high school; MWCD, 2015). They are not civil servants (like public school teachers) but rather "honorary workers" and, as such, receive a monthly honorarium financed by the central and state governments.[2] A recent study found that the multiple tasks for which the workers are responsible often lead them to engage in each task less frequently than expected (Ganimian et al., 2023).

One of the tasks *anganwadi* workers are expected to complete is to visit homes to encourage mothers to "play an effective role in children's growth and development" (MWCD, 2022b). Specifically, workers are expected to conduct four visits *per day*: one to a pregnant mother, one to a lactating mother, and two to homes of children under age 2. With 25 working days per month, workers are expected to complete 100 visits per month. Visits differ in frequency and purpose based on the age of the child, but they largely focus on health (e.g., vaccines), nutrition (e.g., feeding practices), and the caregivers' reproductive health (MWCD, 2022a).[3] Notably absent from these visits is a focus on child-rearing practices more broadly.

2.2 Automated phone calls for caregivers served by ICDS

We conducted the present study in the northern Indian state of Uttarakhand, which has a population of 10.1 million people and about 1.4 million children ages 0 to 6 (MHFW, 2020). Its neonatal mortality rates are among the highest in the country: 30 per 1,000 live births.

[1] Responsibilities include weighing each child at the center monthly and tracking their growth; teaching pre-school education for children ages 3 to 6; organizing supplementary nutrition for children ages 0 to 6; providing health and nutrition education and breastfeeding counseling to young mothers; assisting with immunization and health checkups; helping with iron fortification and vitamin A supplements; and maintaining registers (MWCD, 2022b).

[2] Since 2018, the central government contributes INR 4,500 per worker per month (≈USD 54; AI, 2021). States' top-ups vary widely, from no additional funds to over INR 7,000 (MWCD, 2019). Uttarakhand, the site of the present study, contributes a top-up of INR 3,000 for a total monthly honorarium of INR 7,500 (≈USD 90).

[3] For example, workers are expected to visit children the week after they were born at least twice, and more if they are newborns, to counsel the mother on feeding, warmth, cleanliness, and early disease detection. They are expected to visit 6-to-8-month-olds once a month to counsel on complementary feeding, for vaccination, and weighing them to determine if they need supplementary food at the *anganwadi* center (MWCD, 2022a).

One in three children is stunted (i.e., has a lower-than-expected height for their age), and one in four is underweight (i.e., weighs less than expected for their age; NIPCCD, 2018). Over 95% of four-year-olds are enrolled in some form of pre-school (mostly *anganwadis*). Yet, only three in 10 can count objects and four in 10 can compare quantities; one in two can describe a picture, and one in four can understand a story read to them (ASER, 2019).

To complement the services that caregivers of young children received through ICDS in Uttarakhand, we partnered with a non-profit (*Dost*, which means "friend") to evaluate its flagship program: automated phone calls with child-rearing advice for 6-to-30-month-olds.[4] Since its founding in 2017, *Dost* has offered this program to 100,000 caregivers in four states (typically, mothers without a high school degree living on less than USD 1,500 a year). Descriptive and experimental evidence suggested that the approach would be effective.[5] Further, the intervention could mitigate the effects of interruptions in service delivery, such as the one recently prompted by the COVID-19 pandemic (Angrist et al., 2022, 2023).[6]

Caregivers could sign up for the program with assistance from their *anganwadi* worker. Workers were given a script explaining the program and instructions to enroll (call a free phone number, indicate the child's age, and choose the time of day for the calls).[7] Caregivers were then randomized to receive the program during or after the study (see section 3.1). Those who were assigned to receive it were offered a total of 85 phone calls (of 1-2 minutes each) over 24 weeks, for an average of five phone calls per week.

The content of the calls was based on global evidence on effective child-rearing practices for children under age 3. It drew on multiple frameworks, including the Reach Up home-visitation program in Jamaica (Chang-Lopez et al., 2020), the United Nations Children's Education Fund program guidance for early childhood development (UNICEF, 2017), the World Health Organization's Nurturing Care for Early Childhood Development framework (WHO, 2018), the Center on the Developing Child at Harvard University's theory of change for adult capabilities to improve child outcomes (CDC, 2011), and a text-messaging program for parents evaluated by researchers at Stanford University (Cortes et al., 2021; Doss et al., 2019; York et al., 2019). It covered the themes in the National Council of Educational

[4] *Dost* has chosen to deliver these recordings through the phone because other forms of communication (e.g., WhatsApp) require smartphones, which are less prevalent among low-income households in LMICs. In 2021, only 52% of households with low levels of parental education in rural India had a smartphone (ASER, 2021). Further, more prevalent forms of communication (e.g., text messages) are often ignored (Beam et al., 2022).

[5] A third-party survey of *Dost*'s beneficiaries had found that 60% of those who signed up for the program became highly engaged users, 91% were more confident as parents, and 94% had more knowledge on how to manage their children's behavior (60 Decibels, 2021).

[6] Surveys of frontline workers across seven Indian states indicate several functions of *anganwadis* (e.g., growth monitoring, immunization, food supplementation) were interrupted during most of 2020, and centers varied widely in the extent to which they pursued adaptations (e.g., phone calls for counseling, Avula et al., 2022).

[7] Volunteers from *Dost* provided additional support whenever necessary.

Research and Training's handbook for early childhood care and education (NCERT, 2019), and it was adjusted based on over 1,000 interviews with parents to align it with local needs.

The recordings aimed to create awareness of child care rearing best practices to improve child development. They were organized into 18 modules: (a) the importance of early years of development; (b) embedding talk, care, and play into everyday life; (c) using art as a medium for learning; (d) setting up the home for learning; (e) managing screen time; (f) enabling learning through expeditions; (g) building an emotional bond; (h) creating an emotionally secure environment; (i) caring for parental well being; (j) managing difficult behavior; (k) narrating stories and having conversations; (l) supporting abilities through growth periods; (m) fostering deep and secure sibling bonds; (n) understanding nutritional relationships; (o) learning independence, empathy, and responsibility; (p) promoting physical development through play; (q) imparting experiential learning; and (r) a review of important concepts. Each module required participants to complete four calls at a day/time of their choosing.[8]

Each recording (e.g., managing conflict among siblings) follows the same four-part structure. It begins by introducing a challenge that caregivers may be facing and empathizes with them (e.g., siblings often fight with each other, even when caregivers wish that they did not). Then, it suggests some activities for addressing the challenge at hand in everyday life (e.g., how to discipline a child who is misbehaving without comparing them to their sibling).[9] Next, it reviews common strategies across activities (e.g., remembering each child is unique, focusing on praise, and knowing when to intervene or let siblings work things out). Lastly, it asks caregivers to check their understanding or provide feedback via a touch-tone response (e.g., asking caregivers to press 1 if they learned anything about managing sibling conflict).[10]

To sustain take-up, *Dost* conducted "live" (i.e., non-pre-recorded) calls to keep caregivers engaged with the intervention. These calls were made every Sunday to mothers who had not answered any calls in two weeks and who had not received a live call in the past month. Dost staff encouraged caregivers to listen to the calls and resolved issues that adversely affected take-up (e.g., changes in availability). No content was delivered through these calls.

[8]After the fourth call, some modules also offered additional information in optional calls. If participants did not complete them, they were then offered the required calls for the next module.

[9]These activities might be undifferentiated or specific to children below age 1, 1-year-olds, or 2-year-olds.

[10]These questions are asked for 75% of the calls, towards the end of each call, and they enquire about caregivers' actions (e.g., "do you share your childhood stories and lullabies with your child?") and beliefs (e.g., "do you think children can learn through play?").

3 Research methods

Our design and methods follow a registered pre-analysis plan.[11] This plan prespecified all analyses in the tables and figures in the main text of the paper.

3.1 Sampling, randomization, and implementation quality

Our sample consists of the 2,433 caregivers served by 250 *anganwadi* centers across two blocks (Khatima and Jaspur) of one district (Udham Singh Nagar) of Uttarakhand who signed up for the program.[12] In these centers, we recruited 2,433 caregivers with at least one child (ages 6 to 30 months). Our unit of analysis is the caregiver or their youngest child in that age range (depending on whether we focus on caregiver- or child-level outcomes).[13] We focused on the youngest child because we expected to see larger effects among younger children and we wanted to keep the time costs of the survey manageable for caregivers.

We randomly assigned caregivers to either receive the program during ("treatment group") or after the impact evaluation ("control group") within each center.[14] Each caregiver had an equal probability of being assigned to the treatment or control groups within their center. Following Banerjee et al. (2020), we randomly assigned caregivers to groups multiple times to ensure groups were comparable at baseline. Specifically, we ran our randomization 50 times and chose the assignment that minimized the difference in covariates between groups (this is known as the "minmax method"; see Bruhn and McKenzie, 2009).[15]

Table 1 presents summary statistics on children and caregivers and compares these baseline characteristics across treatment and control groups. Nearly half of the children in our study were born female, and the average child was 18 months old. The typical household in our sample is relatively disadvantaged: nearly all of them have a bathroom (94%) and a bedroom (99%), and most have a TV (70%), but only about half have books for children (56%) or adults (52%), and very few (4.4%) have an Internet connection. We find no systematic differences between the treatment and control groups in child- or caregiver-level outcomes.

The intervention was implemented largely as intended. As Figure 1 shows, by the end of the study, nearly all caregivers in the treatment group completed at least one call, eight in 10

[11] See `https://www.socialscienceregistry.org/trials/10129`.

[12] "Blocks" are administrative units below districts. In Uttarakhand, there are 13 districts and 95 blocks.

[13] In our sample, there are 14 mothers with twins in that age range. In these cases, we focus on the child the mother named first during her baseline interview.

[14] We believed "contamination" across experimental groups was highly unlikely. *Dost* chose to whom it delivered audio recordings and did not do so for the control group until the study ended.

[15] Covariates included baseline measures of caregivers' practices and anxiety, children's sex, age, and overall and language development, and the number of call attempts needed to complete the baseline survey.

completed at least five calls, and six in 10 completed at least 15 calls, indicating caregivers were exposed to a meaningful share of the content of the automated phone calls.

Perhaps unsurprisingly, given the demographics of beneficiaries (who are low-income and time-constrained), the ratio of calls made to calls accepted is large: for the average caregiver, *Dost* had to make roughly six calls to get a caregiver to complete one (see Table A1 in the online appendix). Yet, by the end of the study, the average caregiver had completed nearly 31 calls, and they had been exposed to almost 69 minutes of the program's material. The typical caregiver also remained engaged for most of the study: they went roughly 24 weeks with at least one accepted call and nearly 14 weeks with at least one completed call.

The modules were always presented in the same order, so caregivers were more likely to listen to earlier than later modules: nearly 98% of caregivers listened to the first module, but only 52% reached the last one (see Figure A1, panel A). Accordingly, while nearly all (97%) of caregivers listened to the first module focusing on child development (embedding talk, care, and play into everyday life), and most (72%) listened to the second one (narrating stories and having conversations), only about half (54%) listened to the third one (imparting experiential learning). Lastly, less than half of the caregivers listened to the modules in full: 98% of them listened to part of the first module, but only 47% listened to all of it (panel B).

3.2 Data and attrition

Our data-analytic sample includes caregivers (and their children) surveyed at both baseline and endline. As indicated in our pre-analysis plan, our primary outcomes are children's overall and language development, and our secondary outcomes are caregivers' knowledge of child development, self-efficacy, interactions with their children, and anxiety.

All outcomes were measured via phone surveys administered individually by enumerators at the Abdul Latif Jameel Poverty Action Lab's regional office for South Asia (J-PAL SA) in a baseline prior to randomization (November-December 2022) and in an endline 10 months later (September-October 2023). We used instruments that were previously administered in LMICs. Baseline and endline scores are standardized ($\mu = 0, \sigma = 1$) with respect to the control distribution. The online appendix provides more details on the instruments.[16]

The attrition rate for caregivers in the baseline sample was 22.5%. There were no differences in attrition rates (Table 1, panel D).

[16]We leveraged evidence on how to administer measures of caregiver and child outcomes reliably and validly over the phone, which has grown rapidly in light of the COVID-19 pandemic (Kopper and Sautmann, 2020). We also adhered to J-PAL SA's data-collection procedures, including high-frequency checks for electronic forms, spot checks, and accompaniments, and debriefs for enumerators (Glennerster, 2017; J-PAL, 2017).

3.3 Estimation

We estimate the intent-to-treat (ITT) effect of the offer of the intervention by fitting:

$$Y_{ic}^{t=1} = \alpha_c + \beta T_{ic} + \boldsymbol{\delta}' \boldsymbol{X}_{ic}^{t=0} + \epsilon_{ic} \tag{1}$$

where $Y_{ic}^{t=1}$ is the outcome of interest for caregiver or child i from center c at endline, T_{ic} is an indicator variable for random assignment to the treatment group, and $\boldsymbol{X}_{ic}^{t=0}$ is a vector of baseline covariates at baseline selected through a LASSO procedure, from $Y_{ic}^{t=0}$ (whenever available), the child's age (in months) and sex, the caregiver's highest level of education, whether the caregiver is an adolescent, an index of household assets, and the number of call attempts needed to complete the baseline survey. The α_c parameters are center (i.e., randomization strata) fixed effects. The coefficient of interest is β, which captures the average causal effect of the intervention.

We estimate equation (1) using ordinary least squares regressions. It is common practice to cluster standard errors at the treatment level in randomized evaluations (see Abadie et al., 2022). We do not use clustered standard errors because caregivers were individually randomized into experimental groups and we observe only one child per caregiver.[17]

We account for multiple hypothesis testing by pre-specifying children's overall development and language skills as our two primary outcomes. We report both unadjusted standard errors and q-values that control the false-discovery rate (FDR).[18] These q-values represent, for each corresponding unadjusted p-value, the minimum uncorrected p-value threshold for which that p-value would be in the set of all tests whose p-values are below the generated FDR-corrected p-value threshold, indicating findings that are likely to represent true effects rather than being false positives. We provide further details on our approach to multiple hypothesis testing in the online appendix.

4 Results

4.1 Average effects on caregiver outcomes

We find that the offer of the intervention had a *negative* effect on all caregiver-level outcomes that it sought to improve. As Table 2 indicates, it had a negative but statistically insignificant

[17] de Chaisemartin and Ramirez-Cuellar (2024) suggest making an exception for randomized trials with small randomization strata (such as pairwise randomized trials). Our study does not fall into this category of trials.

[18] Multiple hypothesis testing and advancements over "basic" FDR methods are an active area of research; we chose the step-up procedure developed by Simes (1986).

effect on caregivers' knowledge of child development (-0.11 SDs, $q = 0.11$) and their interactions with children (-0.01 SDs, and $q = 0.80$), and a negative and statistically significant effect on caregivers' self-efficacy (-0.11 SDs, $q = 0.06$).

Conversely, the intervention had a *positive* effect on the one caregiver outcome it sought to reduce. It had a positive and statistically significant effect on caregiver anxiety (0.10 SDs, $q = 0.08$). Together, these results suggest that the automated phone calls did not affect the intermediate outcomes as expected. In fact, they may have had unintended consequences.

4.2 Average effects on child outcomes

Perhaps unsurprisingly, given the pattern of results on caregiver outcomes, as Table 2 also shows, the offer of the intervention had a precisely estimated null effect on children's overall development. In fact, based on the 95% confidence interval, we can rule out effects on the pre-specified primary outcome of interest below -0.05 SDs and above 0.07 SDs.[19] It also had a small, positive, and statistically insignificant effect on children's language and vocabulary skills. These results indicate that the average child benefited little from the phone calls.

4.3 Heterogeneous effects on child outcomes

We did not find that the effect of the intervention varied by children's (age-adjusted) baseline development or sex assigned at birth (i.e., the two dimensions of heterogeneity that we had pre-specified). As Table 3 shows, the effects on overall development and language and vocabulary skills for children in the lowest quartile of baseline development were both negative (-0.09 and -0.03 SDs, respectively) and statistically insignificant. The estimates for girls are closer to zero (0.01 and -0.02 SDs) and statistically insignificant for both outcomes. Thus, the average effects do not mask evidence of heterogeneity for vulnerable sub-groups.

5 Discussion

Our study adds to the already mixed results of randomized evaluations of interventions seeking to improve child-rearing practices through the use of live or automated phone calls—either by themselves or in combination with other supports for caregivers—in LMICs. Of the three studies conducted before ours, Arteaga and Trias (2023) finds positive effects on both caregiver and child outcomes, Smith et al. (2023) finds positive effects on caregiver

[19]This is the only outcome for which we can calculate a 95% confidence interval because the statistical tests for all other outcomes are adjusted for multiple hypothesis testing.

outcomes, and Rafla et al. (2024) finds null effects for both caregiver and child-level outcomes. Our intervention fares worst, with negative effects for caregivers and null results for children.

Our study is not designed for us to ascertain why our intervention worked least well. In this section, we identify some potential reasons by comparing it to the one evaluated by Arteaga and Trias (2023) in Guatemala, which has been the most successful one to date. Some of these reasons are related to the Reach Up curriculum originally deployed through home visits in Jamaica, on which this and the Smith et al. (2023) interventions were based, and others are specific to the way in which the phone calls were carried out in Guatemala. Our purpose is twofold: we seek to understand how the intervention we evaluated could be improved and to highlight questions that we deem important for future research.

One reason for our unexpected results may be the pace at which caregivers were expected to complete the program. The program evaluated by Arteaga and Trias (2023) set the pace at which caregivers had to complete each module (they received five messages every two weeks), whereas the one that we evaluated allowed them to do so at their own pace. This feature of *Dost*'s program may provide caregivers with much-needed flexibility to fit the program around their busy schedules, but it may also allow them to go on multiple days without being exposed to the material, making it challenging to build on prior knowledge.

A second reason may be the extent to which program materials are differentiated by age. The program in Guatemala customized messages based on nine relatively narrow age bands (6-8 months, 9-11 months, 12-14 months, 15-17 months, 18-20 months, 21-23 months, 24-26 months, 27-29 months, and 30 months or older). The one in India offered the same material to all caregivers of children 0-36 months of age. This standardization may make the material easier for the organization to deliver, but it may result in some caregivers being exposed to material that is not particularly relevant to their child's developmental stage.

A third reason may have to do with the domains of child development covered in the calls. The intervention in Guatemala focused on cognitive, language, motor, and to a lesser extent, social-emotional development. The one in India covered a wide range of topics, with three of 19 modules focused on child development.[20] This approach may support caregivers in a wide array of practices, but such breadth may come at the expense of greater depth on child development.[21] This hypothesis is consistent with the null effects of the program in Jordan evaluated by Rafla et al. (2024), in which the domains focused on child development were offered in the later modules and were thus accessed only by relatively few beneficiaries.

[20]These are modules (b) embedding talk, care, and play into everyday life; (k) narrating stories and having conversations; and (q) imparting experiential learning.

[21]Relatedly, while most of the modules of the Guatemalan program emphasize reading, only a few of the ones in the Indian program do, thereby reducing the chances of impacting language and vocabulary skills.

One final reason may relate to the inclusion of activities and/or games in program materials. The Guatemalan program included, and clearly described, one activity or game for each phone call, providing the caregiver with specific recommendations on how to engage with the child. It also offered caregivers guidance on how to develop toys (e.g., dolls or puzzles) for each activity with locally available and affordable materials (e.g., cardboard or plastic bottles).[22] The *Dost* program does not necessarily include an activity or game in every call, and when it does, it presents them as suggestions and describes them at a relatively general level. This approach may result in some caregivers not pursuing such recommendations and/or in others doing so but not implementing the activities/games as intended.

6 Conclusion

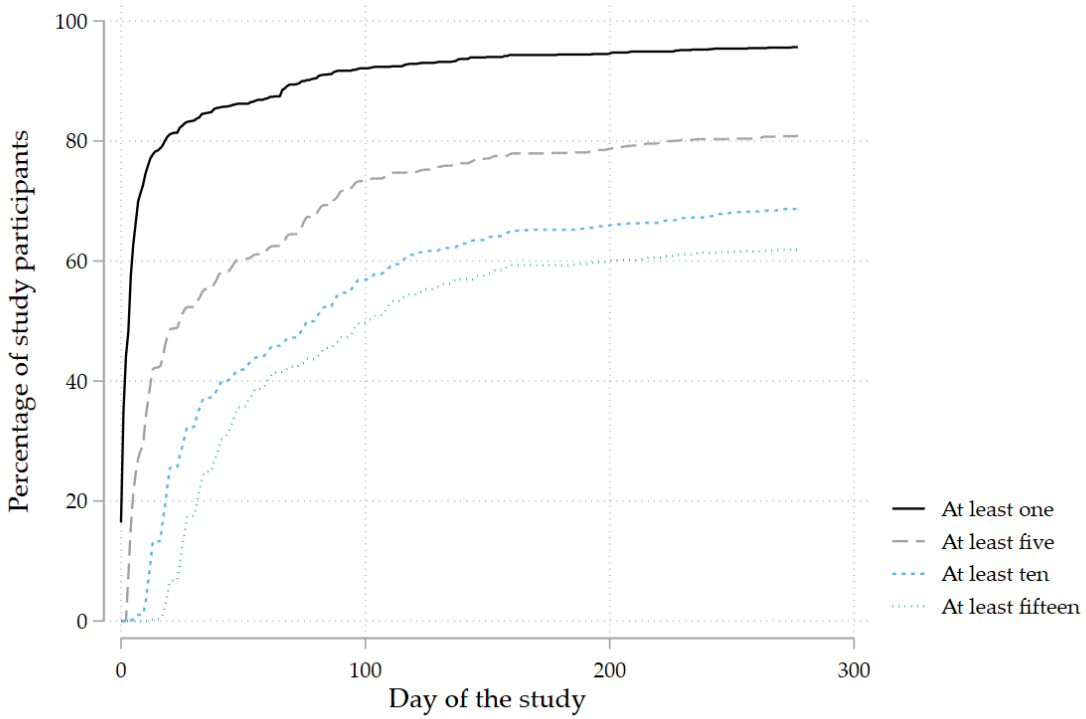
There is growing evidence that improving the quality of interactions children have with their caregivers can improve their school performance, achievement, and long-term outcomes. Yet, there is surprisingly little evidence on how to do so at scale in a cost-effective manner. Existing models rely on trained staff delivering programming in person.

We experimentally evaluated a promising alternative: automated phone calls with evidence-based child-rearing advice for caregivers who benefit from public pre-school centers (*anganwadis*) in the Indian state of Uttarakhand. These calls were meant to offer caregivers a steady source of such advice and complement the home visits conducted by *anganwadi* workers, which focus on health and nutrition and may be irregular due to the many tasks these workers are expected to complete. Unfortunately, however, the calls had mostly adverse effects on caregiver-level outcomes (knowledge of child development, interactions with their children, self-efficacy, and anxiety) and precisely estimated null effects on child-level outcomes (overall development and language and vocabulary skills).

We compare our program to a similar intervention evaluated in another low-income setting (Arteaga and Trias, 2023) and identify four possible reasons for its disappointing results, including: the flexible pace at which caregivers completed the program, the lack of differentiated activities by children's age, an insufficient focus on child development, and the absence of caregiver-child activities or games linked to each call. These hypotheses, however, would need to be empirically evaluated in future research, either through evaluations of adaptations of *Dost*'s intervention or of similar programs that incorporate more of the elements that we deem to be conducive to more positive effects.

[22] In fact, earlier calls in each module tell caregivers which materials they need to prepare toys in later calls.

Figure 1: *Intervention take-up*



Notes: This figure shows, across the study period, the percentage of study participants in the treatment group who completed one, five, ten, and fifteen calls, respectively.

Table 1: *Balancing checks between experimental groups*

	N	(1) Control Mean/SD	N	(2) Treatment Mean/SD	(3) Difference (1)-(2)
Panel A: Background characteristics					
Child is female	1214	0.498	1219	0.489	0.009
		[0.500]		[0.500]	(0.020)
Child age (in months)	1214	18.023	1219	17.808	0.215
		[6.894]		[6.762]	(0.279)
The following is in the home					
Bathroom	1214	0.941	1219	0.941	-0.000
		[0.236]		[0.236]	(0.009)
Bedroom	1214	0.987	1219	0.989	-0.002
		[0.114]		[0.107]	(0.004)
Living room	1214	0.502	1219	0.546	-0.044**
		[0.500]		[0.498]	(0.019)
Books for children	1214	0.556	1219	0.568	-0.012
		[0.497]		[0.496]	(0.020)
Books for adults	1214	0.518	1219	0.503	0.015
		[0.500]		[0.500]	(0.020)
Television	1214	0.704	1219	0.697	0.007
		[0.457]		[0.460]	(0.018)
Computer	1214	0.113	1219	0.103	0.009
		[0.317]		[0.305]	(0.012)
Internet connection	1214	0.044	1219	0.056	-0.011
		[0.206]		[0.230]	(0.009)
Asset index	1214	0.000	1219	0.027	-0.027
		[1.000]		[1.035]	(0.040)
Call attempts needed to complete the baseline	1214	2.452	1219	2.490	-0.038
		[1.727]		[1.752]	(0.068)
Panel B: Child outcomes					
Child development (CREDI score)	1214	0.000	1219	-0.003	0.003
		[1.000]		[0.942]	(0.039)
Language and vocabulary (CDI score)	1214	0.000	1219	-0.030	0.030
		[1.000]		[0.948]	(0.040)
Panel C: Caregiver outcomes					
Caregiver-child interactions (FCI-Play score)	1214	-0.000	1219	0.012	-0.012
		[1.000]		[0.992]	(0.039)
Caregiver anxiety (GAD-7 score)	1213	-0.000	1219	-0.015	0.015
		[1.000]		[0.991]	(0.040)
Panel D: Attrition					
Attrited	1214	0.217	1219	0.232	-0.015
		[0.413]		[0.422]	(0.017)

Notes. This table compares individuals in the control and treatment groups at baseline. It shows the mean and corresponding standard deviations for each variable (in brackets), and it compares both experimental groups, including randomization-strata fixed effects, showing the mean difference and corresponding standard errors (in parentheses). Except for child age and the number of call attempts needed to complete the survey, continuous variables are standardized and centered with respect to the control group. The asset index reflects an inverse-covariance-weighted (ICW) average across the eight yes/no questions about household assets. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.

Table 2: *Intent-to-treat effects*

	Caregiver outcomes				Child outcomes	
	Knowledge	Caregiver-child interactions	Anxiety	Self-efficacy	Child development	Language and vocabulary
	KIDI	FCI-Play	GAD-7	TOPSE	CREDI	CDI
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.11	-0.01	0.10	-0.11	0.01	0.05
	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)
	[0.11]	[0.80]	[0.08]	[0.06]	[0.80]	[0.49]
	1756	1886	1715	1723	1781	1838
R-squared	0.32	0.34	0.39	0.25	0.64	0.42
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports on ITT effects. All regressions include *anganwadi* center fixed effects and a vector of control variables selected via LASSO. All outcome variables are standardized and centered with respect to the control group at baseline. Standard errors in parentheses. Simes FDR q-values in brackets (following a pre-specified order of tests, across the main and heterogeneous impacts investigated in the study).

Table 3: *Heterogeneity in intent-to-treat effects*

	Lowest quartile		Girls	
	Child development CREDI	Language and vocabulary CDI	Child development CREDI	Language and vocabulary CDI
	(1)	(2)	(3)	(4)
Treatment	-0.09	-0.03	0.01	-0.02
	(0.06)	(0.08)	(0.05)	(0.06)
	[0.49]	[0.84]	[0.84]	[0.84]
	1781	1838	1782	1839
	0.64	0.43	0.52	0.41
Controls	Yes	Yes	Yes	Yes

Notes. This table reports on ITT effects for prespecified subgroups of children. All regressions include *anganwadi* center fixed effects and a vector of control variables selected via LASSO. Both outcome variables are standardized and centered with respect to the control group at baseline. "Lowest quartile" refers to the first quartile of baseline scores within CREDI age groups. "Girls" refers to female children. Standard errors in parentheses. Simes FDR q-values in brackets (following a pre-specified order of tests, across the main and heterogeneous impacts investigated in the study).

References

60 Decibels, 2021. Dost education: Impact study. New Delhi, India: 60 Decibels.

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M., 2022. When Should You Adjust Standard Errors for Clustering? The Quarterly Journal of Economics 138, 1–35. doi:10.1093/qje/qjac038.

Abufhele, A., Bravo, D., López Boo, F., Soto-Ramirez, P., 2021. Developmental losses in young children from pre-primary program closures during the COVID-19 pandemic. Washington, DC: Inter-American Development Bank (IDB).

AI, 2021. Integrated Child Development Services (ICDS), GoI, 2020-2021. *Budget Briefs*, 12(4). Delhi, India: Centre for Policy Research.

Andrew, A., Attanasio, O.P., Augsburg, B., Day, M., Grantham-McGregor, S.M., Meghir, C., Mehrin, F., Pahwa, S., Rubio-Codina, M., 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 61, 644–652.

Angrist, N., Ainomugisha, M., Bathena, S.P., Bergman, P., Crossley, C., Cullen, C., Letsomo, T., Matsheng, M., Panti, R.M., Sabarwal, S., Sullivan, T., 2023. Building resilient education systems: Evidence from large-scale randomized trials in five countries. (NBER Working Paper No. 31208). Cambridge, MA: National Bureau of Economic Research (NBER).

Angrist, N., Bergman, P., Matsheng, M., 2022. Experimental evidence on learning using low-tech when school is out. Nature Human Behaviour 6, 941–950.

Arteaga, I., Trias, J., 2023. Can technology narrow the early childhood stimulation gap in rural Guatemala? Results from an experimental approach. *Unpublished manuscript*. Washington, DC: The World Bank.

ASER, 2019. Annual status of education report (rural) 2018: Early years. New Delhi, India: ASER Centre.

ASER, 2021. Annual status of education report (rural) 2021. New Delhi, India: ASER Centre.

Attanasio, O.P., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., Rubio-Codina, M., 2018. Early stimulation and nutrition: The impacts of a scalable intervention. (NBER Working Paper No. 25059). Cambridge, MA: National Bureau of Economic Research (NBER).

Attanasio, O.P., Fernández, C., Fitzsimons, E.O.A., Grantham-McGregor, S.M., Meghir, C., Rubio-Codina, M., 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 349, g5785.

Avula, R., Nguyen, P.H., Ashok, S., Bajaj, S., Kachwaha, S., Pant, A., Walia, M., Singh, A., Paul, A., Singh, A., et al., 2022. Disruptions, restorations and adaptations to health and nutrition service delivery in multiple states across india over the course of the covid-19 pandemic in 2020: an observational study. PLoS One 17, e0269674.

Banerjee, A., Chassang, S., Montero, S., Snowberg, E., 2020. A Theory of Experimenters: Robustness, Randomization, and Balance. American Economic Review 110, 1206–1230. doi:10.1257/aer.20171634.

Beam, E., Mukherjee, P., Navarro-Sola, L., 2022. Lowering barriers to remote education: Experimental impacts on parental responses and learning. (IZA Discussion Paper No. 15596). Bonn, Germany: Institute for the Study of Labor (IZA).

Bruhn, M., McKenzie, D., 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. American Economic Journal: Applied Economics 1, 200–232. doi:10.1257/app.1.4.200.

CDC, 2011. Building adult capabilities to improve child outcomes: A theory of change. Cambridge, MA: Center on the Developing Child at Harvard University. URL: https://developingchild.harvard.edu/resources/building-adult-capabilities-to-improve-child-outcomes-a-theory-of-change/.

de Chaisemartin, C., Ramirez-Cuellar, J., 2024. At What Level Should One Cluster Standard Errors in Paired and Small-Strata Experiments? American Economic Journal: Applied Economics 16, 193–212. doi:10.1257/app.20210252.

Chang-Lopez, S., Walker, S., Grantham-McGregor, S., Powell, C., 2020. Parent manual: Activities for children up to age 3 years. Kingson, Jamaica: Caribbean Institute for Health Research, The University of West Indies.

CISCD, 2000. From neurons to neighborhoods: the science of child development. Committee on Integrating the Science of Child Development. Washington DC: National Academy Press.

Cortes, K.E., Fricke, H., Loeb, S., Song, D.S., York, B.N., 2021. Too little or too much? actionable advice in an early-childhood text messaging experiment. Education Finance and Policy 16, 209–232.

Doss, C., Fahle, E.M., Loeb, S., York, B.N., 2019. More than just a nudge: Supporting kindergarten parents with personalized and differentiated text messages. Journal of Human Resources 56, 567–603.

Engle, P.L., Black, M.M., Behrman, J.R., Cabral de Melho, M., Gertler, P.J., Kapiriri, L., Martorell, R., Young, M.E., the International Child Development Group, 2007. Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. The Lancet 369, 229–242.

Ganimian, A.J., Muralidharan, K., Walters, C.R., 2023. Improving early-childhood human development: Experimental evidence from India. Journal of Political Economy .

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S.M., Grantham-McGregor, S.M., 2014. Labor market returns to an early childhood stimulation intervention in Jamaica. Science 344, 998–1001.

Glennerster, R., 2017. The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency, in: Banerjee, A.V., Duflo, E. (Eds.), Handbook of Economic Field Experiments. Elsevier. volume 1, pp. 175–243. doi:10.1016/bs.hefe.2016.10.002.

González, M., Loose, T., Liz, M., Pérez, M., Rodríguez-Vinçon, J.I., Tomás-Llerena, C., Vásquez-Echeverría, A., 2022. School readiness losses during the covid-19 outbreak. a comparison of two cohorts of young children. Child Development 93, 910–924.

Grantham-McGregor, S.M., Adya, A., Attanasio, O.P., Augsburg, B., Behrman, J.R., Caeyers, B., Day, M., Jervis, P., Kochar, R., Makkar, P., Meghir, M., Phimister, A., Rubio-Codina, M., Vats, K., 2020. Group sessions or home visits for early childhood development in India: A cluster RCT. Pediatrics 146.

Grantham-McGregor, S.M., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., Group, I.C.D.S., et al., 2007. Developmental potential in the first 5 years for children in developing countries. The lancet 369, 60–70.

Grantham-McGregor, S.M., Powell, C.A., Walker, S.P., Himes, J.H., 1991. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. The Lancet 338, 1–5.

Hazarika, G., Viren, V., 2013. The effect of early childhood developmental program attendance on future school enrollment in rural north india. Economics of Education Review 34, 146–161.

J-PAL, 2017. J-PAL Research Protocols. URL: https://drive.google.com/file/d/0B97AuBEZpZ9zZDZZbV9abllqSFk/view.

Kopper, S., Sautmann, A., 2020. Best practices for conducting phone surveys. Abdul Latif Jameel Poverty Action Lab (J-PAL) 20.

Landry, S.H., Smith, K.E., Swank, P.R., 2006. Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. Developmental Psychology 42, 627–642.

Lu, C., Black, M.M., Richter, L.M., 2016. Risk of poor development in young children in low-income and middle-income countries: An estimation and analysis at the global, regional, and country level. The Lancet 4, e916–e922.

Magnuson, K.A., Yoo, P.Y., Duncan, G.J., Yoshikawa, H., Trang, K., Gennetian, L.A., Halperin-Meekin, S., Fox, N.A., Noble, K.G., 2022. Can a poverty reduction intervention reduce family stress and improve stress-related processes among families with infants? An experimental analysis. Madison, WI: Institute for Research on Poverty, University of Wisconsin at Madison.

Masarik, A.S., Conger, R.D., 2017. Stress and child development: A review of the family stress model. Current opinion in psychology 13, 85–90.

McCoy, D.C., Cuartas, J., Behrman, J.R., Cappa, C., Heymann, J., López Bóo, F., Lu, C., Raikes, A., Richter, L., Stein, A., Fink, G., 2021. Global estimates of the implications of COVID-19-related preprimary school closures for children's instructional access, development, learning, and economic wellbeing. Child Development 92, e883–e899.

MHFW, 2020. National family health survey-5 (2019-20). Key indicators from 22 states/UTs from phase I. Delhi, India: Government of India.

MWCD, 2015. Honorarium and appointment of anganwadi workers. New Delhi, India: Ministry of Women and Child Development, Government of India.

MWCD, 2019. Anganwadi sevikas. Delhi, India: Ministry of Women and Child Development. URL: https://bit.ly/3U28fLx. Last accessed: July 16, 2021.

MWCD, 2022a. Home visits planner. New Delhi, India: Ministry of Women and Child Development, Government of India.

MWCD, 2022b. Mission Saksham Anganwadi and Poshan 2.0: Scheme guidelines. New Delhi, India: Ministry of Women and Child Development, Government of India.

Nandi, A., Behrman, J.R., Laxminarayan, R., 2020. The impact of a national early childhood development program on future schooling attainment: Evidence from Integrated Child Development Services in India. Economic Development and Cultural Change 69, 291–316.

NCERT, 2019. Theme-based early childhood care and education programme. New Delhi, India: National Council of Educational Research and Training (NCERT).

NIPCCD, 2018. Handbook 2018: Statistics on children in India. New Delhi, India: National Institute of Public Cooperation and Child Development, Ministry of Women and Child Development, Government of India.

PEO, 2011. Evaluation report on Integrated Child Development Services. New Delhi, India: Programme Evaluation Organisation, Planning Commission, Government of India.

Psacharopoulos, G., Patrinos, H.A., 2004. Returns to investment in education: a further update. Education Economics 12, 111–134.

Rafla, J., Schwartz, K., Yoshikawa, H., Hilgendorf, D., Ramachandran, A., Khanji, M., Abu Seriah, R., Al Aabed, M., Fityan, R., Sloane, P. andAl Aqra, A., Sharawi, T., Molano, A., Foulds, K., Behrman, J.R., Wuermli, A., 2024. Cluster randomized controlled trial of a phone-based caregiver program for Syrian and Jordanian families with young children. New York, NY: Global TIES for Children, New York University (NYU).

Ravindran, S., 2020. Parental investments and early childhood development: Short and long run evidence from India. *Unpublished manuscript*. New York, NY: New York University.

Simes, R.J., 1986. An improved Bonferroni procedure for multiple tests of significance. Biometrika 73, 751–754. doi:10.1093/biomet/73.3.751.

Singh, A., Romero, M., Muralidharan, K., 2024. COVID-19 Learning loss and recovery: Panel data evidence from India. Journal of Human Resources doi:10.3368/jhr.0723-13025R2.

Smith, J.A., Chang, S.M., Brentani, A., Fink, G., Lopez-Boo, F., Torino, B.M., Codina, M.R., Walker, S.P., 2023. A remote parenting program and parent and staff perspectives: A randomized trial. Pediatrics 151, e2023060221F.

UNICEF, 2017. UNICEF's programme guidance for early childhood development. New York, NY: United Nations Children's Education Fund (UNICEF).

UNICEF, 2023. The state of the world's children 2023: For every child, vaccination. Florence, Italy: Global Office of Research and Foresight, UNICEF.

Walker, S.P., Chang, S.M., Powell, C.A., Grantham-McGregor, S.M., 2005. Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted jamaican children: prospective cohort study. The lancet 366, 1804–1807.

Walker, S.P., Wachs, T.D., Gardner, J.M., Lozoff, B., Wasserman, G.A., Pollitt, E., Carter, J.A., Group, I.C.D.S., et al., 2007. Child development: Risk factors for adverse outcomes in developing countries. The Lancet 369, 145–157.

WHO, 2018. Nurturing care for early childhood development: a framework for helping children survive and thrive to transform health and human potential. Geneva, Switzerland: World Health Organization (WHO), United Nations Children's Fund (UNICEF), and the World Bank.

York, B.N., Loeb, S., Doss, C., 2019. One step at a time: The effects of an early literacy text-messaging program for parents of preschoolers. Journal of Human Resources 54, 537–566.

Yoshikawa, H., Wuermli, A.J., Britto, P.R., Dreyer, B., Leckman, J.F., Lye, S.J., Ponguta, L.A., Richter, L.M., Stein, A., 2020. Effects of the global coronavirus disease-2019 pandemic on early childhood development: short-and long-term risks and mitigating program and policy actions. The Journal of Pediatrics 223, 188–193.

Young, M., 2002. From early child development to human development. Washington, DC: The World Bank.

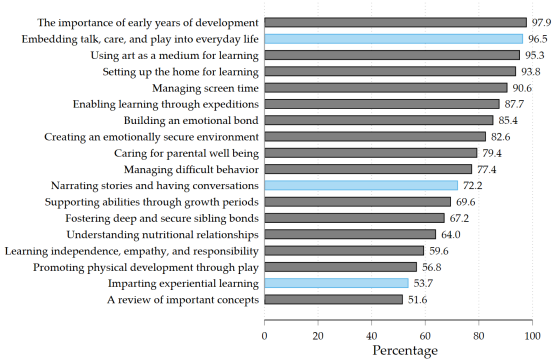
Yousafzai, A.K., Rasheed, M.A., Rizvi, A., Armstrong, R., Bhutta, Z.A., 2014. Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: A cluster-randomised factorial effectiveness trial. The Lancet 384, 1282–1293.

Online Appendix

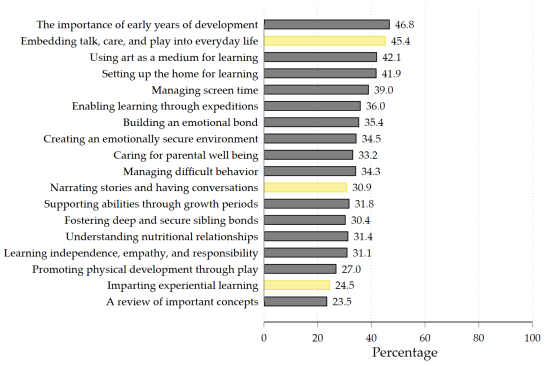
A Additional figures and tables

Figure A1: *Caregivers' completion of program modules*

(a) *At least partially listened to*



(b) *Fully listened to*



Notes. This figure presents the percentage of caregivers in the treatment group who completed each of the program's 18 modules. Subfigure (a) shows the percentage of treatment-group caregivers who at least partially listened to at least one call mapped to a given module; subfigure (b) shows the percentage of treatment-group caregivers who fully listened to at least one call mapped to a given module. Color highlights three modules focused on psycho-social stimulation.

Table A1: *Caregiver exposure to calls*

Indicator	Mean	Standard deviation
No. of calls made to caregiver	409.5	221.86
No. of calls accepted	68.05	29.95
No. of calls fully completed	30.56	26.3
Minutes on calls	68.66	52.43
No. of weeks with at least one accepted call	22.56	8.92
No. of weeks with at least one completed call	13.53	9.09

Notes. This table reports on treatment-group caregivers' overall exposure to calls during the study period.

B Measurement and estimation

B.1 Caregivers' knowledge of child development

We measured caregivers' knowledge of child development using an adaptation of the Knowledge of Infant Development Inventory (KIDI-SF; MacPhee, 1981) at endline. The KIDI-SF contains 20 statements and asks each caregiver whether she agrees with a statement about child development. For example, one item is "If you punish children for doing something naughty, it is okay to give them a piece of candy to stop the crying." Caregivers can respond by indicating agreement, disagreement, or stating that they are not sure.

This instrument had already been administered in India (Karuppannan et al., 2020). Given that it was originally developed in the 1970s, we included new statements related to screen time in children ages two and younger following recommendations from the American Academy of Pediatrics. We estimated each caregiver's score using item response theory and a two-parameter logistic model.

B.2 Caregivers' interactions with their children

We measured caregivers' interactions with their children with the play sub-scale of the Family Care Indicator (FCI; Hamadani et al., 2010) at baseline and endline. The FCI-Play contains six items and asks each caregiver whether they or the child has engaged in specific activities during the week prior to the survey. For example, one question is "Have you told stories to the child last week?" Caregivers can respond affirmatively or negatively. The FCI has been administered in many LMICs, including India (Grantham-McGregor et al., 2020; Luoto et al., 2021).

The play sub-scale includes six items that have been previously administered on their own (Arteaga and Trias, 2023; Tofail et al., 2013; Babikako et al., 2022; Knauer et al., 2016). We estimated each caregiver's score using item response theory and a two-parameter logistic model.

B.3 Caregivers' anxiety

We measured caregivers' anxiety using the General Anxiety Disorder (GAD-7; Löwe et al., 2008) at baseline and endline. The GAD-7 asks each caregiver whether they have been bothered by a set of feelings during the two weeks prior to the survey. For example, one question asks "how often have you felt nervous, anxious, or on edge?" Caregivers can respond using a four-point scale, from 1 ("not at all") to 4 ("nearly every day").

The GAD-7 contains seven questions and has already been administered in India (De Man et al., 2021). We estimated each caregiver's score using item-response theory and a generalized partial credit model.

B.4 Caregivers' self-efficacy

We measured caregivers' self-efficacy using selected items from the Tools of Parents Self-Efficacy (TOPSE; Kendall and Bloomfield, 2005) at endline. The TOPSE asks each caregiver whether they agree with a statement about their perceived capacity to engage in a given parenting behavior. For example, one item is "I can recognize when my child is happy or sad." Caregivers can respond by choosing a number from 0 (indicating that they completely disagree with the statement) to 10 (indicating that they completely agree with the statement).

Following List et al. (2021), we used items for six of the eight sub-scales, which seem to be more relevant. Using results from a validation study in Bangladesh (Ferdowshi et al., 2021), we used the two items with the highest item-total correlation from each sub-scale to construct a short version of the instrument.[23] We estimated each caregiver's score using item response theory and a generalized partial credit model.

B.5 Children's overall development

We measured children's overall development with the Caregiver Reported Early Childhood Development (CREDI; McCoy et al., 2017) at baseline and endline. The CREDI asks each caregiver whether their child can do something that they ought to be able to do, given their age. For example, for children ages 6 to 11 months, one question is "can the child pick up a small object (e.g., a small toy or stone) using just one hand?" Caregivers can respond affirmatively, negatively, or by stating that they do not know.

We used the short form of the CREDI, which produces a single score of overall child development. This form was validated with more than 8,000 children across 17 LMICs, including India (Alderman et al., 2021; McCoy et al., 2018; Waldman et al., 2021). Specifically, the short form contains 20 items that vary by six-month age brackets (i.e., 6-11 months, 12-17 months, etc.). We estimated each child's score using item response theory and a two-parameter logistic model.

B.6 Children's language and vocabulary skills

We measured children's language and vocabulary skills with an adapted version of 50 words or sentences of the MacArthur-Bates Communicative Development Inventory (CDI; Fenson et al., 2000; Jackson-Maldonado et al., 2013) at baseline and endline. The CDI asks each caregiver whether their child can understand and/or state a word or sentence. For example, for children ages 8 to 17 months, one question is "can the child understand and/or say *uh-oh*?" Caregivers can respond by indicating whether the child can understand the prompt, understand and state it, or cannot do either.

Following other adaptations (e.g., Kern, 2007; Floccia et al., 2018), we translated the short form of the English CDI to Hindi and consulted with native speakers to ensure that the list

[23] We thank Sally Kendall for encouraging us to pursue this strategy.

of words presented to caregivers is culturally relevant. We adapted the form for children in three age groups: 6-17, 18-30, and 31-37 months. We estimated each child's score using item response theory and a two-parameter logistic model.

C Empirical analysis

C.1 Rules for handling missing values

We expected to encounter two types of missing data: attrition (i.e., caregivers not participating in the endline) or missing values (i.e., caregivers participating in the endline, but not answering specific questions therein).

We address the first type of missingness as follows. First, we document the overall attrition rate. Then, we investigate whether attrition is systematically related to intervention assignment by fitting a version of equation (1) that replaces the outcome variable with an indicator variable for not participating in the endline.

We address the second type as follows. For missing responses on outcomes, we scale responses using item-response theory models that account for missing values by using concurrent calibration via marginal maximum likelihood estimation (Kolen and Brennan, 2004), given that non-response on specific questions is akin to missingness in any non-equivalent anchor test design in which not all respondents are administered the same questions.

C.2 Multiple hypothesis testing

The CREDI is a "global" measure of early childhood development that also captures a child's language development—therefore, in the group comparison for the CREDI, we do not apply an adjustment for multiple hypothesis testing. In contrast, the group comparison for the CDI is a second test, which warrants such an adjustment.

Next, we hypothesized caregivers' interactions with their children would serve as the most important intermediate outcome and tested for impacts on the FCI play subscale (test number three).

Our analyses of effects among subgroups focused on two additional group comparisons. Starting with the CREDI, we assessed program impacts among the quartile of less-developed children (test four) and impacts among girls (test five).

After that, we explored (average) program impacts on caregivers' beliefs (test six), anxiety levels (test seven), and self-efficacy (test eight).

Lastly, we repeated the above subgroup analyses for the CDI (tests nine and ten). We deprioritized additional subgroup analyses (e.g., group comparisons in the top quartile of baseline child development, among boys, etc.).

We prioritized the study's statistical tests in this order. For example, we accounted for three tests in our analyses of impacts on the FCI play subscale, four tests in the group comparison

of the CREDI among the quartile of less-developed children, five tests in the respective group comparison among girls, etc.

References

Alderman, H., Friedman, J., Ganga, P., Kak, M., Rubio-Codina, M., 2021. Assessing the performance of the Caregiver Reported Early Development Instruments (CREDI) in rural India. Annals of the New York Academy of Sciences 1492, 58–72. doi:10.1111/nyas.14543.

Arteaga, I., Trias, J., 2023. Can technology narrow the early childhood stimulation gap in rural Guatemala? Results from an experimental approach. *Unpublished manuscript*. Washington, DC: The World Bank.

Babikako, H.M., Bourdon, C., Mbale, E., Aber, P., Birabwa, A., Chimoyo, J., Voskuijl, W., Kazi, Z., Massara, P., Mukisa, J., et al., 2022. Neurodevelopment and recovery from wasting. Pediatrics 150.

De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., Johnson, L.C., Thankappan, K.R., Williams, E.D., 2021. Are the PHQ-9 and GAD-7 suitable for use in India? a psychometric analysis. Frontiers in psychology 12, 676398.

Fenson, L., Pethick, S., Renda, C., Cox, J.L., Dale, P.S., Reznick, J.S., 2000. Short-form versions of the macarthur communicative development inventories. Applied psycholinguistics 21, 95–116.

Ferdowshi, N., Imran, M.A., Trishna, T.A., 2021. Adaptation of the tool to measure parenting self-efficacy (TOPSE) in Bangladesh. Dhaka University Journal of Biological Sciences 30, 169–177. doi:10.3329/dujbs.v30i2.54643.

Floccia, C., Sambrook, T., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., et al., 2018. Vocabulary of 2-year-olds learning english and an additional language: norms and effects of linguistic distance. v: General discussion. Monographs of the Society for Research in Child Development 83, 68–80.

Grantham-McGregor, S.M., Adya, A., Attanasio, O.P., Augsburg, B., Behrman, J.R., Caeyers, B., Day, M., Jervis, P., Kochar, R., Makkar, P., Meghir, M., Phimister, A., Rubio-Codina, M., Vats, K., 2020. Group sessions or home visits for early childhood development in India: A cluster RCT. Pediatrics 146.

Hamadani, J.D., Tofail, F., Hilaly, A., Huda, S.N., Engle, P., Grantham-McGregor, S.M., 2010. Use of family care indicators and their relationship with child development in bangladesh. Journal of health, population, and nutrition 28, 23.

Jackson-Maldonado, D., Marchman, V.A., Fernald, L.C., 2013. Short-form versions of the spanish macarthur–bates communicative development inventories. Applied Psycholinguistics 34, 837–868.

Karuppannan, A., Ramamoorthy, T., Rammamoorthi, A., Ravichandran, L., 2020. Mother's knowledge on child's developmental milestones and parenting skills in kanchipuram district, tamilnadu: a descriptive cross sectional study. Int J Health Sci Res [Internet] 10, 242–7.

Kendall, S., Bloomfield, L., 2005. Developing and validating a tool to measure parenting self-efficacy. Journal of advanced nursing 51, 174–181.

Kern, S., 2007. Lexicon development in french-speaking infants. First Language 27, 227–250.

Knauer, H.A., Kagawa, R.M., Garcia-Guerra, A., Schnaas, L., Neufeld, L.M., Fernald, L.C., 2016. Pathways to improved development for children living in poverty: A randomized effectiveness trial in rural mexico. International Journal of Behavioral Development 40, 492–499.

Kolen, M.J., Brennan, R.L., 2004. Test Equating, Scaling, and Linking. 3rd ed., Springer, New York, NY.

List, J.A., Pernaudet, J., Suskind, D.L., 2021. Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. Nature communications 12, 5765.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y., 2008. Validation and standardization of the generalized anxiety disorder screener (gad-7) in the general population. Medical care , 266–274.

Luoto, J.E., Garcia, I.L., Aboud, F.E., Singla, D.R., Fernald, L.C., Pitchik, H.O., Saya, U.Y., Otieno, R., Alu, E., 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, e309–e319.

MacPhee, D., 1981. Knowledge of infant development inventory: Manual. Chapel Hill, NC: Department of Psychology, University of North Carolina .

McCoy, D.C., Sudfeld, C.R., Bellinger, D.C., Muhihi, A., Ashery, G., Weary, T.E., Fawzi, W., Fink, G., 2017. Development and validation of an early childhood development scale for use in low-resourced settings. Population health metrics 15, 1–18.

McCoy, D.C., Waldman, M., Team, C.F., Fink, G., 2018. Measuring early childhood development at a global scale: Evidence from the caregiver-reported early development instruments. Early childhood research quarterly 45, 58–68.

Tofail, F., Hamadani, J.D., Mehrin, F., Ridout, D.A., Huda, S.N., Grantham-McGregor, S.M., 2013. Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-Deficient Children. The Journal of Nutrition 143, 885–893. doi:10.3945/jn.112.160473.

Waldman, M., McCoy, D.C., Seiden, J., Cuartas, J., Fink, G., 2021. Validation of motor, cognitive, language, and socio-emotional subscales using the Caregiver Reported Early Development Instruments: An application of multidimensional item factor analysis. International Journal of Behavioral Development 45, 368–377. doi:10.1177/01650254211005560.