IFS

Institute for Fiscal Studies

Bet Caeyers
Lina Cardona-Sosa
Sarah Cattan
Sonya Krutikova
Abu Siddique

Working paper

Parents in the classroom: strengthening government capacity to deliver early childhood education

26/11

UKRI Economic and Social Research Council

Parents in the Classroom: Strengthening Government Capacity to Deliver Early Childhood Education*

Bet Caeyers[†] Lina Cardona-Sosa[‡] Sarah Cattan[§]

Sonya Krutikova[¶] Abu Siddique[‖]

June 03, 2026

Abstract

Preschool enrollment has expanded rapidly in low-income countries, but improvements in quality have lagged behind. This paper evaluates a scalable model that engages parents to improve early learning in contexts where both schools and homes are constrained. In collaboration with Ghana's Ministry of Education, we study the *Lively Minds* program in rural Ghana that trains mothers from the community to deliver structured, play-based learning activities in public preschools and foster learning and development of their own children at home. Using a cluster randomized controlled trial, we find that the program increases children's cognitive development by 0.11 standard deviations and reduces problem behaviors, particularly among boys. The reductions in problem behaviors are driven by children of participating mothers, consistent with significant improvements in mothers' knowledge of child development and increases in developmentally supportive interactions at home. This pattern highlights the importance of engaging parents rather than relying exclusively on classroom-based inputs delivered by teachers or paraprofessionals. Notably, participation in the program has no adverse effects on mothers' well-being, and we show evidence that implementation quality and mothers' participation have been sustained as the program scaled to roughly one-third of rural districts nationwide in response to the findings of this evaluation. Overall, the results highlight parents as an effective and underutilized resource in poor rural settings for supporting early skill formation beyond the first 1,000 days and strengthening government capacity to deliver quality early childhood education.

JEL Codes: J13, I10, I20

Keywords: early childhood development, preschool quality, parenting, scalability.

*We are very grateful to *Lively Minds* NGO for their collaboration and support. We would like to acknowledge the contributions of Peter Leighton at an early stage of this project and of Salifu Amadu, Mubarik Ahmed, and the Innovations of Poverty Action team in Ghana during the data collection. We are grateful to Prof Elizabeth Spelke for sharing child assessment tools that she developed. We acknowledge contributions by Orazio Attanasio during an earlier stage of this project. We would also like to thank Juanna Joensen, Costas Meghir, Imran Rasul, and seminar and conference participants at the IFS, Royal Holloway, London Behavioural and Experimental Economics Workshop, Lancaster University Economics Workshop, and Research Workshop on Family and Health Economics for many comments and suggestions. This research was funded by the Global Innovations Fund, the Jacobs Foundation, and the ESRC. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing up results. The study protocol was approved by the Ghana Health Service Ethics Review Committee (ref: GHSERC012/07/17) and IRBs at the Innovations for Poverty Action (ref: 14340) and the University College London (ref: 10167/001) in 2017. The trial was registered on ISRCTN (ISRCTN21215509) and AEA RCT Registry (AEARCTR-0002777) in March 2018.

[†]Chr. Michelsen Institute; IFS

[‡]World Bank

[§]IFS; CESifo; HCEO

[¶]University of Manchester; IFS

[‖]Royal Holloway University of London; IFS; IZA@LISER

1 Introduction

Early childhood investments have very high private and social returns, and early skill deficits are difficult and costly to remediate later in life (Heckman and Mosso, 2014; Currie and Almond, 2011; Heckman, 2012; Attanasio et al., 2022; Andrew et al., 2024). Yet governments in low- or lower-middle-income countries (LLMIC) often lack the resources and administrative capacity to deliver high-quality early childhood education at scale. Over the past two decades, enrollment in pre-primary education has increased substantially in LLMIC: from 13% in 2000 to over 30% in 2018 (UNICEF, 2019b). In many settings, however, such expansion has outpaced improvements in service quality: constrained budgets, weak implementation systems, and limited teacher capacity mean that children can spend years in school without mastering basic competencies (World Bank, 2018a). Therefore, scaling preschool access risks generating enrollment without learning, as classrooms become overcrowded and teachers remain poorly trained and under-resourced. At the same time, low levels of parental education and literacy mean that the main caregivers of young children, typically mothers, are often limited in capacity to provide cognitively rich inputs at home. Therefore, the main challenge for many LLMIC is how to improve early learning when both schools and homes are severely constrained.

We study this question in the context of Ghana's public preschool system. Ghana mandated two years of publicly provided pre-primary education in 2008 and now has one of the highest preschool enrollment rates in the region, serving close to 1.3 million children (Ministry of Education, Republic of Ghana, 2025). Despite near-universal enrollment, significant quality challenges persist, particularly in rural and high-poverty areas, including shortages of trained teachers, high teacher turnover and absenteeism, large class sizes, weak pedagogical practices, and limited funding (The World Bank, 2024). Many children also grow up in home environments that are not conducive to early learning, reflecting low parental education and time constraints, limited cognitive stimulation, and widespread use of harsh discipline (UNICEF, 2019a; Lim et al., 2023). In our baseline survey of 2,151 rural preschool-aged children, only 13% of caregivers reported any adult–child play in the previous three days. These constraints are reflected in low learning outcomes, with fewer than half of preschool-aged children developmentally on track in literacy and numeracy, and only about half of Primary 4 pupils meeting minimum competency benchmarks in English and mathematics (Lim et al., 2023; National Council for

Curriculum and Assessment (NaCCA), 2023).

Against this background, we test a novel and scalable approach to improving children's learning and development within this highly constrained setting, being implemented by the Ghana Education Service (GES).[1] It is predicated on the idea that parents of young children are a key underutilized resource in the provision of quality early childhood care and education (ECCE) in LLMICs, particularly in rural communities, where needs are greatest. Accordingly, the Ghanaian Ministry of Education, working through the GES, is implementing a program developed by the NGO Lively Minds that trains and embeds mothers as regular volunteers within public preschools to deliver structured, play-based learning activities for preschool-age children and foster the learning and development of their own children at home.

We study the impact of this approach through a cluster-randomized controlled trial (RCT) conducted across a sample of 80 preschools in two districts in rural Northern Ghana, where the program was not yet operational. Half of the preschools were randomly selected for treatment, while the other half served as the control group and continued with standard preschool provision. In treatment preschools, the GES, with support from Lively Minds, recruited and trained mothers from the communities served by the preschools to deliver play-based learning activities during instructional hours. Interested mothers completed a training course consisting of eight two-hour training sessions led by preschool teachers over three weeks, after which they were assigned to facilitate small-group play sessions (with activities such as counting, sorting, and puzzles) for up to two hours per week over 30 weeks of the school year. In addition to classroom involvement, participating mothers attended monthly parenting and well-being workshops run by the same preschool teachers.

Because the program is designed to change both what happens inside preschools and what participating mothers do at home, we frame its impacts on children's skills formation through two complementary channels: the preschool environment and the home environment. In the preschool, the program directly aims to improve the quality of the learning environment by introducing structured, play-based activities delivered in small groups by trained community mothers. This represents a direct investment in preschool-based early childhood education, which may improve cognitive skills through increased instructional time, more individualized

[1] The Ghana Education Service (GES) is the implementing agency of the Ministry of Education responsible for the delivery of public pre-primary, primary, and secondary education.

attention, and exposure to play-based curricula (Grantham-McGregor et al., 1991; Currie, 2001; Hirsh-Pasek et al., 2009; Cascio and Schanzenbach, 2013). Structured, adult-supervised play and routines may also reduce problem behaviors among children by strengthening self-regulation, such as turn-taking and inhibitory control (Diamond and Lee, 2011; Healey and Healey, 2019). Beyond these direct classroom effects, a novel feature of this program is that it uses mothers as the marginal input in preschools, so implementation itself becomes an investment in parenting human capital with potential spillovers to the home environment. The training is designed not only to equip mothers to support learning in the classroom, but also to strengthen caregiving and learning interactions at home. Through training and repeated classroom practice, mothers acquire pedagogical knowledge, skills, and habits that can be applied both in preschool and in the household, where they shape parenting inputs and investments in the child. This engagement may update maternal information and beliefs about returns to investment during early childhood (Cunha et al., 2022; Dizon-Ross, 2019). Therefore, if parental participation increases cognitively rich interactions and supportive parenting at home, children's skills should also improve through the home environment channel.

Our primary outcome of interest is children's cognitive and socioemotional development. Cognitive development is measured through direct one-to-one assessment by independent assessors and captures foundational literacy and numeracy skills, as well as executive functions. Socioemotional development is measured through direct assessment of emotional awareness as well as maternal report of problem behaviors. Our design also allows us to estimate the impacts of the program on the knowledge and practices of the participating mothers and the development of their children. To do so, identical mobilization and sign-up processes for participating in the program were conducted in the treatment and control communities, with the exception that in the control communities mothers were asked to sign up for participation in the following year (after the RCT), when the program was planned to be rolled out there. Comparing mothers who expressed interest in participating and their children in the treatment preschools to those who expressed interest in participating in the following year in the control preschools allows us to isolate the causal effects of intention to participate.

First, we find that early-childhood development (ECD) knowledge and skills of parents with minimal education and prior engagement in child development can be significantly improved by

this relatively low-intensity program implemented within a highly constrained education system. In particular, we find that participating mothers' knowledge of child development increases by nearly half of a standard deviation (SD). They also report substantially more engagement in play and learning activities with their child at home. There is some indication from enumerator observations of mother-child play that the quality of mothers' engagement with their children also improved. These effects are distinct from the broader effects of the program on parents in the community. This group primarily updated beliefs on easily observable aspects of preschool quality, not broader knowledge about child development. Non-participating mothers also do not increase the quality or quantity of engagement with their child and even show some evidence of reduced investments in play materials at home, perhaps perceiving these as less critical in light of improvements in quality at preschools.

Second, we find that following the training, engaging mothers in running small-group play-schemes during instructional hours in public preschools leads to economically meaningful gains in cognitive skills of the children enrolled in these preschools at baseline. Overall, there is a 0.11 SD increase in the aggregate cognitive development index, with improvements concentrated in emergent numeracy and executive function domains. The gains are broad-based across gender and are robust to alternative score construction that reduces measurement error. These average gains mask important distributional patterns: the program raises the probability that children score above the 25^{th}, 50^{th}, and 75^{th} percentiles of the control-group cognitive distribution, with little movement at the extreme tails.

The program also reduced behavioral difficulties among children. There is a marginally significant reduction of nearly 10% of a SD overall, which is within the range of average effects of 0.1-0.15 reported in a meta-analysis of effects of preschool interventions on socioemotional skills (Holla et al., 2021). This effect is driven by a reduction specifically in externalizing behaviors (disruptive hyperactive behaviors) among those with a high problem load. It is concentrated among boys, who, in line with findings in the wider literature, are also more likely to exhibit externalizing behaviors. These impacts align well with the program's emphasis on structured, adult-supervised small-group activities and routines, shown in the literature to be an effective approach to reducing externalizing behavioral difficulties.

Third, we find evidence that mobilizing parents as the program implementers generates

additional benefits for the children of participating mothers, beyond the effects of improving classroom activities alone. Comparing children exposed to the play schemes whose mothers did not participate with children exposed to the same play schemes whose mothers did participate, we find that while cognitive gains are similar across the two groups, socioemotional improvements are concentrated among children whose mothers participated. We also find suggestive, though statistically imprecise, evidence of positive spillovers on the development of younger siblings of children whose mothers participate in the program.

The program is relatively low-cost, requiring $37 per child (in implementation-year USD) in additional funding beyond existing education-system resources used to deliver the program. This includes one-off set up costs with significantly lower ongoing additional costs following set up. Following the evidence generated by this study on the effectiveness of the program, Government of Ghana scaled it to 60 districts. At this much larger scale additional funding required for implementation went down significantly as the program became more integrated within core public preschool provision. Additional funds required during the set-up phase at scale were $17 per child while running costs went as low as $3 per child. Key to keeping additional costs low is reliance on mothers volunteering their time and existing school infrastructure and staff rather than additional salaried staff or material inputs. However, reliance on the existing workforce and volunteers to deliver the program raises concerns about potential hidden costs. Participating mothers are predominantly illiterate women living in poor rural areas who tend to be disempowered and marginalized (see, for instance, Malapit and Quisumbing (2015)). Like many rural women in low-income settings, they face substantial demands on their time from income-generating activities, farm work, and domestic responsibilities (Dinkelman and Ngai, 2022), making additional time demands potentially costly. Such reliance on low-status, poorly paid, overstretched female frontline workers and mothers is a common feature of ECD programs. In fact, there is a growing concern that the global push to expand early childhood provision may have unintended adverse consequences on the welfare of these already vulnerable groups (Ballard et al., 2023; Sullivan et al., 2024). Our analysis offers encouraging evidence that this is not a concern with this program. We find no adverse effects of the program on the mental health and self-esteem of the participating mothers. This is consistent with the program's modest time demands on the participating mothers, about 1-2 hours per week, and emphasizing their

well-being in the parenting workshops.

In sum, our findings show that a government-led model that mobilizes mothers as a low-cost input into public preschools can raise early human capital in settings where both school and home inputs are severely constrained. The program leads to gains in children's cognitive development, while further improvements in socioemotional outcomes are concentrated among children whose mothers participate in running the program, consistent with observed program-induced improvements in the ECD knowledge and practices of these mothers. These meaningful gains are achieved at relatively low cost and with no detectable reductions in the psychosocial well-being of participating mothers.

Related literature. We contribute to two strands of literature. First and foremost, we add to a nascent evidence base on scalable, cost-effective government-led approaches to improving early childhood education. While Ghana is a trailblazer, having been the first to provide universal access to public preschool in the region nearly 20 years ago, the challenges it faces with ensuring adequate quality are far from unique. We have seen these problems on a mass scale during the expansion of primary schooling (World Bank, 2018b) and are starting to see them at the preschool level as more countries join Ghana in expanding preschool provision (Blimpo et al., 2022; Bouguen et al., 2018). In all likelihood, as this continues, a growing number of countries will face the same challenges and require viable approaches to addressing them. The need for evidence on such approaches is all the more pressing as shrinking donor funding, rising debt, and growing student numbers are resulting in significant contractions of per-pupil education budgets across Sub-Saharan Africa, including Ghana (World Bank and UNESCO, 2024).

Existing evidence suggests that there are some promising approaches, but it is dominated by programs operated by nongovernmental organizations (e.g., see Dean and Jayachandran (2019); Martinez et al. (2017)). Two very recent studies suggest that teaching assistant (TA) models can be an effective approach to improving the quality of government early childhood education provision, though accompanied by adequate training for the teachers to ensure that the TAs are deployed appropriately in the classroom (Ganimian et al., 2024; Andrew et al., 2024). In line with these studies, we also find that bringing additional support into the classroom can be an effective way of improving children's outcomes in government preschools. However, we provide evidence on the potential for low educated often illiterate parents, who are easily dismissed

as unlikely educators, rather than additional salaried staff, to provide this support in a way that is not detrimental to their welfare. In terms of magnitudes, our 0.11 SD gain on an aggregate cognitive index is comparable to the ITT effect reported in a recent TA intervention (Ganimian et al., 2024), which finds a 0.11 SD gain on a composite learning measure in household assessments. Our effects are also closer in magnitude to the estimated impacts from extensive-margin expansions in preschool access (Bassi et al., 2025, 0.16 SD gain on a skills index). The latter benchmark is informative because it suggests that, in low-capacity systems, relatively low-cost quality improvements within existing preschools (e.g., the *Lively Minds* program) can generate learning gains comparable to those achieved by large-scale expansions in preschool access.

By focusing on parents, our study also contributes to the literature on parenting interventions. There is a large literature showing that parenting programs can be effective at improving parenting and child development for very young children (in the first 1,000 days) (Jeong et al., 2021; Carneiro et al., 2024). However, these programs tend to involve much more intensive training of the implementing workforce as well as parents themselves than the Lively-Minds model. Moreover, low-educated parents in settings such as ours may struggle more to support the development of early learning skills among preschool-age children than the development of children earlier on. Attempts to engage parents in the development of preschool-age children in poor, low-education settings have so far yielded mixed results. For example, Angrist et al. (2025) evaluates an NGO-led program providing one-to-one training to parents of preschool-age children in a low-education setting but finds no impacts on parental knowledge or behavior. In comparison, Özler et al. (2018) offers more encouraging evidence on a group-based parenting program for parents of preschool-age children. Our study adds to this evidence, demonstrating not only the feasibility of improving ECD knowledge and practices among parents of preschool-age children in poor, rural, low-education settings, but also doing so in a government-led way within the existing public education infrastructure.

Structure. The paper is organized as follows: Section 2 describes the Ghanaian context and the L*ively Minds* program in detail. Section 3 outlines the experimental design, sampling, data, and empirical strategy. Section 4 presents the main results. Section 5 provides evidence on the program's opportunity costs. Section 6 assesses program costs, documents progress with the

ongoing scale-up of the program across around 40% of rural districts in Ghana and concludes. All tables and figures are provided at the end.

2 Setting and intervention

2.1 Ghana Education System

Ghana introduced two years of publicly provided pre-primary education in 2008 and now has one of the highest preschool enrollment rates in West Africa, serving close to 1.3 million children nationwide (Ministry of Education, Republic of Ghana, 2025). Despite near-universal access, quality remains uneven, particularly in rural and high-poverty areas, where schools face persistent challenges including shortages of trained teachers, high turnover and absenteeism, large class sizes, and limited instructional resources (The World Bank, 2024). At the same time, many children grow up in home environments that provide limited cognitive stimulation, reflecting low parental education, time constraints, and widespread use of harsh discipline (UNICEF, 2019a; Lim et al., 2023). These constraints are reflected in weak early learning outcomes: fewer than half of preschool-aged children are developmentally on track in literacy and numeracy, and only about half of Primary 4 pupils meet minimum competency benchmarks in English and mathematics (Lim et al., 2023; National Council for Curriculum and Assessment (NaCCA), 2023).

The Government of Ghana has pursued a series of policy reforms aimed at strengthening early childhood education. For example, the Early Childhood Education Policy Framework published in 2021 provides a systematic, government-led structure for planning, curriculum implementation, teacher preparation, and quality assurance in KG services (**?**). Nevertheless, there remains a need for scalable, evidence-based interventions that can operationalize these policy frameworks. To date, much of the available evidence on effective ECD programs in contexts comparable to Ghana has come from small-scale initiatives implemented by non-governmental organizations outside of routine government systems.

Implementation of basic education in Ghana is administratively decentralized. While national policy and curriculum are set by the Ministry of Education and national agencies, day-to-day management and delivery of preschool services are largely the responsibility of district Ghana Education Service (GES) offices. District directorates oversee teacher deployment, su-

pervision, and school monitoring; organize in-service training; and are responsible for ensuring compliance with national guidelines. As a result, interventions to improve the quality of public preschools need to engage closely with the capacity, incentives, and implementation practices of district-level GES teams.

2.2 The *Lively Minds* Program

Against this backdrop, we evaluate the *Lively Minds* program, which is explicitly designed to strengthen the capacity of public preschools within Ghana's existing education system. Rather than operating as a parallel NGO-led model, the program is implemented through district GES structures, working with district teams, preschool teachers, and local communities to improve the quality of play-based learning in routine public preschool provision.

Program description. The *Lively Minds* program, developed by the Lively Minds NGO, seeks to improve ECD outcomes of pre-school children by mobilizing and training women in the community to run weekly, play-based learning activities in classrooms alongside teachers and to improve their ECCE practices at home. We refer to these women as "participating mothers" (PMs) since they tend to be mothers of preschool-age children. During each Play Scheme session, five mothers each facilitate a simple indoor "play station" (covering matching/sorting, numeracy, sizes/colours/senses, books, and building) using discovery-based activities in small groups of no more than five children. Preschool teachers supervise the sessions, while additional children participate in structured outdoor play led by another mother. The intended intensity is one indoor session per child per week, with outdoor play two to three times per week. In exchange for their time, participating mothers are offered monthly parenting and well-being workshops delivered by the preschool teachers.

Implementation model. The program is embedded within the Ghana Education Service (GES) and implemented at the district level. District GES teams are first oriented to the program, after which two teachers, the headteacher and the PTA chair from each public preschool in the district receive six days of training on play-based pedagogy, classroom organization, and how to train participating mothers. Teachers then organize a community meeting in which they introduce the program and invite women from the community to enroll. The women who

express interest then have to complete a training course consisting of eight sessions, two hours each, over the course of three weeks, led by the teachers. This training course is designed for low-literacy participants. Typically, 30–40 mothers are trained from the communities served by each participating preschool. Following the training, PMs are divided into four groups, and each group is given a different day on which they teach at the pre-school for one hour (with extra time for preparation and debrief). During sessions run by the PMs in the preschools, 25 children are arranged in small groups (maximum of 5) and rotate around the following 5 indoor play stations: matching/sorting; numeracy; sizes, colors, senses; books; building. One PM runs each play station, and they teach using discovery-based teaching methods, rather than rote learning, which is the norm in formal education settings in Ghana. The remaining KG children play outdoor games, led by PMs. The KG teachers supervise the sessions.

In exchange for their time running the Play Schemes and to reinforce new behaviors, PMs are offered monthly parenting workshops by the preschool teachers following a curriculum developed by Lively Minds NGO on various topics related to child development and their own well-being.[2]

GES officials monitor the Play Schemes as part of their normal supervisory functions, and Lively Minds NGO conducts some additional unannounced monitoring visits. GES officials and Lively Minds NGO staff have monthly meetings to track the progress of the Play Schemes and identify corrective measures. There are monthly "top-up" training workshops for preschool teachers, delivered by district GES staff, where they discuss problems, share successes, and are trained to provide the PMs with the monthly Parenting Workshop.

Innovation. The *Lively Minds* model introduces a classroom experience that is markedly different from business-as-usual in rural Ghanaian pre-schools by combining small-group instruction, simple educational games, and classroom organization around rotating play stations. The program's design explicitly addresses common barriers to parental engagement: training is deliberately simple, highly practical, and demonstration-based (e.g., through role play), avoiding complex parenting curricula. A central aim is to build mothers' confidence in their ability to support children's learning despite limited formal education. The time burden is kept intentionally low—each participating mother volunteers for about 1-2 hours per week, rotating

[2]Topics include nutrition, hygiene, child rights, play, communication, malaria prevention, financial awareness, self-esteem, and inclusive education.

across four groups on different days; participation is encouraged through free monthly parenting and well-being workshops rather than financial payments. Finally, both the training and play curriculum were developed with and in rural communities of Northern Ghana and refined through nearly a decade of prior implementation, ensuring that they are context-appropriate and feasible in predominantly rural settings where women often have flexible farm-based work schedules.

3 Study Design, Data, and Empirical Strategy

3.1 Experimental Design, Timeline, and Sample

In order to evaluate the *Lively Minds* program, we conducted an RCT between 2017 and 2019. The randomization was conducted at the school level: 80 schools in 2 districts of Northern Ghana were randomized into two equally sized control and treatment groups. The selected districts had not yet introduced the *Lively Minds* program in their preschools. As described in the previous section, the business-as-usual implementation approach for this program is to implement it in all preschools once the district enrolls in the program. In order to ensure adequate power, implementation in the RCT districts was restricted to the schools randomly selected for treatment.

The RCT districts, Bongo and Tolon, are located in the Upper East and Northern regions, respectively. At the time of this study, these were among the most deprived areas of Ghana, with very high rates of poverty —50.4% in the Northern Region and 44.4% in the Upper East, compared to a national poverty rate of 16.5% — and some of the slowest long-run progress in poverty reduction since the early 1990s (Cooke et al., 2016). Consistent with these patterns, household survey evidence shows low levels of education, especially among women, in northern Ghana, with a high share of women reporting no formal schooling and low literacy (Ghana Statistical Service and UNICEF, 2018). The two study regions differ along several dimensions, including language, religion, culture, and school quality, providing an opportunity to test the program in a heterogeneous set of communities that reflect the diversity of rural Ghana.

The randomization was conducted using two levels of stratification - circuit and school size.[3]

[3]A circuit is a geographical cluster of around 10 schools that falls under one supervisor from the Ghana Education Service (GES.

The final sample includes 38 schools in the Bongo District, and 42 Schools in the Tolon District. Among the 40 schools allocated to the treatment group, 21 schools were located in Tolon and 19 schools in Bongo. The population of interest is children starting pre-school in the academic year 2017/18. In order to ensure that baseline data collection was completed before the start of intervention activities, sampling and baseline activities had to start before the beginning of the school year in July 2017. This meant that it was not possible to use school enrollment lists to identify the population of interest and do the sampling. We therefore conducted a census of households located in the vicinity of the sampled schools to identify those with children due to start pre-school. The households closest were enumerated first, gradually moving to those further away to reach 150 households.[4] Of the 12,000 households listed in this way, a total of 4,486 eligible households were identified. We then drew a random sample of 2,407 children (approximately 30 per school) for the study.

Baseline data was collected over the course of three months between September and November 2017. The baseline survey included direct assessment of all of the children in the sample by trained interviewers, interviews with the primary caregiver of each of these children, the main KG teacher in each of the 80 KGs, as well as members of the community leadership team from the main community that each school served. Details of the baseline instruments can be found in (Amadu et al., 2018).

Endline data collection took place between September and November 2018, at the beginning of the next school year (2018/19) and following a two-month break in the Play Schemes and parenting workshops over the summer school holidays. Endline was completed before the program was rolled out to control schools in November 2018. We aimed to re-interview all of the children and primary caregivers in the baseline sample, including those who had moved to a different community since the baseline. We were able to collect complete endline data for 2,151 of the 2,407 children in our baseline sample, yielding an attrition rate of 11%, which is uncorrelated with treatment status (see Table B.1 and Table B.2.)

Table B.3 and Table B.4 present key baseline characteristics of primary carers and target children, respectively, among the sample of non-attritors by treatment status. The sample is well-balanced across key household, caregiver and child characteristics.

[4]See Amadu et al. (2018) for further details.

3.2 Ethics

We received ethical clearance for the project from the Ghana Health Service Ethics Review Committee (ref: GHSERC012/07/17), Innovations for Poverty Action (ref: 14340), and University College London IRB (ref: 10167/001).

3.3 Implementation and Compliance

All intervention activities took place between September 2017 and July 2018. Initial mobilization and introduction of district-level GES officials to the program took place in September 2017. The district team then implemented training of pre-school teachers, headteachers and PTA leads in treatment pre-schools. The teachers conducted recruitment and training of participating mothers (PMs) in October 2017.

A distinguishing feature of the Lively Minds program is its reliance on parents rather than teaching assistants, making it important to assess effects on both children and participating mothers. Because participation was voluntary, we need a comparison group for participating mothers who would have volunteered if given the chance in the control group. We identify these women by replicating mobilization activities in control communities. In treatment areas, women were invited to participate in the current academic year, while in control areas, they were invited to sign up for the following year, when the program would be rolled out. Our main study sample includes 437 children whose mothers either signed up to run the play-schemes in the treatment schools (263) or signed up to run the play-schemes in the following year in the control group (174).We use these two groups to estimate the Intent to Treat impact of Lively Minds program on PMs. Note that while we refer to these women as "Participating Mothers", they are actually mothers who signed up to participate during the initial mobilization stage. A small proportion (34 out of 263) of these mothers in the treatment group then do not report having participated in the program, suggesting that there may be some drop-out between expression of interest and start of the program. It is important that these 34 mothers are included in our analysis for comparability with the control group since we do not know which of the mothers who expressed interest in the control group would have dropped out before the start of the program.

The key assumption for identification of program impacts on (potential) PMs through comparison of the women who signed up to participate in the treatment and control schools is that

selection into signing up is the same in treatment and control preschools, even though in the latter sign-up was for a later start. We view this assumption as plausible for three reasons. First, the mobilization and sign-up activities were conducted by the same team in both treatment and control communities using an identical process, with the only difference being the timing of program rollout. Second, because children spend two years in pre-school and our sample consists of children entering their first year of KG, mothers in both treatment and control communities had a similar opportunity to participate while their own child was enrolled. Third, and most importantly, we verify that potential PMs in treatment and potential PMs in control are well balanced on baseline characteristics: Table B.5 and Table B.6 show no systematic differences across a wide range of household, caregiver, and child characteristics within this PM sample. We therefore treat this sub-sample as a credible comparison group for estimating impacts on PMs.

Play Schemes started to run in treatment schools in November 2017 and continued during term time until the end of the school year in July 2018. The Play Schemes were designed to run 4 times per week in each school during school term with different children and mothers attending each session. Each child would participate in indoor games at least once per week and in outdoor play 2 or 3 times per week. Over the course of the academic year, the Play Schemes ran over three terms: first term 1st of November 2017 - 10th of December 2017 (6 weeks), second term 15th of January 2018 - 15th of April 2018 (13 weeks) and, the last 14th of May 2018 - 24th of July 2018 (11 weeks). At full implementation, the Play Schemes would have operated for 30 weeks in total, providing each child with approximately 30 hours of participation over the academic year.

Implementation data collected by supervisors from the GES and Lively Minds NGO teams during random visits to preschools in the treatment group suggest that compliance was high. The Play Schemes were found to be running in 80% of all monitoring visits. The main reasons for the Play Scheme not running at a particular visit were bad weather, community funerals, other school/community event or low attendance of PMs, the latter being the most frequent reason recorded by the monitoring team. Using this as a proxy of the actual proportion of Play Schemes that were running suggests that children were exposed to an average of 23-25 hours of Play Schemes during the school year. On most monitoring visits, there were 2 teachers

present, an average of 8 PMs, and 24 children. In most cases mats were organized as expected, and in more than half of the cases, the practice of discovery-based teaching was observed. In addition, each school received an average of between 4 and 5 monitoring visits at the monthly parent workshops organised by the preschool teachers as part of the program over the course of the academic year. In 88% of the visits, the workshop was taking place. In most cases, both teachers were present with an average attendance of 30 mothers per session.

Teacher data collected as part of the endline survey provides additional evidence on compliance. Table 1 shows that teachers in treatment preschools were almost four times more likely than control teachers to agree with the statement that parents actively contribute to preschool activities (Column 1), 38 percentage points more likely to report having received active parental support during the academic year (Column 2), and 29 percentage points more likely to report that toys, games, or other equipment were accessible to children during free play on a typical preschool day (Column 3).

3.4 Outcomes

We pre-specified child development in the cognitive and socioemotional domains as the primary outcome (see AEA RCT registry no. AEARCTR-0002777).

3.4.1 Child development

At both baseline and endline we administered the International Development and Early Learning Assessment (IDELA) to measure cognitive and socio-emotional development, alongside selected items from the Strengths and Difficulties Questionnaire (SDQ) to capture behavioral difficulties (Goodman, 1997). The IDELA was administered directly to child by trained interviewers and covers emergent literacy, emergent numeracy, executive functioning (EF), and emotional awareness. The SDQ was administered to the child's primary caregiver to elicit reports of behavioural difficulties (Hoosen et al., 2018a). At endline, we supplemented these measures with a set of child tasks developed at the psychology Laboratory of Development Studies at Harvard and piloted in Ghana (Coffey and Spelke, 2024). These tasks were added to increase the discriminatory power of our measures while mapping onto the same core domains captured by the IDELA and, like the IDELA, were administered directly to children by trained enumerators. The scores

used in the main analysis are constructed following developer-prescribed scoring rules. Baseline scores of these measures are standardized to have mean=0 and standard deviation=1. Endline scores are standardized relative to the control group distribution. Appendix A provides further details about the assessments, construction of scores for the main analysis and robustness to alternative ways of scoring.

To study potential intra-household spillovers, we pre-specified developmental outcomes of the target child's siblings as secondary outcomes and collected child development data for one older and one younger sibling of each target child, where applicable. Younger siblings' development was captured using the Caregiver-Reported Early Development Index (CREDI) short form, a caregiver-report instrument designed for population-level measurement of early development from birth to age three in low-resource settings; scores were constructed following the developers' age-standardization guidance and then standardized relative to the control group (McCoy et al., 2017). Older siblings completed direct, one-to-one assessments consisting of a curriculum-aligned literacy and numeracy test developed for the Ghanaian context within the Ministry of Education, complemented by selected tasks from the battery developed by Harvard Laboratory of Development Studies and used with the target child. See Amadu et al. (2020) for further details.

3.4.2 ECD Knowledge, Practices and the Home Environment

The aim of *Lively Minds* is to improve child development through improving ECD care and education skills of parents and engaging them more actively in supporting the development of young children in their communities. Our pre-specified set of secondary outcomes, therefore, includes a rich set of measures capturing different dimensions of ECD knowledge and practices among parents of children in the sample collected through interviews of the primary caregivers of the children in our sample. First, knowledge about ECD is assessed using a short subset of the Knowledge of Infant Development Inventory (KIDI), a widely used instrument that elicits caregivers' beliefs about children's developmental capabilities and appropriate expectations (MacPhee, 1981a). Second, we measure caregivers' understanding of pre-school quality using a contextually tailored tool in which parents were shown illustrated vignette pairs depicting contrasting classroom practices and asked to select the option more conducive to child devel-

opment.

We capture ECD practices by combining caregiver reports with direct observation. Our first measure of ECD practices was collected through observing how mothers interacted with their own children during structured play tasks and coding a range of supportive and less supportive instructional behaviors, drawing on items from the UNICEF Multiple Indicator Cluster Surveys (MICS) (Bornstein et al., 2010). We also capture ECD behavior through measuring material and time inputs into the child using the Family Care Indicators (FCI), which records the availability of play and learning materials in the home and recent play and learning activities with the child (Hamadani et al., 2010). Baseline total scores for these measures are standardized to have mean=0 and standard deviation=1 while endline scores are standardized relative to the control group distribution.

3.5 Empirical strategy

The experimental design allows us to study the causal impact of *being eligible* for the *Lively Minds* program (intent-to-treat effect – ITT). In other words, we compare outcomes of children who were reported during the pre-baseline census as planning to attend one of the "treatment" pre-schools to those reported as planning to attend one of the "control" pre-schools. In line with the pre-analysis plan, we condition the outcomes at endline on a set of observable variables (including the outcome at baseline) in order to improve the precision of our estimates. Formally, we estimate the ITT effect of the *Lively Minds* program on the outcomes of interest by estimating the following regression:

$$Y_{is,1} = \alpha + \beta Treat_{s,0} + \sigma Y_{is,0} + \gamma X_{is,0} + \theta Strata_s + \delta District_s + \epsilon_{is,1} \tag{3.1}$$

where $Y_{is,1}$ is the outcome of interest for child i intending to go to pre-school s measured at endline; $Treat_{s,0}$ is a dummy equal to 1 if the pre-school s received the *Lively Minds* program; $Y_{is,0}$ is the same outcome measured at baseline[5]; as pre-specified $X_{is,0}$ is a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline; $Strata_s$ is a fixed effect for the randomization

[5]There are some outcomes for which there is no exact corresponding baseline measure. In these cases we add controls that are likely to constitute good proxies for a direct baseline measure of the outcome.

strata of school s; $District_s$ is an indicator for whether the school is located in Bongo or Tolon district; finally $\epsilon_{is,1}$ is the random error term, clustered at the school level (our unit of randomization).

We estimate equation (3.1) by OLS so that β is the estimated average ITT impact of *Lively Minds* on outcome $Y_{is,1}$. Given the number of outcomes examined across cognitive subdomains, socioemotional subdomains, and parental inputs, we account for multiple inference using Romano-Wolf stepdown adjusted p-values within pre-specified outcome families (Romano and Wolf, 2005). We report the unadjusted p-values in all of the tables and the adjusted p-values wherever they are relevant.

4 Results

4.1 Knowledge and practices of parents

The approach developed by *Lively Minds* is that it mobilizes low-educated, often illiterate, mothers as classroom facilitators. The key assumption underlying this model is that it is possible to train the mothers for this role through a short (8 sessions) group-based training course led by preschool teachers, supplemented by a monthly parenting workshop.

We, therefore, start by showing estimates of impacts of the *Lively Minds* program on (i) mothers' ECD knowledge, (ii) investments in child development at home, and (iii) quality of interaction with children. See Section 3 for details on how these are measured. As discussed in Section 3, we use outcome measures using raw scores in our estimates and present estimates constructed using IRT in Appendix B. We present impacts separately for mothers who sign up to participate in the program and those who do not, but whose children were registered as planning to go to a participating preschool during the baseline. We present impacts on this latter group to capture any indirect impacts of the program on mothers in the community. Differences in impacts and p-values for tests of statistical significance on these differences are shown at the bottom of the table.

ECD Knowledge. Table 2, shows clear improvements in knowledge about child development among PMs as a result of the program: a large and statistically significant treatment effect on PMs' knowledge about child development, which increases by just under half of a SD (column

1, $p = 0.012$). There is also some suggestive evidence that the knowledge about what makes a good pre-school improved among PMs, but it is not significant at conventional significance levels (column 2, $p = 0.537$).

As discussed in Section 2, the approach developed by *Lively Minds* differs markedly from business-as-usual methods in Ghanaian public preschool by including small-group play, an emphasis on simple educational games with props, and classroom organization designed to support children's learning. It is, thus, plausible that even parents who are not directly participating in the program update their beliefs about child development and preschool quality. We see some evidence of this among non-PMs in Table 2. While, in contrast to the PMs, we observe no effect on knowledge of non-PMs about child development (column 1, $p = 0.866$), there is a significant increase of 0.14 SD (column 2, $p = 0.096$) in knowledge about preschool quality among non-PMs. This effect size for non-PMs is also similar in magnitude to that for PMs (column 2, Panel C, $p = 0.980$), though, due to the much smaller sample size for the PM group, it is not significant for them. These patterns are consistent with differences in exposure to program content by PM status. Non-PMs appear to update beliefs that map to more easily observable changes in what happens at preschools as a result of the program, while PMs also acquire better overall knowledge on child development.

ECD practices. Better knowledge about ECD is important because it can lead to changes in parents' behavior that are beneficial for child development (Cunha et al., 2022; Carneiro et al., 2024). We now examine the impacts of *Lively Minds* on these behaviors. While the training, parenting workshops, and experience running the Play Schemes may have influenced the ECD practices of PMs, their low levels of education and the relatively light-touch nature of the intervention make the impact less certain.

Results in columns 3-4 of Panel A in Table 2 show evidence of improvements in ECD practices among PMs. There is a significant increase in the frequency of play and engagement in other developmentally supportive activities with children of PMs by adults in the households of the PMs. Among the PMs in the control group, only 36% reported that anyone in the households had conducted any play activities with the sampled child in the last 3 days. Among PMs in the treatment group, this proportion increases by 50% (18 p.p.) (column 4, $p = 0.014$). Effects are concentrated on the time rather than the material investment margin: there is no evidence of

any change in investment in play materials for the children (column 3, $p = 0.786$).

The large and significant increase in play activities without an accompanying increase in material investments aligns well with the adoption of low-cost, routine-based ways of engaging children that were emphasized through the training and parenting workshops; PMs may have implemented what they learned on embedding learning opportunities into everyday household chores. Disaggregating the results by specific play activities in Table B.10, we see increases in reading to the child, telling them stories, playing, and drawing with them. We do not see impacts on singing, going out, or playing with objects.

Results in columns 5-6 Table 2 further provide suggestive evidence of an improvement in the quality of PMs' engagement with children. This was measured through enumerators' observations of the quality of PMs' interactions with their own children (see Section 3 for details on this measure). We see an increase of 0.20 SD in positive parenting behaviors (supportive, responsive, and developmentally appropriate), but it is not statistically significant, especially once we adjust for multiple hypothesis testing (column 5, Panel A, $p = 0.467$).

In light of evidence of changes in PM parenting behavior as well as impacts on non-PMs' beliefs about pre-school quality, it is possible that the *Lively Minds* program also had indirect impacts on the parenting of non-PMs. Columns 3-6 in Table 2 for non-PMs show some suggestive evidence of this, though in a way that is very different from the PMs. Specifically, there is no impact on the time spent with children of non-PMs or the quality of engagement with children, in contrast to the large positive effects among PMs; the differences between treatment effects on time-investment outcomes for PMs and non-PMs are statistically significant (column 4, 'Difference in treatment' row). There is, however, some suggestive evidence of a *reduction* in investment in play materials at home (column 3), which we do not see for the PMs. Among non-PMs, investment in play materials at home decreased by close to a fifth of a standard deviation, though this result is not statistically significant once we adjust for multiple hypothesis testing ($p = 0.170$). This reduction is driven by negative impacts of the program on several items, including household objects used as toys, books, and toys to help with moving (Table B.9). The combination of the effects on non-PMs suggests that they were aware of improvements that happened in the preschools because of Lively Minds and responded by reducing investments at home. It is plausible that non-PM households viewed the preschool program as a substitute

for home inputs, reducing investments in play materials because they perceived that children's developmental needs were being met at school.

4.2 Child development

Cognitive development. In line with evidence of the success of *Lively Minds* in training mothers to strengthen their ECD knowledge and practices, along with high compliance in the delivery of Play Schemes in the classrooms by the trained mothers, reported in Section 3, we find that the program produced economically meaningful gains in children's cognitive development. Table 3 shows an ITT treatment effect of 0.11 standard deviations (SD; column 1, $p = 0.049$) on the aggregate measure of cognitive development, which combines emergent numeracy, emergent literacy, and executive functions. Disaggregating by subdomain, the gains are concentrated in emergent numeracy (column 3, $p = 0.038$) and executive functions (column 4, $p = 0.038$).

The 'Cognition' panel of Table 4 further shows that the program was not equally effective at all points in the cognitive skill distribution. It significantly increased the probability of getting a score above the 25^{th}, 50^{th}, and 75^{th} percentile of the control group distribution, with effects getting bigger at higher levels of skills, but had no impacts at the tails—on the probability of getting a score above the 10^{th} or the 90^{th} percentile. There is some indication that the distributional pattern of treatment effects is different from this for the executive functioning domain: intervention was most effective at increasing executive functions among those at the lower end of the executive functions distribution.

The program, however, was equally effective at improving the cognitive skills of boys and girls (column 1, Panel B of Table 3). Estimated treatment effects are similar in magnitude for boys ('boy' row, $p = 0.086$) and girls ('girl' row, $p = 0.044$), and we cannot reject equality of effects across genders ('difference' row, $p = 0.743$). It was also equally effective among children of PMs and non-PMs (Table 5). This result suggests that the benefits of the program on children's cognitive skills are driven by changes that take place in the pre-schools as a result of the program—likely the introduction of the Play-Schemes—rather than any changes in the home learning environment of the children of PMs.

Socioemotional development. Table 6 reports effects of the *Lively Minds* program on children's socioemotional development. A positive coefficient in column 1 would indicate an im-

provement in emotional awareness, and negative coefficients in columns 2–4 would imply a reduction in behavioral difficulties.

We find little evidence of program impact on emotional awareness (column 1, $p = 0.560$) and across the emotional awareness score distribution (see Table 4). This finding holds equally for boys and girls (see Panel B of Table 6, column 1). There is, however, some suggestive evidence that the program reduced children's behavioral difficulties. Column 2 in Table 6 shows an effect of -0.098 SD, although the estimate falls just short of conventional significance levels (column 2, $p = 0.102$). Distributional analysis of the treatment effects suggests that the program significantly improved problem behavior among children with above-average problem load at the 75^{th} and 90^{th} percentiles of the outcome distribution (Table 4).

The reduction in behavioral difficulties appears to be driven by a reduction in externalizing behaviors (column 3, 0.11 SD reduction, $p = 0.116$), with significant impacts at the top half of the externalizing problems distribution (Table 4). These reductions are consistent with the program's emphasis on structured, adult-supervised small-group play, which promotes rule-following and turn-taking. This is in line with evidence that play-based activities can strengthen children's self-regulation skills (Diamond and Lee, 2011; Healey and Healey, 2019), and with a broader literature showing that weaknesses in self-regulation are closely linked to externalizing difficulties such as hyperactivity and inattention (Sher-Censor et al., 2016; Coelho et al., 2023).

Analysis of treatment effects on boys and girls separately shows that the reductions in problem behaviors are concentrated among boys; *Lively Minds* reduces problem behaviors among boys by just over a fifth of a standard deviation (see Panel B of Table 6, column 3, $p = 0.004$). This is consistent with broader evidence that boys tend to exhibit higher levels of externalizing problem behaviors than girls in early childhood—in our control group, on average, girls' problem behaviors score is 0.11 SD lower than boys'—and that the effect of *Lively Minds* is concentrated at the higher end of the problem score distribution (Keenan and Shaw, 1997).

In comparison to the findings for cognitive development, results in Table 7 suggest that improvements in problem behaviors are concentrated among children of PMs. Impacts among this group are large and significantly different from the small and statistically insignificant impacts on problem behaviors among children of non-PMs. Column 2 shows that among children of PMs, the program reduced behavioral difficulties by 0.26 SD (column 2, $p = 0.023$). This is a

sizeable effect, much larger than the average effect of 0.1-0.15 SD reported in a meta-analysis of preschool interventions (Holla et al., 2021). This effect is driven by a similar size and significant reductions in both internalizing and externalizing behaviors.

This finding suggests that the program's benefits for children's socioemotional development are likely driven by improvements in the home environment and caregiving practices among PMs, stemming from gains in their child development knowledge and behaviors. The absence of effects for children of non-PMs implies that changes occurring in preschools as a result of the program, while beneficial for cognitive outcomes, are insufficient on their own to improve children's socioemotional development in the absence of complementary improvements in the home environment.

4.3 Robustness

Selection of controls. Our main specifications follow our pre-specified approach to selection of controls, as described in Section 3. As a robustness check, we also estimate impacts on the cognitive aggregate score as well as the emotional awareness and behavioral difficulties scores, selecting controls using the post-double selection LASSO method (Belloni et al., 2014). The treatment effect estimates, shown in Table B.13, are similar in magnitude and significance to the main estimates reported above.

Measurement error. We re-estimate the main specifications using cognitive scores constructed with Item Response Theory (IRT) to assess the sensitivity of our results to measurement error in the assessment scores. As discussed in Section 3 and Appendix B, IRT explicitly models measurement error and allows items to differ in difficulty and discrimination, rather than imposing equal weights across all items as in simple averages. Table B.11 shows that the estimated treatment effects on the aggregate cognitive score, as well as on emergent numeracy and executive function, are very similar in magnitude and statistical significance to our main estimates using raw scores. Correcting for measurement error using IRT leads to effects broadly similar in magnitude, with somewhat larger estimated effects on emergent literacy (from 0.05 to 0.071 SD, $p = 0.11$), while executive function remains significant and externalizing behavior is marginal. Table B.12 shows the effects on the socioemotional outcomes.

Measurement invariance. A concern in interpreting treatment effects is that the intervention may affect how some assessment items function rather than the underlying skills themselves, which could bias comparisons between treated and control groups. We, therefore, test for measurement invariance for all latent constructs using standard multi-group factor models. We begin with a configural model that allows all parameters to vary by group and then sequentially test metric, scalar, and strict invariance by imposing equality restrictions on factor loadings, item intercepts, and residual variances, respectively, comparing model fit at each step. In practice, full scalar or strict invariance rarely holds in applied early-childhood settings, so we allow for partial invariance, freeing a small number of parameters where score tests indicate the largest violations while keeping the remaining items constrained to anchor the scale. Across our main constructs (cognitive skills, socioemotional skills, parental knowledge, and parenting behaviors), we find that while full invariance is often rejected, models with partial metric or partial scalar invariance fit the data as well as fully unconstrained models. For constructs where only partial invariance holds, we verify that our results are robust to using factor scores derived from the partially invariant model. As shown in Table A.5, most estimated impacts are nearly identical or larger and more statistically significant than the main results, which use a simple average of raw item scores for measures of child development and IRT scores constructed assuming full measurement invariance. See Appendix A for a more in-depth discussion and analysis of this issue.

5 Opportunity costs and household spillovers

A common feature of early childhood and related social programs in low-resource LMIC settings is reliance on low-paid (or unpaid), often marginalised women as frontline implementers. A prominent example is the community health worker (CHW) model: in many settings, CHW cadres are predominantly female and are tasked with delivering a broad portfolio of services, often with limited compensation and high workload, raising concerns about exploitation and burnout (Jain et al., 2022; Astale et al., 2023; Ballard et al., 2023). In India, for instance, the all-female ASHA workforce has been widely debated in relation to incentive-based pay, stress, and excessive workload (Shrivastava et al., 2023; Dhaliwal et al., 2025). Pakistan's Lady Health Worker program provides another example of a large female cadre responsible for multiple

maternal and child health tasks in underserved communities (Jalal, 2011). The *Lively Minds* program similarly depends on unpaid, low-educated mothers who already shoulder substantial domestic and care responsibilities. While such models can be cost-effective, a key welfare and scalability question is whether they impose hidden costs on participating women or their families. In spite of the critical role that such women play in the success of the ongoing global effort to improve early childhood provision, there is little evidence on how participation in the delivery of ECD programs affects them (Ballard et al., 2023; Sullivan et al., 2024).

In standard economic models of time allocation, volunteering is not "free": any additional program-related time must be accommodated through reductions in paid work, leisure, or other household activities (Becker, 1965; Gronau, 1977). This concern is particularly salient in low-income settings where women already face high unpaid care burdens and may be "time poor" (Bardasi and Wodon, 2010; Dinkelman and Ngai, 2022). If participation in the *Lively Minds* program meaningfully tightened these constraints, we would expect this to be reflected in mothers' psychosocial well-being.

Time requirements and psychosocial costs. In practice, the program's direct time requirement was modest—typically around 1–2 hours per week at the preschool—and predictable, making it easier to absorb through small adjustments in daily routines rather than major changes in labor supply. Moreover, many women's economic activities in this rural setting are informal, seasonal, and flexible, which likely reduced the likelihood of large crowd-out effects.

Consistent with this, we find no evidence that the program adversely affected participating mothers' mental health or self-esteem (Table 8).[6] A plausible interpretation is that the limited time commitment did not materially increase net time pressure. In addition, any incremental burden may have been offset by potential psychosocial benefits from volunteering, social interaction, and a greater sense of purpose or community engagement—channels that have been linked to better well-being in other contexts (Piliavin and Siegl, 2007; Hussam et al., 2022). *Lively Minds* program training and monthly parenting workshops also explicitly emphasized maternal well-being, which may have further mitigated potential stress.[7]

[6] See Section 3 for details on how these are measured.

[7] In fact, there is some suggestive evidence that the program may have had positive impacts on the mental health and self-esteem of participating mothers (PMs) in analyses that draw on an augmented booster sample of PMs.

Spillovers on siblings. We next examine whether the program generated spillovers within the household. This is in line with evidence that parent-focused interventions often generate broader familial impacts (Carneiro et al., 2023). Spillovers could be positive if (i) target children transmit what they learned at preschool to siblings, or (ii) participating mothers apply new knowledge and practices to all children in the household. Conversely, spillovers could be negative if mothers reallocated attention toward the target child at the expense of siblings.

To examine these possibilities, we estimate impacts on one younger and one older sibling of each target child (see Section 3) and report them in Table 9. The sample includes 749 younger and 1,083 older siblings. We find no evidence of program impacts on older siblings. For younger siblings, we observe a positive and economically meaningful impact ($0.21SD$), but the effect is imprecisely estimated (column 2, $p = 0.244$) due to the small sample. Overall, the results offer no evidence of indirect spillovers from children's participation, but improvements in parenting among PMs may benefit younger siblings. Importantly, we find no evidence of adverse effects on any children in the household.

6 Conclusion

Many LLMICs have expanded access to pre-primary schooling in recent years, but face financial and administrative constraints in improving what children actually learn in these settings. Returns to investing early are well established: skill formation is dynamic, early investments raise the productivity of later investments, and early deficits are costly to remediate, implying potentially high returns to improving the quality of early learning environments (Heckman, 2012; Gertler et al., 2014). Yet the binding policy question is not whether to expand access, but how to improve the quality of existing preschools at scale in systems where teachers are overstretched, learning materials are scarce, education budgets are shrinking, and many caregivers have limited education themselves.

This paper studies a government-led approach to this problem in rural Ghana. In partnership with Ghana's Ministry of Education and the NGO Lively Minds, we evaluate a program using a cluster-RCT that recruits mothers from the local community, trains them using simple, scripted materials that do not require literacy, and embeds them as regular volunteers in public preschools to deliver structured, play-based learning activities in small groups. An important innovation is

that the program uses mothers, rather than hired paraprofessionals or teachers, as the marginal classroom input, so that implementation itself can build parenting human capital and potentially improve the home environment concurrently.

We observe economically meaningful gains in children's cognitive development. The program increases an aggregate cognitive index by 0.11 SD, with impacts concentrated in emergent numeracy and executive function. Socioemotional impacts, however, are more selective: reductions in behavioral difficulties are concentrated among boys only. We also find that cognitive gains accrue broadly to all children exposed to improved classroom activities, suggesting that low-educated mothers can effectively augment teacher capacity when provided with structured training. On the other hand, reductions in behavioral difficulties are driven entirely by children of participating mothers, consistent with large improvements in mothers' knowledge of child development, greater engagement in play activities at home, and increases in observed positive parenting behaviors. This implies that engaging mothers as active contributors, rather than relying exclusively on classroom-based inputs, generates complementary benefits through the home environment channel, particularly for socioemotional outcomes. Importantly, we find no adverse effects on mothers' mental health or self-esteem, addressing concerns about the hidden costs of volunteer-based models in settings where women already face substantial care burdens.

The program is relatively low-cost, requiring $37 per child (in implementation-year USD) in additional funding beyond existing education-system resources used to deliver the program. This includes one-off set-up costs with significantly lower ongoing additional costs following set-up. Following this trial, the Government of Ghana expanded the model across 60 rural districts across Ghana. Administrative records from 46 of these districts (2,543 preschools) show high volunteer training completion and re-enrollment of mothers, and unannounced monitoring visits from the government indicate that both parenting workshops and play scheme sessions are typically running and are rated satisfactory when observed. Using unannounced GES monitoring visits data during the scale-up, Panel A of Figure 1 documents whether Play Scheme (PS) and Parenting Workshop (PW) sessions were running at the time of the visit. For PS, GES officials classified a session as running if more than five PMs were present and at least one teacher was present. For PW, a workshop was classified as successfully running if more than 75% of the expected PMs were present. The results suggest high and stable implementation

coverage overall. Panel B of Figure 1 further shows that satisfactory quality for both of the PS components appears to be around the 90%+ range throughout. The close tracking of PS and PW quality over time suggests that when schools and PMs were able to run sessions, implementation quality was consistently strong. Overall, these patterns provide reassuring evidence that the scale-up maintained not only session continuity but also a good standard of delivery for both play-based classroom activities and parenting engagement. The additional funds required to implement the program fell sharply at scale to roughly $17.4 per child (compared to $37 during the trial) during the set-up phase and as low as $3 per child post set-up. This decrease is due predominantly to greater integration of the program within the core preschool delivery system.

Two features of the *Lively Minds* program help rationalize why this model performs well under severe implementation constraints. First, the "production technology" is intentionally simple relative to standard parenting programs: it does not require literacy, relies on repeatable activity scripts, and is delivered to groups of parents rather than through intensive one-to-one coaching at home. Second, and more unusually, the main play-activity input is provided by low-educated, often illiterate mothers from the community. Given limited prior evidence on whether parents with very low schooling in poor rural LLMIC settings can effectively support children's learning in formal classroom environments, our results suggest that — with appropriate structure and support — they can. A likely reason for this success is that the program was developed iteratively over more than a decade in Northern Ghana, adapting the model to local constraints rather than transplanting a first-best design into weak systems.

Whether these results generalize to other low-capacity countries or contexts remains an open question. The model's reliance on volunteer mothers may be more feasible in contexts where women have flexible work schedules, strong community ties, and intrinsic motivation to support children's education - conditions that may not hold uniformly across LLMICs. More work is also needed on equilibrium and welfare effects, including how volunteer models affect women's labor supply and intrahousehold allocations over longer horizons, and whether communities experience volunteer fatigue or selection in volunteer supply as programs mature.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Amadu, S., O. Attanasio, B. Caeyers, L. Cardona Sosa, S. Cattan, S. Krutikova, P. Leighton, L. Masselus, and M. Yakubu (2018). Improving early childhood development in rural ghana through scalable low-cost community-run play schemes: Baseline report. IFS Report R144.1, Institute for Fiscal Studies.

Amadu, S., O. Attanasio, B. Caeyers, L. Cardona Sosa, S. Cattan, S. Krutikova, P. Leighton, and M. Yakubu (2020). Improving early childhood development in rural ghana through scalable community-run play schemes: Programme impact evaluation report. Technical report, Institute for Fiscal Studies, London, UK. Accessed: 30 December 2025.

Andrew, A., O. P. Attanasio, R. Bernal, L. C. Sosa, S. Krutikova, and M. Rubio-Codina (2024). Preschool quality and child development. *Journal of Political Economy 132*(7), 2304–2345.

Angrist, N., S. Kabay, D. Karlan, L. Lau, and K. Wong (2025). Human capital at home: Evidence from a randomized evaluation in the philippines. NBER Working Paper 33574, National Bureau of Economic Research.

Asiedu, E., D. Karlan, M. Lambon-Quayefio, and C. Udry (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences 118*(29), e2024570118.

Astale, T., T. Abebe, and G. Mitike (2023). Workload and emerging challenges of community health workers in low-and middle-income countries: A mixed-methods systematic review. *PLoS One 18*(3), e0282717.

Attanasio, O., S. Cattan, and C. Meghir (2022). Early childhood development, human capital, and poverty. *Annual Review of Economics 14*(1), 853–892.

Ballard, M., A. Olaniran, M. M. Iberico, A. Rogers, A. Thapa, J. Cook, Z. Aranda, M. French, H. E. Olsen, J. Haughton, et al. (2023). Labour conditions in dual-cadre community health worker programmes: a systematic review. *The Lancet Global Health 11*(10), e1598–e1608.

Bardasi, E. and Q. Wodon (2010). Working long hours and having no choice: Time poverty in guinea. *Feminist Economics 16*(3), 45–78.

Bassi, M., B. Besbas, L. I. D. Diaz, S. Ravindran, and A. Reynoso (2025). From access to achievement: The primary school-age impacts of an at-scale preschool construction program in highly deprived communities. *NBER working paper*.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal 75*(299), 493–517.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Blimpo, M. P., P. Carneiro, P. Jervis, and T. Pugatch (2022). Improving access and quality in early childhood development programs: Experimental evidence from the gambia. *Economic Development and Cultural Change 70*(4), 1479–1529.

Bornstein, M. H., L. R. Cote, O. M. Haynes, C.-S. Hahn, and Y. Park (2010). Parenting knowledge: experiential and sociodemographic factors in european american mothers of young children. *Developmental Psychology 46*(6), 1677–1693.

Bornstein, M. H. and W. A. Rothenberg (2022). The unicef multiple indicator cluster surveys and the early childhood development index. In M. H. Bornstein, W. A. Rothenberg, R. H. Bradley, K. Deater-Deckard, J. E. Lansford, A. Bizzego, G. Esposito, D. L. Putnick, and S. Zietz (Eds.), *Parenting and Child Development in Low- and Middle-Income Countries*. Routledge.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau (2018). Preschool and parental response in a second best world: Evidence from a school construction experiment. *Journal of Human Resources 53*(2), 474–512.

Brown, A., E. Weiss, N. M. Suntheimer, R. Appiah, E. Aurino, and S. Wolf (2025). Factorial and cultural validity of a social and emotional behavior measure in northern ghana. *Applied Developmental Science*. Advance online publication.

Carneiro, P., E. Galasso, I. L. Garcia, P. Bedregal, and M. Cordero (2024). Impacts of a large-scale parenting program: Experimental evidence from chile. *Journal of Political Economy 132*(4), 1113–1161.

Carneiro, P., I. Rasul, and F. Salvati (2023). Families as drivers of inequality: Experimental evidence from an early childhood intervention. *Working Paper*.

Cascio, E. U. and D. W. Schanzenbach (2013). The impacts of expanding access to high-quality preschool education. *Brookings Papers on Economic Activity*, 127–178.

Coelho, V., J. Cadima, and A. I. Pinto (2023). Child functioning in inclusive preschools: Associations between self-regulation, engagement, prosociality, and hyperactivity. *Journal of Applied Developmental Psychology 86*, 101518.

Coffey, J. R. and E. S. Spelke (2024). Direct cognitive testing in the field at scale: Leveraging core knowledge to assess the development of young children in ghana. Unpublished manuscript / preprint.

Cooke, E., S. Hague, and A. McKay (2016, March). The ghana poverty and inequality report: Using the 6th ghana living standards survey.

Cunha, F., I. Elo, and J. Culhane (2022). Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *Journal of Econometrics 231*(1), 3–32.

Currie, J. (2001). Early childhood education programs. *Journal of Economic perspectives 15*(2), 213–238.

Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of Labor Economics*, Volume 4, pp. 1315–1486. Elsevier.

Dean, J. T. and S. Jayachandran (2019). Attending kindergarten improves cognitive but not socioemotional development: Evidence from india. Working paper.

Dhaliwal, B. K., M. Gupta, A. Nadda, S. Singh, A. Shet, K. Scott, A. Dutta, and S. Closser (2025). Urban community health workers in punjab, india: A qualitative study of ashas' roles in the health system. *PLOS Global Public Health 5*(8), e0004629.

Diamond, A. and K. Lee (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science 333*(6045), 959–964.

Dinkelman, T. and L. R. Ngai (2022). Time use and gender in africa in times of structural transformation. *Journal of Economic Perspectives 36*(1), 57–80.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–2765.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2024). Augmenting state capacity for child development: Experimental evidence from india. *Journal of Political Economy 132*(5), 1565–1602.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science 344*(6187), 998–1001.

Ghana Statistical Service and UNICEF (2018). Ghana multiple indicator cluster survey 2017/18, survey findings report. Ghana Statistical Service in collaboration with Ministry of Health, Ministry of Education, and UNICEF.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry 38*(5), 581–586.

Goodman, R. (2001, November). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry 40*(11), 1337–1345.

Grantham-McGregor, S. M., C. A. Powell, S. P. Walker, and J. H. Himes (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the jamaican study. *The Lancet 338*(8758), 1–5.

Gronau, R. (1977). Leisure, home production, and work—the theory of the allocation of time revisited. *Journal of Political Economy 85*(6), 1099–1123.

Hamadani, J. D., F. Tofail, A. Hilaly, S. N. Huda, P. Engle, and S. M. Grantham-McGregor (2010). Use of family care indicators and their relationship with child development in bangladesh. *Journal of Health, Population and Nutrition 28*(1), 23–33.

Healey, D. and M. Healey (2019). Randomized controlled trial comparing the effectiveness of structured-play (engage) and behavior management (triple p) in reducing problem behaviors in preschoolers. *Scientific Reports 9*(1), 3497.

Heckman, J. J. (2012). Invest in early childhood development: Reduce deficits, strengthen the economy. *The Heckman Equation 7*(1-2), 1–2.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). The weirdest people in the world? *Behavioral and Brain Sciences 33*(2-3), 61–83.

Hirsh-Pasek, K., R. M. Golinkoff, L. E. Berk, and D. G. Singer (2009). *A mandate for playful learning in preschool: Applying the scientific evidence*. Oxford University Press.

Holla, A., M. Bendini, L. Dinarte Diaz, and I. Trako (2021, June). Is investment in preprimary education too low? lessons from (quasi) experimental evidence across countries. Policy Research Working Paper 9723, World Bank.

Hoosen, N., E. L. Davids, P. J. de Vries, and M. Shung-King (2018a). The strengths and difficulties questionnaire (sdq) in africa: a scoping review of its application and validation. *Child and Adolescent Psychiatry and Mental Health 12*(1), 6.

Hoosen, N., E. L. Davids, P. J. de Vries, and M. Shung-King (2018b). The strengths and difficulties questionnaire (sdq) in africa: A scoping review of its application and validation. *Child and Adolescent Psychiatry and Mental Health 12*(6).

Hussam, R., E. M. Kelley, G. Lane, and F. Zahra (2022). The psychosocial value of employment: Evidence from a refugee camp. *American Economic Review 112*(11), 3694–3724.

Jain, M., Y. Caplan, B. M. Ramesh, H. Kemp, B. Hammer, S. Isac, J. Blanchard, V. Namasivayam, and S. K. Sgaier (2022). Improving community health worker compensation: a case study from india using quantitative projection modeling and incentive design principles. *Global Health: Science and Practice 10*(3).

Jalal, S. (2011). The lady health worker program in pakistan—a commentary. *European Journal of Public Health 21*(2), 143–144.

Jeong, J., H. O. Pitchik, and G. Fink (2021). 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).

Keenan, K. and D. Shaw (1997). Developmental and social influences on young girls' early problem behavior. *Psychological Bulletin 121*(1), 95.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. T. Normand, E. E. Walters, and A. M. Zaslavsky (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine 32*(6), 959–976.

Lim, I. E., I. Mahmud, M. U. Chowdhury, A. Kaiser, F. A. Bonny, S. Akanbonga, and M. T. Hasan (2023). Early childhood development (ecd) in ghana: Assessing the status and determinants of the literacy–numeracy, physical, social–emotional, and learning domains. *International Journal of Early Childhood*. Published online 28 Feb 2023.

MacPhee, D. (1981a). *Knowledge of Infant Development Inventory (KIDI)*. Unpublished manual, University of North Carolina.

MacPhee, D. (1981b). Manual: Knowledge of infant development inventory. Unpublished manuscript, Department of Psychology, University of North Carolina at Chapel Hill.

Malapit, H. J. L. and A. R. Quisumbing (2015). What dimensions of women's empowerment in agriculture matter for nutrition in ghana? *Food Policy 52*, 54–63.

Martinez, S., S. Naudeau, and V. Pereira (2017). Preschool and child development under extreme poverty: Evidence from a randomized experiment in rural mozambique. Working Paper 8290, World Bank, Washington, DC.

McCoy, D. C., C. R. Sudfeld, D. C. Bellinger, A. Muhihi, G. Ashery, T. E. Weary, W. Fawzi, and G. Fink (2017). Development and validation of an early childhood development scale for use in low-resourced settings. *Population Health Metrics 15*(1), 3.

Ministry of Education, Republic of Ghana (2025). 2025 Programme Based Budget Estimates: Ministry of Education. Programme based budget estimates, Ministry of Education, Republic of Ghana.

National Council for Curriculum and Assessment (NaCCA) (2023). 2022 P4 NST Report. National standardised test report, National Council for Curriculum and Assessment. Findings from the 2022 National Standardised Test for Basic Four pupils, administered and analysed by NaCCA.

Özler, B., L. C. Fernald, P. Kariger, C. McConnell, M. Neuman, and E. Fraga (2018). Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial. *Journal of Development Economics 133*, 448–467.

Piliavin, J. A. and E. Siegl (2007). Health benefits of volunteering in the wisconsin longitudinal study. *Journal of Health and Social Behavior 48*(4), 450–464.

Romano, J. P. and M. Wolf (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton university press.

Schmitt, D. P. and J. Allik (2005). Simultaneous administration of the rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology 89*(4), 623–642.

Sher-Censor, E., T. Y. Khafi, and T. M. Yates (2016). Preschoolers' self-regulation moderates relations between mothers' representations and children's adjustment to school. *Developmental Psychology 52*(11), 1793.

Shrivastava, R., L. Sharma, M. Jolly, R. Ahuja, R. Sharma, J. A. Naslund, J. Agrawal, R. Shidhaye, S. Mehrotra, S. D. Hollon, et al. (2023). "we are everyone's ashas but who's there for us?" a qualitative exploration of perceptions of work stress and coping among rural frontline workers in madhya pradesh, india. *Social Science & Medicine 336*, 116234.

Sullivan, J., Z. Aranda, M. Adusumilli, A. Martens, and A. Jiménez (2024). Perceptions around occupational mental well-being of community health workers and an intervention package for its promotion: A mixed-methods study in rural chiapas, mexico. *SSM-Mental Health 6*, 100360.

The World Bank (2024). Ghana Accountability for Learning Outcomes Project (P165557) – Implementation Status & Results Report. Implementation status & results report, World Bank. Implementation Status & Results Report for the Ghana Accountability for Learning Outcomes Project as of December 25, 2024.

UNICEF (2019a). Ghana multiple indicator cluster survey 2017/18, survey findings report. Technical report, Ghana Statistical Service, Accra, Ghana. Co-published by UNICEF.

UNICEF (2019b). A world ready to learn: Prioritizing quality early childhood education. Global report, United Nations Children's Fund (UNICEF), New York, NY.

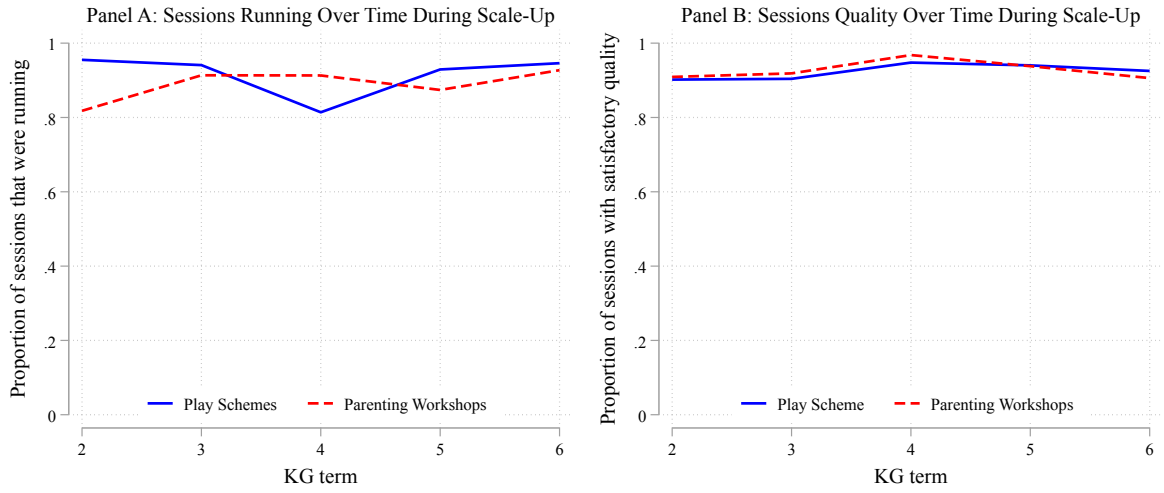
World Bank (2018a). The world development report 2018 - learning to realize education's promise. Wdr, World Bank, New York, NY.

World Bank (2018b). *World Development Report 2018:* Learning to Realize Education's Promise. World Bank.

World Bank and UNESCO (2024, January). Education Finance Watch 2024. Technical report, World Bank and UNESCO Institute for Statistics.

Youth in Mind (2020). Strengths and difficulties questionnaire (sdq). https://www.sdqinfo.org/. Accessed 2026.

Figure 1: Implementation and compliance during scale-up



Note: Panel A shows the proportion of sessions observed to be running during unannounced visits. Panel B shows the proportion of observed sessions meeting a satisfactory quality threshold, from the 2nd to the 6th kindergarten term. Note that KG Term 1 is the set-up phase during which they recruit and train the PMs; hence, no session data are available.

Table 1: Teacher reports related to program compliance

	Parents Contribute	Parents support at least once a month	Any games for free play	Any play in small groups
	(1)	(2)	(3)	(4)
Treatment	4.088	0.382	0.291	0.052
	(1.671)	(0.081)	(0.089)	(0.041)
p-val	*0.001*	*0.000*	*0.002*	*0.213*
Control mean endline	2.154	0.372	0.526	0.885
Strata Fixed Effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	156	156	156	156

Notes: The sample includes teachers who work in one of the pre-schools randomized into treatment or control at baseline and who were sampled to answer the teacher survey at endline. In Column (1), the coefficient is an odds ratio from an ordered logit model. In Columns (2)–(4), OLS estimates are presented. Across all columns, standard errors are clustered at the community level, and p-values are reported in italics. The model controls for strata fixed effects and an indicator for whether there is a clothing shop in the community. The outcomes and baseline outcomes are standardised to have mean 0 and standard deviation 1 in the control group. At the foot of each column we report the mean outcome among controls tracked to endline.

Table 2: Impacts on parental knowledge, play investments, and behaviors

	I. Knowledge		II. Play Investments		III. Behaviors	
	Child Development	Pre-school Quality	Play Materials	Any play activities	Positive Behaviors	Negative Behaviors
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment among PMs	0.459	0.122	-0.038	0.181	0.196	-0.026
	(0.131)	(0.106)	(0.135)	(0.049)	(0.136)	(0.139)
p-val	*0.001*	*0.253*	*0.780*	*0.000*	*0.153*	*0.850*
Adj. p-val	*0.012*	*0.537*	*0.966*	*0.012*	*0.467*	*0.966*
Treatment among non-PMs	-0.044	0.138	-0.202	-0.031	0.048	-0.012
	(0.105)	(0.060)	(0.104)	(0.036)	(0.076)	(0.095)
p-val	*0.673*	*0.024*	*0.056*	*0.390*	*0.528*	*0.901*
Adj. p-val	*0.866*	*0.096*	*0.170*	*0.762*	*0.816*	*0.866*
Difference in treatment	0.503	-0.016	0.164	0.213	0.147	-0.014
	(0.119)	(0.113)	(0.130)	(0.042)	(0.105)	(0.136)
p-val	*0.000*	*0.889*	*0.211*	*0.000*	*0.165*	*0.916*
Adj. p-val	*0.002*	*0.980*	*0.387*	*0.002*	*0.387*	*0.980*
Control mean, non-PMs	0.034	-0.009	-0.091	0.316	0.006	0.018
Control mean, PMs	0.063	0.246	-0.061	0.364	0.184	-0.037
Baseline outcome	Yes	No	Yes	Yes	No	No
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151	2,151	2,151

Notes: All outcomes are raw scores standardised to have mean 0, std 1 in control group. The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented where standard errors are clustered at the community level, and p-values are reported in italics. Column (1) uses an outcome from the KIDI scale. Column (2) uses a vignette-based instrument measuring parental knowledge about different aspects of pre-school quality. Columns (3) and (4) use outcomes from the Family Care Indicator. Columns (5) and (6) use outcomes from a surveyor-reported instrument on parenting behaviors toward the target child. Throughout we estimate fully interacted models that control for the baseline outcome (if available), strata fixed effects, child's age in months at endline, a gender indicator, and the number of clothing stores in the community, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother signed up to participate in the play schemes (during the intervention in the treatment group, in the following year in the control group). Control means are reported for each subgroup at the bottom of the table.

Table 3: Treatment effects on cognitive outcomes

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
A. Full sample				
Treatment	0.107	0.050	0.108	0.127
	(0.053)	(0.057)	(0.051)	(0.052)
p-val	*0.049*	*0.383*	*0.038*	*0.017*
Adj. *p*-val	-	*0.269*	*0.038*	*0.038*
Baseline outcome	0.450	0.368	0.381	0.231
	(0.025)	(0.025)	(0.022)	(0.022)
p-val	*0.000*	*0.000*	*0.000*	*0.000*
B. Heterogeneity by gender				
Treatment among boys	0.104	0.056	0.092	0.138
	(0.060)	(0.069)	(0.060)	(0.061)
p-val	*0.086*	*0.421*	*0.132*	*0.028*
Adj. *p*-val	-	*0.317*	*0.142*	*0.064*
Treatment among girls	0.122	0.058	0.129	0.133
	(0.060)	(0.060)	(0.057)	(0.065)
p-val	*0.044*	*0.334*	*0.027*	*0.044*
Adj. *p*-val	-	*0.305*	*0.044*	*0.050*
Difference	0.018	0.002	0.037	-0.005
	(0.054)	(0.063)	(0.058)	(0.074)
p-val	*0.743*	*0.975*	*0.519*	*0.949*
Adj. *p*-val	-	*0.988*	*0.854*	*0.988*
Control mean endline (full sample)	0.000	0.000	0.000	0.000
Control mean endline (boys)	0.067	0.009	0.090	0.054
Control mean endline (girls)	0.015	0.012	0.022	-0.006
Strata fixed effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: Standard errors are reported in parentheses. *p*-values are reported in italics. Adjusted *p*-values account for multiple hypothesis testing. The sample includes children attending pre-schools randomised into treatment or control at baseline and successfully tracked at endline. OLS estimates are reported with standard errors clustered at the community level. The outcome in Column (1) is the average of scores on IDELA and Spelke tasks measuring emergent literacy, numeracy, and executive function. Outcomes in Columns (2)–(4) correspond to the three subdomains. Outcomes are standardised to have mean 0 and standard deviation 1 in the control group.

Table 4: Distributional treatment effects: probability of scoring above selected control-group percentiles

	Dependent variable: Indicator for score above:				
	10^{th} perc	**25^{th} perc**	**Median**	**75^{th} perc**	**90^{th} perc**
	(1)	(2)	(3)	(4)	(5)
Cognition	0.008	0.047	0.028	0.064	0.000
	(0.013)	(0.015)	(0.025)	(0.027)	(0.017)
p-val	*0.509*	*0.003*	*0.259*	*0.019*	*0.984*
Emergent Literacy	0.002	-0.001	0.035	0.019	0.006
	(0.013)	(0.020)	(0.021)	(0.026)	(0.015)
p-val	*0.903*	*0.954*	*0.103*	*0.464*	*0.688*
Emergent Numeracy	0.007	0.033	0.038	0.045	0.016
	(0.009)	(0.016)	(0.024)	(0.025)	(0.016)
p-val	*0.435*	*0.049*	*0.113*	*0.075*	*0.328*
Executive Function	0.024	0.051	0.017	0.031	0.025
	(0.013)	(0.017)	(0.023)	(0.028)	(0.018)
p-val	*0.068*	*0.004*	*0.475*	*0.268*	*0.172*
Emotional Awareness	-0.009	0.014	0.029	0.015	0.008
	(0.013)	(0.019)	(0.026)	(0.032)	(0.021)
p-val	*0.500*	*0.465*	*0.270*	*0.648*	*0.713*
Emotional and Behavioural Pbms	0.005	-0.021	0.002	-0.069	-0.054
	(0.016)	(0.029)	(0.027)	(0.024)	(0.016)
p-val	*0.738*	*0.476*	*0.928*	*0.004*	*0.001*
Externalising Behaviours	0.006	-0.010	-0.058	-0.045	-0.033
	(0.009)	(0.025)	(0.031)	(0.025)	(0.015)
p-val	*0.515*	*0.692*	*0.063*	*0.072*	*0.031*
Internalising Behaviours	0.008	0.019	0.019	-0.032	-0.053
	(0.010)	(0.024)	(0.032)	(0.027)	(0.022)
p-val	*0.425*	*0.445*	*0.543*	*0.250*	*0.018*
Strata fixed effects	Yes	Yes	Yes	Yes	Yes
Baseline outcome	Yes	Yes	No	Yes	No
Characteristics	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151	2,151

Notes: The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented; standard errors (in parentheses) are clustered at the community level. *P*-values are reported in italics. The dependent variable is an indicator for whether the measure of development is above a particular percentile of the endline score distribution. Models control for strata fixed effects, child age (in months) at endline, a gender dummy, and a variable for the number of clothing stores in the community; specifications also include baseline outcomes as indicated.

Table 5: Impacts on cognitive outcomes, by PM status

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.117	0.007	0.137	0.127
	(0.070)	(0.070)	(0.081)	(0.091)
p-val	*0.099*	*0.919*	*0.096*	*0.166*
Adj. p-val		*0.922*	*0.194*	*0.230*
Treatment, non-PM children	0.091	0.046	0.085	0.126
	(0.054)	(0.059)	(0.054)	(0.050)
p-val	*0.098*	*0.434*	*0.121*	*0.015*
Adj. p-val		*0.339*	*0.136*	*0.028*
Difference in treatment	0.026	-0.039	0.052	0.002
	(0.061)	(0.061)	(0.083)	(0.080)
p-val	*0.672*	*0.527*	*0.530*	*0.984*
Adj. p-val		*0.842*	*0.842*	*0.974*
Control mean non-PM children	0.016	-0.008	0.031	0.004
Control mean PM children	0.140	0.083	0.155	0.104
Baseline outcome	Yes	Yes	Yes	Yes
Strata fixed effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The sample includes children attending pre-schools randomised into treatment or control at baseline and successfully tracked at endline. OLS estimates are reported with standard errors clustered at the community level, and p-values are reported in italics. The outcome in Column (1) is the average score on IDELA and Spelke tasks measuring emergent literacy, numeracy, and executive function. Columns (2)–(4) report the corresponding subdomains. All specifications estimate fully interacted models controlling for baseline outcomes, strata fixed effects, child's age in months, gender, number of clothing stores in the community, and interactions of these variables with treatment and an indicator for maternal participation in the Lively Minds programme. Control means at the bottom of each column correspond to children in the control group whose mothers did or did not sign up to participate in the programme in the following year.

Table 6: Impacts of intervention on children's socio-emotional outcomes

	Emotional Awareness	**Emotional Behavioural Problems**		
		Overall	**Externalising Behaviours**	**Internalising Behaviours**
	(1)	(2)	(3)	(4)
Panel A - Full sample				
Treatment	0.042	-0.098	-0.110	-0.061
	(0.073)	(0.059)	(0.061)	(0.060)
p-val	*0.560*	*0.102*	*0.074*	*0.314*
Adj. *p*-val			*0.116*	*0.238*
Baseline outcome	0.229	0.065	0.093	0.038
	(0.024)	(0.023)	(0.022)	(0.028)
p-val	*0.000*	*0.007*	*0.000*	*0.173*
Panel B - Heterogeneity by gender				
Treatment among boys	0.071	-0.211	-0.187	-0.178
	(0.079)	(0.070)	(0.071)	(0.073)
p-val	*0.372*	*0.004*	*0.011*	*0.017*
Adj. *p*-val			*0.026*	*0.026*
Treatment among girls	0.023	0.015	-0.038	0.059
	(0.082)	(0.077)	(0.077)	(0.076)
p-val	*0.775*	*0.845*	*0.626*	*0.441*
Adj. *p*-val			*0.601*	*0.601*
Difference	-0.048	0.226	0.149	0.237
	(0.072)	(0.089)	(0.086)	(0.089)
p-val	*0.508*	*0.013*	*0.085*	*0.009*
Adj. *p*-val			*0.064*	*0.010*
Control mean endline	0.000	0.000	-0.000	-0.000
Strata fixed effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The sample includes children attending preschools randomised into treatment or control at baseline and successfully tracked at endline. OLS estimates are reported with standard errors clustered at the community level, and *p*-values are reported in italics. The outcome in Column (1) is the average score on IDELA and Spelke tasks measuring emotional awareness. Column (2) reports an index based on SDQ items capturing emotional, peer, conduct, and hyperactivity problems. Columns (3) and (4) report indices for externalising (conduct and hyperactivity) and internalising (emotional and peer problems) behaviours, respectively. All specifications control for strata fixed effects, child's age in months at endline, gender, and the number of clothing stores in the community. Column (1) additionally controls for baseline emotional awareness; Columns (2)–(4) control for the corresponding baseline outcome. Outcomes are standardised to mean zero and unit variance in the control group. Panel A includes the baseline outcome as a control. Panel B interacts all regressors with an indicator for the child being a girl. Control means are reported at the bottom of each column.

Table 7: Impacts on children's socio-emotional outcomes, by PM status

	Emotional Awareness	**Emotional Behavioural Problems**		
		Overall	**Externalising Behaviours**	**Internalising Behaviours**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.006	-0.261	-0.274	-0.181
	(0.088)	(0.113)	(0.111)	(0.112)
p-val	*0.948*	*0.023*	*0.016*	*0.110*
Adj. p-val			*0.038*	*0.098*
Treatment, non-PM children	0.035	-0.064	-0.082	-0.030
	(0.074)	(0.056)	(0.056)	(0.064)
p-val	*0.641*	*0.257*	*0.146*	*0.635*
Adj. p-val			*0.202*	*0.603*
Difference in treatment	-0.029	-0.197	-0.192	-0.151
	(0.075)	(0.102)	(0.096)	(0.111)
p-val	*0.701*	*0.057*	*0.050*	*0.178*
Adj. p-val			*0.078*	*0.130*
Control mean non-PM children	0.015	-0.035	-0.033	-0.026
Control mean PM children	0.051	-0.014	-0.047	0.020
Baseline outcome	Yes	Yes	Yes	Yes
Strata fixed effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The sample includes children attending pre-schools randomised into treatment or control at baseline and successfully tracked at endline. OLS estimates are reported with standard errors clustered at the community level, and p-values are reported in italics. The outcome in Column (1) is the average score on IDELA and Spelke tasks measuring emotional awareness. Column (2) reports an index based on SDQ items capturing emotional, peer, conduct, and hyperactivity problems. Columns (3) and (4) report indices for externalising (conduct and hyperactivity) and internalising (emotional and peer problems) behaviours, respectively. All specifications estimate fully interacted models controlling for baseline outcomes, strata fixed effects, child's age in months, gender, number of clothing stores in the community, and interactions of these variables with treatment and an indicator for maternal participation in the Lively Minds programme. Control means correspond to children in the control group whose mothers did or did not sign up to participate in the programme in the following year.

Table 8: Psychological impacts of intervention on PMs

	Depressive symptoms	**Self-esteem**
	(1)	(2)
Treatment	0.114	0.023
	(0.128)	(0.112)
p-val	*0.374*	*0.835*
Baseline outcome	0.145	-0.016
	(0.054)	(0.048)
p-val	*0.009*	*0.743*
Control mean endline (PM sample)	0.018	0.055
Strata fixed effects	Yes	Yes
Characteristics	Yes	Yes
Observations	437	437

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The sample includes primary caregivers of children attending pre-schools randomised into treatment or control at baseline who were successfully tracked at endline and volunteered to run the play schemes. OLS estimates are reported with standard errors clustered at the community level, and p-values are reported in italics. In Column (1), the outcome is the mean score across the 20 items of the SRQ-20 Depression Scale. In Column (2), the outcome is the mean score across the items of the Rosenberg Self-Esteem Scale. Both outcomes are standardised to have mean zero and unit variance in the control group. All models control for the baseline outcome, strata fixed effects, child's age in months at endline, gender, and the number of clothing stores in the community. Control means correspond to the mean outcome among PMs in the control group at endline.

Table 9: Spillover effects of the intervention on siblings' development

	Older siblings	**Younger siblings**
	(1)	(2)
Treatment among PM children	-0.029	0.205
	(0.165)	(0.175)
p-val	*0.863*	*0.244*
Treatment among non-PM children	-0.006	0.028
	(0.084)	(0.096)
p-val	*0.946*	*0.770*
Difference in treatment	-0.023	0.177
	(0.164)	(0.206)
p-val	*0.889*	*0.394*
Control mean PM children	0.178	-0.061
Control mean non-PM children	-0.053	0.017
Baseline outcome	Yes	Yes
Strata fixed effects	Yes	Yes
Characteristics	Yes	Yes
Observations	1,083	749
Observations non-PM	834	589
Observations PM	249	160

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The sample consists of the next older and younger siblings of children who attended pre-schools randomised into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are reported with standard errors clustered at the community level, and p-values are reported in italics. In Column (1), the outcome is a latent factor score constructed using the older sibling's performance on literacy, numeracy, two maths tasks, and an emotional awareness task. In Column (2), the outcome is the younger sibling's age-adjusted score on the CREDI scale of child development. All outcomes are standardised to have mean zero and unit variance in the control group. All specifications estimate fully interacted models controlling for baseline outcomes (for older siblings, an aggregate literacy and numeracy score), strata fixed effects, child's age in months at endline, and the number of clothing stores in the community, along with interactions of these variables and the treatment indicator with maternal participation in the Lively Minds programme. The younger sibling's age is not observed and is therefore not included as a control. Control means correspond to children in the control group whose mothers did or did not sign up to participate in the programme in the following year.

A Measures of child development and parental investments

This appendix section provides a description of the measures of child development and parental investments that we use in the analysis presented in this paper. It establishes the internal validity of these measures and describes how we construct the outcomes we use in the analysis. As part of this, we discuss the notion of measurement invariance and the test we perform to ensure that the measures of child and parental outcomes we use can be compared across relevant groups.

A.1 Instruments to measure child and parental outcomes

Table A.1 lists the instruments used to measure the development of the child and their siblings, as well as parental knowledge, investment and parenting behaviours. The last two columns also indicate whether we collected the instrument at baseline and/or at endline.

Table A.1: Instruments used to measure child and parental outcomes

Construct	**Instrument**	**Baseline**	**Endline**
Development of target child	IDELA	Y	Y
	Spelke Tasks		Y
	Strengths and Difficulties Questionnaire (SDQ)	Y	Y
Development of younger sibling	CREDI		Y
Development of older sibling	Literacy and numeracy tests		Y
	Spelke tasks		Y
Parental investments	Knowledge of Infant Development Inventory (KIDI)	Y	Y
	Knowledge of pre-school quality		Y
	Play materials and activities (Family Care Indicators)	Y	Y
	Parenting behaviours		Y
Maternal wellbeing	Rosenberg of self-esteem	Y	Y
	SRQ-10 scale of depressive symptoms	Y	Y

When selecting measurement tools to measure the target child's developmental outcomes, we took into account three key considerations. First of all, we selected measures that allowed us to study impacts on specific sub-domains of children's development, rather than overall aggre-

gate measures. For example, we specifically selected instruments that allowed us to separately measure skills related to emergent numeracy, or emergent literacy, as opposed to just overall cognitive development. Second, we wanted to make sure that we used internationally validated instruments that have previously also been used in Ghana and in comparable contexts. A third and final key consideration was to ensure that our child development assessment is complete, so that it covers all important aspects of child development and also that it was robust to potential "teaching to the test" effects of the Lively Minds curriculum. For these reasons, we used three main instruments: the International Development and Early Learning Assessment (IDELA), tasks developed by Elizabeth Spelke's Lab ("Spelke Tasks" henceforth), and the Strengths and Difficulties Questionnaire (SDQ).

For siblings' and parental outcomes, we selected, to the best extent possible, already validated measures that had been used in low-income contexts and complemented these measures with newly created measures when available measures did not measure the constructs of interest comprehensively enough. We now discuss each measure in turn.

A.1.1 Instruments to measure the target child's development

IDELA The IDELA is a free instrument developed by *Save the Children* to measure the early learning and development of children from 3.5 to 6 years old. It was purposefully designed to be internationally applicable in LMIC settings and comparable and has already been used and validated in many contexts, including Ghana.

The IDELA assessment takes about 35 minutes per child. It consists of a series of tasks, which are conducted with the child in a 1-to-1 setting and scored by the data collector. The tasks cover a range of cognitive and socio-emotional dimensions, which include:

- *emergent numeracy*, i.e. the ability to do basic math which for the target age-group is measured for example by recognition of numbers, shapes and sizes
- *emergent literacy*, i.e. the ability to read or to recognise letters
- *fine motor skills*, i.e. the mental control and self-regulation, for example the ability to memorise, or plan and control impulses
- *socio-emotional skills*, specifically around the child's expression and management of their

own emotion, their ability to empathise with others and to establish positive rewarding relationships with them.

The IDELA was piloted extensively prior to the baseline survey, leading to a small number of context-specific adaptations to improve children's comprehension and engagement. These changes primarily involved replacing unfamiliar items with locally relevant alternatives, simplifying task structure, and adding one domain-relevant question.[8]

Harvard Development Studies Lab tasks ('Spelke's tasks') Based on our third consideration around the comprehensiveness of our child development assessment and the risk of "teaching to the test", at endline, we complemented the IDELA tool with a battery of assessments developed for and used in the Ghanaian by the psychology Laboratory of Development Studies at Harvard (Coffey and Spelke, 2024). With the exception of motor skills, these tasks measure similar developmental domains as the IDELA.

The tasks that we administered covered a range of skills, including emergent math, emergent literacy, executive function, and emotional awareness. For emergent math, we administered the following tasks:

- *Geometric Intruder*, which tests whether a child can identify which shape, among four presented to them, presents a unique difference
- *Panamath*, which tests whether the child can quickly determine which of two boxes on a laminated card has more dots (to test numerical estimation)
- *Point to number*, which tests whether a child can identify particular number of toys and, then, to cards with particular arabic numbers on them
- *Extra number*, which tests whether the child can identify which of two plates has more toys in them and whether the child can add the number of toys there are on the two plates

For emergent literacy, we administered a single task called *Vocabulary assessment* which tested the size of the child's vocabulary. In this task, the child is presented with a set of

[8]Specifically, one item was dropped as it proved too difficult for nearly all children during piloting (Item 1(f)); apples were replaced with mangoes to reflect locally familiar objects (Item 7); a four-piece puzzle was introduced as an example to support understanding of a more complex task (Item 8); an image was adapted to better reflect the local context (Item 11); a triangle was replaced with a square due to greater familiarity among children (Item 21); a hopping task was replaced with a pencil tap game to better capture the intended skill (Item 24); and one additional health and hygiene question was incorporated into the assessment (Item 25).

laminated cards with pictures of different objects. The assessor mentions an object and asks the child to point to the card with represents that object.

For executive function, we administered two tasks which required the child to play with objects on a mat:

- *Attention Switching*, which tests whether the child can find an object after the surveyor taps on an incorrect location.
- *Mental Simulation/Rotation*, which tests whether the child can find objects placed under cups after the cups slide or are switched on the map.

For emotional awareness, we administered a task called *Point to Emotion*, where the child is presented with several pictures of human faces showing different emotions and asked to point to the image showing the emotions described by the assessor.

Strengths and Difficulties Questionnaire The third instrument we administered to measure child developmental outcomes is a parent-reported behavioural assessment informed by the structure and content of the Strengths and Difficulties Questionnaire (Goodman, 2001) [9] The instrument includes 25 items describing child behaviours, with responses recorded on a 3-point Likert scale: (0) not true, (1) somewhat true, and (2) certainly true.[10] These items map onto key domains of socio-emotional development: conduct problems, hyperactivity, emotional problems, peer problems, and prosocial behaviour. Following this framework, conduct problems and hyperactivity are combined as indicators of externalising behaviours, while emotional and peer problems represent internalising behaviours. Our analysis focuses on 20 items capturing internalising and externalising problem behaviours (Achenbach, 1991).

A.1.2 Instrument to measure the development of younger siblings

We measured the development of younger siblings of the target child using the Caregiver-Reported Early Development Index (CREDI) Short Form (**?**). The CREDI consists of a set of population-level measures designed to assess early childhood development for children from

[9] The SDQ was initially developed and validated in Western high-income country settings (Goodman, 1997; Henrich et al., 2010), but has since been used internationally with versions available in over 90 languages (Youth in Mind, 2020), including in Sub-Saharan Africa and Ghana (Hoosen et al., 2018b; Brown et al., 2025).

[10] Instrument available on request

birth to age three using caregiver reports. Developed to be feasible for use in large-scale surveys and in low-resource settings, the CREDI focuses on core developmental domains including cognition, language, motor skills, and socio-emotional development. The instrument is administered through brief, age-specific questionnaires that ask caregivers about observable behaviors and milestones, rather than requiring direct child assessment. The CREDI Short Form, used in this study, consists of 20 caregiver-reported items for each child, designed as a fixed-length set to rapidly assess overall early childhood developmental status in population-level surveys and monitoring efforts. This contrasts with the Long Form, which includes up to 100 items and allows more detailed, domain-specific scoring.

A.1.3 Instruments used to measure the development of older siblings

We measured literacy and numeracy of older siblings with an adapted version of a test previously used in the Ghanaian context and developed by the Ministry of Education officials from the National Council for Curriculum and Assessments (NaCCA). The test items are aligned with the Ghana Education Service curriculum. The instrument includes tasks related to subtraction, addition and fractions.[11]

At endline we complemented these literacy and numeracy tests with a set of tasks also developed by the Harvard Development Studies Lab. Some of these tasks were similar to those administered to the target child and adjusted for the age of the older sibling. Specifically, the tasks tested math (*Geometric intruder*), literacy (*Vocabulary assessment*), emotional awareness (*Point to emotion*), and executive function. For executive function, the task tested the child's ability to hit the key on the same side or opposite side as an object.

A.1.4 Instruments to measure parental knowledge

Knowledge of Infant Development Inventory (KIDI) We measured the primary carer's knowledge of stimulation and care practice and her beliefs regarding the importance of these for children's development using the Knowledge of Infant Development (KIDI). The KIDI is a widely used self-report measure of parental knowledge about child developmental processes, norms and milestones, caregiving practices, and health and safety guidelines (MacPhee, 1981b). In its classic form, the KIDI consists of 75 items, covering a broad range of content relevant to

[11] Instrument available on request

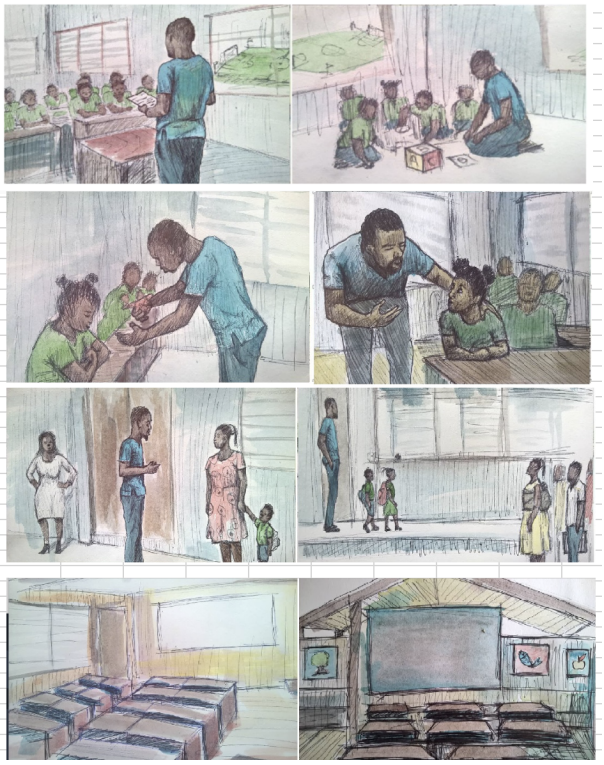


Figure A.1: Vignettes used to measure parental knowledge about pre-school quality

infants and young children, with respondents indicating whether statements are agree, disagree, or unsure (and in some milestone items, options about the age at which behaviors are expected). In the study, we used an abbreviated six-item version of the KIDI.

Knowledge about pre-school quality The second measure of parental knowledge hones in specifically on parental beliefs about what constitutes a developmentally appropriate pre-school environment. To assess this, we presented 4 pairs of visual vignettes designed by a local artist to the parents. As shown in Figure A.1, each vignette showed contrasting classroom scenarios. These include:

- Teacher directed whole class teaching versus play-based small-group interactions
- Authoritarian, top-down discipline versus guidance-based authoritative discipline
- Parent-teacher interactions versus separation between parents and teachers
- Classrooms high in structural quality versus ones with lower structural quality but more developmentally appropriate environment (e.g. pictures on walls)

For each pair, we asked parents to point to the image they thought was associated with higher quality pre-school education. The development of this tool was motivated by the finding during the baseline that existing survey tools for measuring parental beliefs about school quality yielded little variation in this study context in ways that raised concerns about strong experimenter demand effects.

A.1.5 Instruments to measure play materials and play activities

We used the Family Care Indicators (FCI) instrument, which is itself derived from the Home Observations for Measurement of the Environment for use in large-scale population surveys in LMICs (Hamadani et al., 2010). The FCI scale captures the availability of play and learning materials in the household, as well as educational and play activities conducted with the child in the previous 3 days by any household member over the age of 15. Enumerators recorded whether each of 13 categories of play and learning materials were available. Maternal report was used to establish whether and how many times each of the several learning and play activities took place over the preceding 3 days.

A.1.6 Instrument to measure parenting behaviours

Parenting behaviours were assessed using two observational approaches conducted by trained enumerators. First, enumerators directly observed mothers' teaching behaviors while they interacted with their child during two structured activities. Mothers were instructed on how to carry out two simple guessing and sorting games and then asked to play these games with their child, providing an opportunity for the enumerator to assess parenting style. Enumerators recorded the presence or absence of 15 specific behaviors reflecting instructional practices during the activities (coded 0 = not observed; 1 = observed). These behaviors captured both supportive and less supportive teaching strategies, such as adapting instructional approaches when initial strategies were ineffective, tailoring guidance to the child's prior actions, limiting the child's opportunities for independent problem-solving, or displaying low confidence while completing the task.

Second, enumerators documented mother–child interactions observed throughout the interview process. During the survey, enumerators coded whether mothers engaged in nonviolent or violent disciplinary practices toward the child. Nonviolent discipline included behaviors such

as removing privileges, explaining inappropriate behavior, or redirecting the child's attention. Violent discipline encompassed psychological aggression (e.g., shouting or using disparaging language) and physical punishment (e.g., shaking, spanking, hitting, or slapping the child, with or without an object). These items were drawn from a subset of the discipline measures used in the UNICEF Multiple Indicator Cluster Surveys (MICS) (Bornstein and Rothenberg, 2022), and enumerators were trained to systematically observe and code these behaviors during the course of the survey.

We grouped items from both sets of data about parenting behaviours and appraoches into two groups: one group of more supportive or positive parenting behaviours and another group of less supporting or more negative parenting behaviours.

A.1.7 Instruments to measure caregiver well-being

The well-being of primary caregivers of the target children (predominantly mothers) was measured in two ways. First, we administered the Rosenberg Self-Esteem scale (Rosenberg, 1965). This 10-item scale is globally one of the most widely used measures of self-esteem including in LMICs (Schmitt and Allik, 2005). Responses are on a 4-point Likert scale.

The second measure of mothers' wellbeing captured mental health using the SRQ-20 Psychological Distress Scale. This scale combines brevity with strong psychometric properties and has been used globally as part of the WHO World Mental Health Surveys (Kessler et al., 2002). It consists of 10 items asking about the frequency with which respondents experience a range of symptoms. Responses are on a 5-point Likert scale.

A.2 Constructing measures of child development and parental beliefs and investments

Constructing measures of child development and parental beliefs and investments used in the analysis involved two main steps. First, we mapped individual items to constructs of interest and validated such mapping by checking the construct's internal consistency. Second, we created aggregate scores for each of the constructs. We go through each of these steps in turn.

A.2.1 Step 1: Mapping measures to constructs and internal consistency of constructs

Table A.2 summarises how individual items from the IDELA, the Harvard Lab tasks, and the SDQ are mapped into three broad domains of child development: cognitive skill, emotional awareness, and emotional and behavioural problems. This classification is guided by the theoretical frameworks and intended constructs underlying each instrument, as defined by the test developers.

As shown in the table, items from the IDELA and the Harvard tasks cover closely related cognitive sub-dimensions, including emergent literacy, emergent numeracy, and executive function, as well as aspects of emotional awareness. In contrast, emotional and behavioural problems are measured exclusively using SDQ items, which capture dimensions such as conduct problems, emotional symptoms, hyperactivity, and peer relationships. These items are further grouped into two sub-constructs: externalising behaviours (combining conduct problems and hyperactivity) and internalising behaviours (combining peer and emotional problems).

Table A.3 shows how individual items from the KIDI, FCI, and the instruments we constructed to measure parental knowledge about pre-school quality and to measure parenting behaviours in the study are mapped into the domains of parental inputs we focus on in the paper: parental knowledge (about child development and pre-school quality), parental investment in play, and parenting behaviours. A

Table A.2: Mapping of items to the dimensions of child development they measure

Dimensions	Sub-dimensions	Measurements
Cognitive skill	**Emergent math**	Geometric intruder (Spelke task)
		Comparison by size and length (IDELA 2)
		Sorting and classification (IDELA 3)
		Shape identification (IDELA 4)
		Puzzle completion (IDELA 8)
		Panamath (Spelke task)
		Point to number (Spelke task)
		Number identification (IDELA 5)
		One-to-one correspondence (IDELA 6)
		Extra number (Spelke task)
		Addition and subtraction (IDELA 7)
	Emergent literacy	Vocabulary assessment (Spelke task)
		Expressive vocabulary (IDELA 15)
		Print awareness (IDELA 16)
		Letter identification (IDELA 17)
		First letter sounds (IDELA 18)
		Emergent writing (IDELA 19)
		Oral comprehension (IDELA 20)
	Executive function	Attention switching (Spelke task)
		Mental simulation / rotation (Spelke task)
		Short-term memory (IDELA 13)
		Inhibitory control (IDELA 14)
Emotional awareness		Point to emotion (Spelke task)
		Self-awareness (IDELA 1)
		Friends (IDELA 9)
		Emotional awareness/regulation (IDELA 10)
		Empathy/perspective taking (IDELA 11)
		Solving conflict (IDELA 12)
Emotional and behavioural problems	**Externalising behaviours**	SDQ tempers, obeys, fights, lies, steals
		SDQ restless, fidgety, distracted, thinks before acting, good attention
	Internalising behaviours	SDQ somatic symptoms, worries, unhappy, nervous in new situations, many fears
		SDQ solitary, has good friend, generally liked, bullied, help out

Table A.3: Mapping of parental measures to constructs

Dimensions	Sub-dimensions	Measurements
Parental knowledge	**Knowledge of infant and child development**	Abbreviated KIDI (all 6 items)
	Knowledge of parental quality	Instrument constructed for study using vignettes around rote learning, harsh discipline, parent-teacher interactions, and the presence of a developmentally appropriate environment
Parental investment in play	**Questions about play materials from the Family Care Indicator (FCI)**	*Number of toys/objects the surveyor can observed in the following categories:*
		Toys made at home
		Toys that have been bought (single toys/pack of toys)
		Household objects that the child plays with
		Toys to play music
		Building blocks
		Things for drawing, painting and/or writing
		Toys that induce constant physical movement
		Dolls and other objects that aid role play and fantasy games
		Children's books or to paint
		Toys to learn shapes and/or colour, including blocks and wall charts
		Picture charts to help the child learn vocabulary
		Other
	Questions about play activities from the Family Care Indicator (FCI)	*In the last 3 days until yesterday, did you or any other adult older than 15 years of age perform any of the following activities with the child?*
		Read or look at picture books together
		Tell stories
		Sing to child
		Go out to some place outside the house
		Play together with toys
		Make drawings, painting, writing, or play with paper and pencil
		Play to name objects, colours, count objects, or say numbers

Table A.3: Mapping of parental measures to constructs (cont.)

Dimensions	Sub-dimensions	Measurements
Parenting behaviours	**Positive/supporting parenting behaviours from instrument constructed for study**	***Surveyor's observations while the primary carer plays a game with the child***
		Ways the mother encouraged good behaviour
		Praised with words such as well done, good work
		Showed physical signs of affection, e.g. kiss, cuddle
		Rewarded by giving them something that is desired, such as a sweet
		Other behaviours the primary carer did during the task
		Informed the child that she is there to provide help if needed
		Responded positively when child completes part of the task
		Focused the child when they lose interest
		Provided instructions which are contingent upon child's previous action
		Tried new strategies for teaching if one was not working
		Provided feedback (e.g., okay, that's it, try again)
		Persisted despite difficulties; did not give up
		Appeared shy, timid, or under-confident in performing the task
		Set rules/told child how they expect him/her to behave
		Broke the task into small steps
		Turned the task into a game for child
		Used descriptive words for objects or actions
		Surveyor's observations if the primary carer did any of the following to any child in the house, at any time (excluding during the play task)
		Told child name of object or person during visit
		spontaneously praised child at least twice
		Voice conveyed positive feelings toward child
		Caressed or kissed child at least once
		Kept children in visual range, looked at often
		Responded verbally to child's vocalisations

Table A.3: Mapping of parental measures to constructs (cont.)

Dimensions	Sub-dimensions	Measurements
Parenting behaviours (cont.)	**Negative/harsh parenting behaviours from instrument constructed for study**	***Surveyor's observations while the primary carer plays a game with the child***
		Ways the mother dealt with bad behaviour
		Took away privileges, forbade something they liked, or did not allow child to leave house
		Shook child
		Shouted, yelled at, or screamed at child
		Gave child something else to do
		Spanked, hit, or slapped child on the bottom with bare hand
		Hit child on the bottom or elsewhere on the body with hard object
		Called child dumb, lazy, or another name like that
		Hit or slapped child on the face, head, or ears
		Other behaviours the primary carer did during the task
		Seemed largely disinterested in the task
		Displayed anger or hostility towards child
		Acted intrusively, not allowing opportunities for child to solve on their own
		Surveyor's observations if the primary carer did any of the following to any child in the house, at any time (excluding during the play task)
		Took away privileges, forbade something they liked, or did not allow him/her to leave house
		Shook him/her
		Shouted, yelled at, or screamed at him/her
		Gave him/her something else to do
		Spanked, hit, or slapped him/her on the bottom with bare hand
		Hit him/her on the bottom or elsewhere on the body with with hard object
		Called him/her dumb, lazy, or another name like that
Parental mental health	**Depressive symptoms**	SRQ-20 scale
	Self-esteem	Rosenberg scale of self-esteem

While these mappings follow the conceptual guidance provided by the test developers, we also assess its empirical validity by examining the internal consistency of items assigned to each (sub-)construct. Specifically, we analyse pairwise correlations across items (not reported here, but available upon request) and compute Cronbach's alpha, which measures the extent to which a set of items captures a common underlying construct.

Table A.4: Internal consistency of constructs at baseline and endline

Construct	Cronbach's Alpha	
	Baseline	**Endline**
Target child outcomes		
Cognitive skill	0.81	0.88
Emergent math	0.70	0.83
Emergent literacy	0.54	0.70
Executive function	0.48	0.42
Emotional awareness	0.61	0.66
Emotional and behavioural problems	0.81	0.75
Externalising behaviours	0.63	0.57
Internalising behaviours	0.72	0.62
Younger sibling outcomes		
CREDI		
Older sibling outcomes		
Parental outcomes		
Knowledge of child development	0.81	0.69
Knowledge of pre-school quality		0.32
Play materials	0.47	0.45
Play activities	0.48	0.48
Positive parenting behaviours		0.81
Negative parenting behaviours		0.69
Depressive symptoms	0.89	0.88
Self-esteem	0.69	0.60

Table A.4 reports Cronbach's alpha coefficients for all multi-item constructs at baseline and endline. Overall, the measures of cognitive skill exhibit high internal consistency, with alpha values above 0.8 for the aggregate index at both time points. Internal consistency is more moderate for the cognitive sub-dimensions, particularly for emergent literacy and executive

function, which is consistent with the greater heterogeneity of tasks used to capture these domains. Among socio-emotional measures, emotional awareness shows moderate reliability, while the SDQ-based measure of emotional and behavioural problems displays good internal consistency overall, though somewhat lower values for the externalising and internalising subscales than for the overall scales.

For parental outcomes, internal consistency varies more substantially across measures. Knowledge of child development and maternal depressive symptoms show relatively high reliability, while measures of play materials and play activities exhibit lower alpha coefficients. This likely reflects the very low values and lack of variation on many of the FCI items in our sample, which motivates us to only compute an aggregate measure of play materials. For play activities, our main measure is whether the child engages in any of the play activities included in the scale. s.

Taken together, the results indicate that the main aggregate constructs—particularly cognitive skills and emotional and behavioural problems—are measured with acceptable to high reliability, supporting their use in the main analysis. In a few cases, constructs display relatively low internal consistency, which suggests that results for these measures should be interpreted with appropriate caution.

A.2.2 Aggregating items into measures of underlying constructs

The analysis above indicates that, with a few exceptions, the items assigned to each construct exhibit sufficient internal consistency to be interpreted as measuring a common underlying dimension. We now describe how these items are aggregated into measures of latent constructs.

To fix ideas, we adopt a standard latent factor framework, in which each observed item is a function of an unobserved construct and an idiosyncratic error term. Observed responses depend on both the level of the latent trait θ and item-specific parameters, including factor loadings and intercepts. This framework provides a unifying representation of alternative approaches to index construction. Specifically, assume for all $j = 1, \ldots, J$, we can write $Y_{i,j}$, the j-th measure of the underlying θ for individual i, as a linear function of the underlying construct and measurement error $\epsilon_{i,j}$:

$$Y_{i,j} = \alpha_j + \lambda_j \theta_i + \epsilon_{i,j} \tag{1}$$

where α_j is referred to as an intercept and λ_j is referred to as a factor loading.

A central concern in our setting is the comparability of these measures across groups. Our analysis focuses on differences in average levels of child development and parental investments between treated and control units, and between participating and non-participating mothers. For such comparisons to be meaningful, the measurement model must be invariant across groups. Absent measurement invariance, differences in observed scores may reflect variation in item functioning rather than differences in the underlying construct.

Measurement invariance is typically characterised by a sequence of increasingly restrictive conditions. Metric invariance requires that factor loadings are equal across groups, ensuring that items load onto the latent construct on the same scale. Scalar invariance additionally requires equality of item intercepts, which is necessary for comparisons of latent means. Strict invariance further imposes equality of residual variances. While strict invariance is not required for our purposes, scalar invariance is essential for interpreting differences in average outcomes.

In practice, full invariance often fails, particularly in applied settings where interventions may affect specific items differently. We therefore allow for partial invariance, whereby equality constraints are imposed on a subset of parameters while others are allowed to vary across groups. Provided that at least one item remains invariant, the latent construct can still be meaningfully identified and compared across groups.

To guide our choice of measurement approach, we conduct standard tests of measurement invariance. We begin with a fully unconstrained (configural) model and sequentially impose equality of factor loadings (metric invariance) and intercepts (scalar invariance), evaluating whether these restrictions lead to a deterioration in model fit. We report p-values from likelihood ratio tests comparing nested models, where a high p-value indicates that the additional restrictions are not rejected.

A priori, for child outcomes, we would expect measurement invariance to be stronger for cognitive outcomes and emotional awareness than for emotional and behavioural problems, given that the former are based on direct observations and the later is based on maternal reports. For parental outcomes, we do not have clear hypothesis given that all the constructs reflect parental knowledge and parental behaviours, which will have been potentially been differently affected by the intervention and whether a mother was participating in the LM scheme.

Table A.5 reports the results of these tests for child outcomes. For several constructs,

including executive function, emotional awareness, and internalising behaviours, both metric and scalar invariance are not rejected, supporting the comparability of these measures across groups. For the overall cognitive construct and emergent numeracy, metric invariance is initially rejected; however, partial metric invariance can be achieved by freeing a small number of factor loadings, after which scalar invariance is not rejected. For emergent literacy, scalar invariance is rejected, but partial scalar invariance can be obtained by relaxing a limited number of intercepts.

For socio-emotional outcomes, a similar pattern emerges. While metric invariance is not rejected for the overall measure of emotional and behavioural problems, scalar invariance is rejected, but partial scalar invariance can be achieved with minor adjustments. For externalising behaviours, both metric and scalar invariance are rejected; however, allowing a subset of factor loadings and intercepts to vary across groups restores partial invariance. Overall, these results suggest that, although full invariance does not hold uniformly across all constructs, departures from invariance are limited and can be addressed through modest relaxations of the measurement model.

Table A.6 presents the corresponding results for parental outcomes. For parental knowledge of child development, both metric and scalar invariance are not rejected. For parenting behaviours, metric invariance is not rejected for positive behaviours, and scalar invariance is only marginally rejected, with partial scalar invariance achieved by freeing a small number of intercepts. For negative parenting behaviours, metric invariance is rejected, but partial metric and scalar invariance can be achieved by allowing selected parameters to vary across groups, indicating that non-invariance is concentrated in specific items.

For maternal mental health, the results again indicate that full invariance does not always hold, but that partial invariance provides a good approximation. For depressive symptoms, metric invariance is not rejected, while scalar invariance is rejected but can be restored under partial invariance. For self-esteem, metric invariance is not rejected, and partial scalar invariance can be achieved with minimal adjustments. These patterns are consistent with the expectation that self-reported measures may exhibit some heterogeneity in how items are interpreted across groups.

Taken together, the results indicate that most constructs satisfy either full or partial measurement invariance, with only limited deviations from the baseline model. This supports the

use of a common latent framework for comparing outcomes across groups.

Based on these results, we construct three sets of measures. First, we estimate factor scores from models allowing for partial invariance, thereby accommodating differences in item functioning across groups. Second, for constructs where invariance is largely supported, we estimate factor scores imposing equality of parameters across groups. Third, for transparency and comparability with existing work, we construct indices based on simple averages of item scores.

In the main analysis, we report results for child development outcomes based on averaged indices, reflecting both their simplicity and their alignment with standard practice. For parental investment measures—where guidance from test developers is more limited and several measures are constructed by the authors—we instead use factor scores estimated under the assumption of invariance across groups.

We assess the robustness of our results to these alternative measurement choices by re-estimating the main specifications using the different sets of measures described above. [include tables with alternative results] The estimates are quantitatively and qualitatively similar across specifications, including when using factor scores derived from models allowing for partial invariance, indicating that our conclusions are not sensitive to the choice of aggregation method.

Table A.5: Measurement invariance tests: Child outcomes

Construct	Test	p-value	Conclusion	Parameters freed
Cognitive skill (overall)	Metric vs. Configural	0.004	Metric invariance rejected	
	Partial metric vs. Configural	0.388	Partial metric invariance retained	Factor loadings: Extra number, IDELA 2, IDELA 17, IDELA 18
	Scalar vs. Configural	0.347	Scalar invariance not rejected	
Emergent literacy	Metric vs. Configural	0.114	Metric invariance not rejected	
	Scalar vs. Configural	0.035	Scalar invariance rejected	
	Partial scalar vs. Configural	0.165	Partial scalar invariance retained	Intercepts: IDELA 18
Emergent numeracy	Metric vs. Configural	0.031	Metric invariance rejected	
	Partial metric vs. Configural	0.216	Partial metric invariance retained	Factor loading: IDELA 2
	Scalar vs. Configural	0.441	Scalar invariance not rejected	
Executive function	Metric vs. Configural	0.902	Metric invariance not rejected	
	Scalar vs. Configural	0.758	Scalar invariance not rejected	
Emotional awareness	Metric vs. Configural	0.978	Metric invariance not rejected	
	Scalar vs. Configural	0.837	Scalar invariance not rejected	
Emotional and behavioural problems (overall)	Metric vs. Configural	0.259	Metric invariance not rejected	
	Scalar vs. Configural	0.042	Scalar invariance rejected	
	Partial scalar vs. Configural	0.152	Partial scalar invariance retained	Intercepts: SDQ attends
Externalising behaviours	Metric vs. Configural	0.000	Metric invariance rejected	
	Partial metric vs. Configural	0.074	Partial metric invariance retained	Factor loadings: SDQ steals, attends, fidgety, lies, reflect, obey
	Scalar vs. Configural	0.022	Scalar invariance rejected	
	Partial scalar vs. Configural	0.168	Partial scalar invariance retained	Intercepts: SDQ attends, restless, fidgety, reflect
Internalising behaviours	Metric vs. Configural	0.666	Metric invariance not rejected	
	Scalar vs. Configural	0.381	Scalar invariance not rejected	

Notes: The table reports p-values from likelihood ratio tests comparing nested measurement models across groups. Each test evaluates the null hypothesis that the more constrained model fits the data as well as the less constrained model. A high p-value indicates that the corresponding level of invariance is not rejected. The "Conclusion" column summarises the outcome of each test. When full invariance is rejected, we consider models allowing for partial invariance, in which selected factor loadings and/or intercepts are freed across groups. These parameters are chosen based on score tests and are reported in the final column.

Table A.6: Measurement invariance tests: Parental outcomes

Construct	Test	p-value	Conclusion	Parameters freed
Parental knowledge of child development	Metric vs. Configural	0.348	Metric invariance not rejected	
	Scalar vs. Configural	0.163	Scalar invariance not rejected	
Positive parenting behaviours	Metric vs. Configural	0.458	Metric invariance not rejected	
	Scalar vs. Configural	0.073	Scalar invariance not rejected	
	Partial scalar vs. Configural	0.176	Partial scalar invariance retained	Intercepts: signs of affection, convey positive feelings, keep child in visual range
	Strict vs. Configural	0.000	Strict invariance rejected	
Negative parenting behaviours	Metric vs. Configural	0.000	Metric invariance rejected	
	Partial metric vs. Configural	0.148	Partial metric invariance retained	Factor loadings: hit, spank, called names, hit, shouted at (activity), shouted at (interview), anger, intrusive behaviour, shook (interview), gave something to do, slap, disinterested, shook (activity)
	Partial scalar vs. Configural	0.271	Partial scalar invariance retained	Intercepts: hit, spank, called names, hit, shouted at (activity), shouted at (interview), anger, intrusive behaviour, shook (interview), gave something to do, slap, disinterested, shook (activity)
	Strict vs. Configural	0.000	Strict invariance rejected	
Depressive symptoms	Metric vs. Configural	0.298	Metric invariance not rejected	
	Scalar vs. Configural	0.000	Scalar invariance rejected	
	Partial scalar vs. Configural	0.224	Partial scalar invariance retained	Intercepts: always tired, poor digestion, difficulty with daily activities, frightened, sleep problems, tired easily
	Strict vs. Configural	0.492	Strict invariance not rejected	
Self-esteem	Metric vs. Configural	0.630	Metric invariance not rejected	
	Scalar vs. Configural	0.016	Scalar invariance rejected	
	Partial scalar vs. Configural	0.315	Partial scalar invariance retained	Intercepts: proud
	Strict vs. Configural	0.369	Strict invariance not rejected	

Notes: The table reports p-values from likelihood ratio tests comparing nested measurement models across groups. Each test evaluates the null hypothesis that the more constrained model fits the data as well as the less constrained model. A high p-value indicates that the corresponding level of invariance is not rejected. The "Conclusion" column summarises the outcome of each test. When full invariance is rejected, we consider models allowing for partial invariance, in which selected factor loadings and/or intercepts are freed across groups. These parameters are chosen based on score tests and are reported in the final column.

B Additional Figures and Tables

Table B.1: Attrition: household and primary carer characteristics

	Endline sample		Attrited sample		p-value	p-value	
	Control	Treated	Control	Treated	attrit/endline	diff treated/control	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Households							
Household size	8.965	10.321	9.992	10.367	0.002	0.250	2407
Number of children 16 or under	3.676	4.417	4.205	4.474	0.009	0.602	2407
Number of children 6 or under	1.332	1.617	1.536	1.639	0.045	0.667	2407
Wealth index (z-score)	-0.212	0.025	0.025	-0.025	0.067	0.361	2394
Main income source:							
Farming own land	0.359	0.446	0.435	0.438	0.044	0.842	2407
Wage work	0.328	0.274	0.276	0.284	0.665	0.221	2407
Livestock	0.047	0.033	0.035	0.033	0.036	0.075	2407
Profits from small enterprise	0.219	0.183	0.183	0.191	0.254	0.815	2407
B. Primary carers							
Age	36.648	36.435	36.092	36.833	0.687	0.744	2407
Target child's biological mother	0.656	0.790	0.789	0.763	0.000	0.949	2407
Some education	0.289	0.196	0.211	0.201	0.004	0.575	2407
Illiterate	0.859	0.926	0.915	0.923	0.003	0.589	2407
Raven's score	45.280	43.178	43.094	43.717	0.385	0.480	2407
Depression score	6.949	8.340	8.332	8.048	0.019	0.492	2407
Self-esteem score	18.637	17.944	18.105	17.928	0.000	0.014	2407

Notes: The sample is households and primary caregivers in the baseline sample. Columns (1) and (2) report means of baseline characteristics among observations in the endline sample, by treatment status. Columns (3) and (4) report the same means among observations that attrited at endline. We run a regression of each characteristic on an indicator for attrition (not being in the endline sample), a treatment indicator, and their interaction. Column (5) reports the p-value on the attrition indicator, and Column (6) reports the p-value on the interaction. Column (7) reports the full sample size.

Table B.2: Attrition: target child characteristics

	Endline sample		Attrited sample		p-value	p-value	
	Control	Treated	Control	Treated	attrit/endline	diff treated/control	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age (months)	55.311	56.325	56.209	56.226	0.104	0.487	2405
Gender	0.551	0.492	0.498	0.499	0.081	0.483	2407
Currently in school	0.766	0.781	0.779	0.780	0.870	0.734	2407
In the last 30 days:							
Any cough	0.512	0.549	0.567	0.522	0.975	0.281	2406
Any stomach pain	0.477	0.508	0.532	0.478	0.677	0.122	2406
Any high fever	0.340	0.340	0.359	0.321	0.651	0.522	2407
Any diarrhea	0.287	0.320	0.313	0.321	0.116	0.251	2395

Notes: The sample is the target children observed in the baseline sample. Columns (1) and (2) report means of baseline characteristics among observations in the endline sample, by treatment status. Columns (3) and (4) report the same means among observations that attrited at endline. We run a regression of each characteristic on an indicator for attrition (not being in the endline sample), a treatment indicator, and their interaction. Column (5) reports the p-value on the attrition indicator, and Column (6) reports the p-value on the interaction. Column (7) reports the full sample size.

Table B.3: Baseline balance: households and primary carer characteristics

	Full sample	Control	Treatment	p-value	Adj. p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Households						
Household size	10.321	10.497	10.150	0.208	0.671	2151
	(6.369)	(6.756)	(5.967)			
Number of children 16 or under	4.417	4.573	4.264	0.045	0.228	2151
	(3.568)	(3.800)	(3.321)			
Number of children 6 or under	1.617	1.676	1.560	0.140	0.561	2151
	(1.825)	(1.893)	(1.755)			
Wealth index (z-score)	0.025	-0.025	0.074	0.283	0.729	2139
	(2.114)	(2.067)	(2.159)			
Main income source:						
Farming own land	0.446	0.447	0.445	0.933	0.990	2151
Wage work	0.274	0.274	0.274	0.998	0.994	2151
Livestock	0.033	0.034	0.031	0.721	0.970	2151
Profits from small enterprise	0.183	0.189	0.177	0.458	0.882	2151
B. Primary carers						
Age	36.435	36.948	35.935	*0.050*	*0.251*	2151
	(11.996)	(12.156)	(11.823)			
Target child's biological mother	0.790	0.774	0.806	*0.063*	*0.291*	2151
Some education	0.196	0.192	0.200	*0.652*	*0.896*	2151
Illiterate	0.926	0.930	0.922	*0.466*	*0.882*	2151
Raven's score (%)	43.178	43.473	42.890	*0.413*	*0.882*	2151
	(16.524)	(17.090)	(15.956)			
Depression score (0-20)	8.340	8.189	8.486	*0.193*	*0.593*	2151
	(5.288)	(5.278)	(5.295)			
Self-esteem score (0-30)	17.944	17.912	17.975	*0.684*	*0.896*	2151
	(3.584)	(3.656)	(3.514)			
Knowledge of child development (z-score)	-0.021	0.000	-0.042	*0.315*	*0.754*	2151
	(0.965)	(1.000)	(0.931)			
Play materials (z-score)	0.088	0.000	0.173	*0.000*	*0.004*	2151
	(1.074)	(1.000)	(1.136)			
Any play activities with child (0/1)	0.132	0.120	0.143	*0.108*	*0.445*	2151

Notes: The sample includes households and primary carers whose children attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. Columns (1)–(3) report baseline means for the full sample, control group, and treatment group (with standard deviations in parentheses underneath the means for non-binary variables). Columns (4) and (5) report the p-value and adjusted p-value for the difference in means between control and treatment; adjusted p-values use the Romano and Wolf (2005) procedure with 500 bootstrap samples. Column (6) reports the full sample size.

Table B.4: Baseline balance: target child characteristics

	Full sample	Control	Treatment	p-value	Adj. p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Characteristics						
Age (months)	56.325	56.276	56.372	0.816	0.996	2151
	(9.499)	(9.481)	(9.521)			
Gender	0.492	0.496	0.489	0.754	0.996	2151
Currently in school	0.781	0.783	0.779	0.809	0.996	2151
In the last 30 days:						
Any cough	0.549	0.525	0.572	0.025	0.144	2150
Any stomach pain	0.508	0.481	0.535	0.011	0.068	2150
Any high fever	0.340	0.319	0.361	0.036	0.154	2151
Any diarrhea	0.320	0.321	0.320	0.975	0.996	2144
B. Cognitive outcomes (z-scores)						
Aggregate score	-0.002	0.000	-0.004	*0.929*	*0.998*	2151
	(1.052)	(1.000)	(1.100)			
Emergent literacy	-0.005	0.000	-0.011	*0.808*	*0.988*	2151
	(1.031)	(1.000)	(1.061)			
Emergent numeracy	0.017	0.000	0.033	*0.462*	*0.910*	2151
	(1.045)	(1.000)	(1.088)			
Executive function	-0.029	0.000	-0.057	*0.202*	*0.681*	2151
	(1.034)	(1.000)	(1.065)			
C. Socioemotional outcomes (z-scores)						
Emotional awareness	0.018	0.000	0.036	*0.415*	*0.910*	2151
	(1.038)	(1.000)	(1.073)			
Strengths and Difficulties:						
Emotional and behavioural problems	0.003	0.000	0.005	*0.907*	*0.998*	2151
	(0.987)	(1.000)	(0.975)			
Externalising behaviours	0.012	0.000	0.024	*0.568*	*0.936*	2151
	(0.983)	(1.000)	(0.967)			
Internalising behaviours	-0.006	0.000	-0.013	*0.767*	*0.988*	2151
	(0.992)	(1.000)	(0.985)			

Notes: The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. Columns (1)–(3) report baseline means for the full sample, control group, and treatment group (with standard deviations in parentheses underneath the means for non-binary variables). Columns (4) and (5) report the p-value and adjusted p-value for the difference in means between control and treatment; adjusted p-values use the Romano and Wolf (2005) procedure with 500 bootstrap samples. Column (6) reports the full sample size.

Table B.5: Baseline balance using sample of PMs: households and primary carers characteristics

	Full sample	Control	Treatment	p-value	Adj. p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Households						
Household size	10.288	10.713	10.008	*0.215*	*0.693*	437
	(5.809)	(6.237)	(5.502)			
Number of children 16 or under	4.494	4.816	4.281	*0.096*	*0.399*	437
	(3.283)	(3.525)	(3.101)			
Number of children 6 or under	1.595	1.672	1.544	*0.441*	*0.814*	437
	(1.706)	(1.767)	(1.666)			
Wealth index (z-score)	-0.018	0.106	-0.101	*0.288*	*0.784*	436
	(1.987)	(1.969)	(1.999)			
Main income source:						
Farming own land	0.465	0.494	0.445	*0.312*	*0.784*	437
Wage work	0.238	0.201	0.262	*0.142*	*0.557*	437
Livestock	0.034	0.040	0.030	*0.582*	*0.814*	437
Profits from small enterprise	0.178	0.178	0.179	*0.988*	*0.984*	437
B. Primary carers						
Age	36.586	37.402	36.046	*0.184*	*0.800*	437
	(10.450)	(11.087)	(9.991)			
Target child's biological mother	0.842	0.828	0.852	*0.500*	*0.972*	437
Some education	0.183	0.190	0.179	*0.773*	*0.972*	437
Illiterate	0.936	0.920	0.947	*0.256*	*0.820*	437
Raven's score (%)	42.429	42.385	42.459	*0.963*	*0.972*	437
	(16.065)	(17.782)	(14.856)			
Depression score (0–20)	0.092	0.132	0.066	*0.493*	*0.972*	437
	(0.985)	(1.008)	(0.971)			
Self-esteem score (0–30)	-0.055	0.085	-0.147	*0.014*	*0.098*	437
	(0.973)	(1.073)	(0.890)			
Knowledge of child development (z-score)	-0.066	-0.033	-0.088	*0.570*	*0.972*	437
	(0.980)	(1.092)	(0.900)			
Play materials (z-score)	0.136	-0.001	0.226	*0.042*	*0.259*	437
	(1.140)	(1.009)	(1.212)			
Any play activities with child (0/1)	0.121	0.098	0.137	*0.220*	*0.820*	437

Notes: The sample includes participating mothers' households and primary carers whose children attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. Columns (1)–(3) report baseline means for the full sample, control group, and treatment group (with standard deviations in parentheses underneath the means for non-binary variables). Columns (4) and (5) report the p-value and adjusted p-value for the difference in means between control and treatment; adjusted p-values use the Romano and Wolf (2005) procedure with 500 bootstrap samples. Column (6) reports the full sample size.

Table B.6: Baseline balance using sample of PMs: target child characteristics

	Full sample	Control	Treatment	p-value	Adj. p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Characteristics						
Age (months)	55.684	55.713	55.665	*0.960*	*0.996*	437
	(9.649)	(9.410)	(9.821)			
Gender	0.497	0.506	0.490	*0.756*	*0.978*	437
Currently in school	0.835	0.833	0.837	*0.931*	*0.996*	437
In the last 30 days:						
Any cough	0.558	0.511	0.589	*0.109*	*0.441*	437
Any stomach pain	0.533	0.460	0.582	*0.012*	*0.074*	437
Any high fever	0.380	0.339	0.407	*0.154*	*0.513*	437
Any diarrhea	0.338	0.356	0.326	*0.509*	*0.934*	435
B. Cognitive outcomes (z-scores)						
Aggregate score	0.048	0.109	0.008	*0.340*	*0.760*	437
	(1.087)	(1.098)	(1.080)			
Emergent literacy	0.056	0.087	0.035	*0.616*	*0.920*	437
	(1.051)	(1.065)	(1.044)			
Emergent numeracy	0.052	0.092	0.026	*0.526*	*0.910*	437
	(1.073)	(1.059)	(1.084)			
Executive function	-0.001	0.099	-0.067	*0.104*	*0.415*	437
	(1.046)	(1.058)	(1.035)			
C. Socioemotional outcomes (z-scores)						
Emotional awareness	0.008	-0.008	0.018	*0.803*	*0.920*	437
	(1.081)	(1.073)	(1.089)			
Strengths and Difficulties:						
Emotional and behavioural problems	0.010	0.031	-0.004	*0.718*	*0.920*	437
	(0.998)	(1.007)	(0.993)			
Externalising behaviours	0.028	-0.042	0.074	*0.235*	*0.629*	437
	(1.000)	(1.017)	(0.988)			
Internalising behaviours	-0.008	0.085	-0.070	*0.112*	*0.415*	437
	(0.997)	(1.022)	(0.977)			

Notes: The sample includes participating mothers' children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. Columns (1)–(3) report baseline means for the full sample, control group, and treatment group (with standard deviations in parentheses underneath the means for non-binary variables). Columns (4) and (5) report the p-value and adjusted p-value for the difference in means between control and treatment; adjusted p-values use the Romano and Wolf (2005) procedure with 500 bootstrap samples. Column (6) reports the full sample size.

Table B.7: Impacts on parental knowledge, play investments, and behaviors using IRT scores

	I. Knowledge		**II. Play Investments**		**III. Behaviors**	
	Child Development	Pre-school Quality	Play Materials	Any play activities	Positive Behaviors	Negative Behaviors
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment among PMs	0.423	0.198	-0.082	0.181	0.214	0.043
	(0.130)	(0.120)	(0.143)	(0.049)	(0.125)	(0.139)
p-val	*0.002*	*0.104*	*0.567*	*0.000*	*0.091*	*0.758*
Adj. p-val	*0.020*	*0.327*	*0.786*	*0.014*	*0.327*	*0.786*
Treatment among non-PMs	-0.033	0.184	-0.247	-0.031	0.040	0.022
	(0.102)	(0.059)	(0.108)	(0.036)	(0.072)	(0.103)
p-val	*0.749*	*0.003*	*0.024*	*0.390*	*0.578*	*0.832*
Adj. p-val	*0.912*	*0.014*	*0.082*	*0.772*	*0.878*	*0.912*
Difference in treatment	0.456	0.013	0.165	0.213	0.174	0.021
	(0.117)	(0.120)	(0.126)	(0.042)	(0.100)	(0.140)
p-val	*0.000*	*0.911*	*0.193*	*0.000*	*0.086*	*0.882*
Adj. p-val	*0.004*	*0.970*	*0.377*	*0.002*	*0.238*	*0.970*
Control mean, non-PMs	0.034	0.008	-0.118	0.316	0.009	0.036
Control mean, PMs	0.055	0.283	-0.057	0.364	0.177	-0.046
Baseline outcome	Yes	No	Yes	Yes	No	No
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151	2,151	2,151

Notes: All outcomes are measured as IRT scores standardised to have mean 0 and standard deviation 1 in the control group. The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented where standard errors are clustered at the community level, and p-values are reported in italics. Throughout we estimate fully interacted models that control for the baseline outcome (if available), strata fixed effects, child's age in months at endline, a gender indicator, and the number of clothing stores in the community, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother signed up to participate in the play schemes (during the intervention in the treatment group, in the following year in the control group). Control means are reported for each subgroup at the bottom of the table.

Table B.8: Parental knowledge about preschool quality

	Teacher discipline	**Rote learning**	**Parents' involvement**	**Buildings**
Treatment among PMs	0.007	0.102	-0.053	0.059
	(0.043)	(0.061)	(0.047)	(0.035)
p-val	*0.870*	*0.100*	*0.268*	*0.100*
Adj. *p*-val	*0.842*	*0.317*	*0.425*	*0.317*
Treatment among non-PMs	0.042	0.089	-0.023	0.021
	(0.030)	(0.027)	(0.028)	(0.016)
p-val	*0.162*	*0.001*	*0.421*	*0.195*
Adj. *p*-val	*0.365*	*0.006*	*0.365*	*0.365*
Difference in treatment	-0.035	0.012	-0.030	0.037
	(0.040)	(0.059)	(0.049)	(0.040)
p-val	*0.389*	*0.834*	*0.547*	*0.355*
Adj. *p*-val	*0.760*	*0.818*	*0.760*	*0.760*
Control Mean PMs	0.824	0.732	0.833	0.902
Control Mean non-PMs	0.753	0.623	0.820	0.856
Baseline outcome	No	No	No	No
Strata Fixed Effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented with standard errors clustered at the community level, and p-values are reported in italics. Throughout we estimate fully interacted models that control for the baseline outcome (if available), strata fixed effects, child's age in months at endline, a gender indicator, and the number of clothing stores in the community, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother signed up to participate in the play schemes (during the intervention in the treatment group, in the following year in the control group). At the foot of each column, we report the mean outcome among controls in each subgroup.

Table B.9: Impacts on play materials

	Home toys	Toy pack	Single toy	HH object used as toy	Music toy	Pack of blocks	Single block	Drawing toy pack	Single drawing toy	Toy to help moving	Dolls	Books	Toys to learn shapes	Picture charts	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Treatment PMs	-0.024	0.014	-0.034	-0.070	-0.031	0.003	0.000	0.017	-0.001	0.004	0.023	0.010	-0.015	0.014	0.030
	(0.075)	(0.007)	(0.038)	(0.061)	(0.037)	(0.003)	(0.000)	(0.016)	(0.036)	(0.060)	(0.058)	(0.025)	(0.012)	(0.008)	(0.013)
p-val	*0.746*	*0.045*	*0.371*	*0.251*	*0.408*	*0.300*	*1.000*	*0.280*	*0.989*	*0.942*	*0.691*	*0.699*	*0.197*	*0.077*	*0.025*
Treatment non-PMs	0.011	-0.007	-0.008	-0.079	-0.034	-0.001	-0.002	-0.018	-0.033	-0.066	-0.035	-0.035	-0.007	-0.003	0.017
	(0.045)	(0.005)	(0.027)	(0.030)	(0.026)	(0.002)	(0.004)	(0.012)	(0.032)	(0.039)	(0.036)	(0.019)	(0.005)	(0.004)	(0.010)
p-val	*0.799*	*0.178*	*0.757*	*0.011*	*0.192*	*0.418*	*0.724*	*0.126*	*0.307*	*0.091*	*0.324*	*0.065*	*0.148*	*0.519*	*0.105*
Difference in treatment	-0.036	0.021	-0.025	0.009	0.003	0.004	0.002	0.035	0.032	0.071	0.058	0.045	-0.009	0.017	0.013
	(0.063)	(0.008)	(0.046)	(0.062)	(0.031)	(0.003)	(0.004)	(0.019)	(0.037)	(0.054)	(0.051)	(0.022)	(0.011)	(0.008)	(0.015)
p-val	*0.571*	*0.012*	*0.579*	*0.888*	*0.913*	*0.192*	*0.724*	*0.070*	*0.390*	*0.197*	*0.260*	*0.047*	*0.451*	*0.034*	*0.400*
Control Mean PMs	0.286	0.009	0.204	0.600	0.119	0.002	0.000	0.037	0.112	0.336	0.249	0.034	0.009	0.011	0.014
Control Mean non-PMs	0.289	0.014	0.184	0.643	0.088	0.001	0.008	0.036	0.137	0.286	0.225	0.033	0.010	0.007	0.014
Baseline outcome	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151	2,151

Notes: The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented where standard errors are clustered at the community level, and *p-val*ues are reported in italics. Throughout we estimate fully interacted models that control for the baseline outcome (if available), strata fixed effects, child's age measured in months at endline, a gender indicator, and the number of clothing stores in the community, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother has signed up to participate in the play schemes (during the intervention in the treatment group, in the following year in the control group). At the foot of each column, we report the mean outcome among controls in each subgroup.

Table B.10: Impacts on play activities

	Read	Telling stories	Singing	Going out	Playing	Drawing	Playing with objects
	(16)	(17)	(18)	(19)	(20)	(21)	(22)
Treatment among PMs	0.062	0.111	0.025	0.015	0.029	0.031	0.020
	(0.020)	(0.031)	(0.025)	(0.036)	(0.011)	(0.013)	(0.015)
p-val	*0.003*	*0.001*	*0.328*	*0.683*	*0.009*	*0.025*	*0.192*
Treatment among non-PMs	-0.001	0.001	-0.023	-0.011	0.003	-0.010	-0.004
	(0.018)	(0.019)	(0.018)	(0.031)	(0.005)	(0.008)	(0.007)
p-val	*0.940*	*0.978*	*0.208*	*0.729*	*0.573*	*0.208*	*0.536*
Difference in treatment	0.063	0.111	0.048	0.026	0.026	0.041	0.025
	(0.023)	(0.028)	(0.025)	(0.031)	(0.012)	(0.014)	(0.018)
p-val	*0.009*	*0.000*	*0.062*	*0.409*	*0.041*	*0.006*	*0.162*
Control Mean non-PMs	0.089	0.131	0.086	0.102	0.016	0.020	0.015
Control Mean PMs	0.080	0.140	0.098	0.133	0.023	0.025	0.023
Baseline outcome	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151	2,151	2,151	2,151

Notes: The sample includes children who attended one of the pre-schools randomized into treatment or control at baseline and who were successfully tracked at endline. OLS estimates are presented where standard errors are clustered at the community level, and *p-val*ues are reported in italics. Throughout we estimate fully interacted models that control for the baseline outcome (if available), strata fixed effects, child's age measured in months at endline, a gender indicator, and the number of clothing stores in the community, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother has signed up to participate in the play schemes (during the intervention in the treatment group, in the following year in the control group). At the foot of each column, we report the mean outcome among controls in each subgroup.

Table B.11: Impacts on children's cognitive outcomes using factor scores

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
Treatment	0.091	0.071	0.094	0.121
	(0.052)	(0.054)	(0.050)	(0.051)
p-val	*0.086*	*0.193*	*0.064*	*0.020*
Adj. *p*-val		*0.110*	*0.062*	*0.030*
Baseline outcome	0.487	0.359	0.439	0.257
	(0.025)	(0.025)	(0.023)	(0.022)
p-val	*0.000*	*0.000*	*0.000*	*0.000*
Control mean endline	0.000	-0.000	0.000	-0.000
Strata Fixed Effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

Notes: All outcomes measured as IRT scores standardised to have mean 0, std 1 in the control group. See Table 3 note for more details.

Table B.12: Impacts of intervention on children's socio-emotional outcomes using factor scores

	Emotional Awareness	**Emotional Behavioural Problems**		
		Overall	**Externalising Behaviours**	**Internalising Behaviours**
	(1)	(2)	(3)	(4)
Treatment	0.041	-0.096	-0.099	-0.072
	(0.074)	(0.064)	(0.054)	(0.072)
p-val	*0.584*	*0.137*	*0.068*	*0.316*
Adj. *p*-val			*0.082*	*0.224*
Baseline outcome	0.233	0.048	0.063	0.044
	(0.024)	(0.026)	(0.022)	(0.029)
p-val	*0.000*	*0.068*	*0.005*	*0.133*
Control mean endline	-0.000	0.000	-0.000	0.000
Strata Fixed Effects	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,151

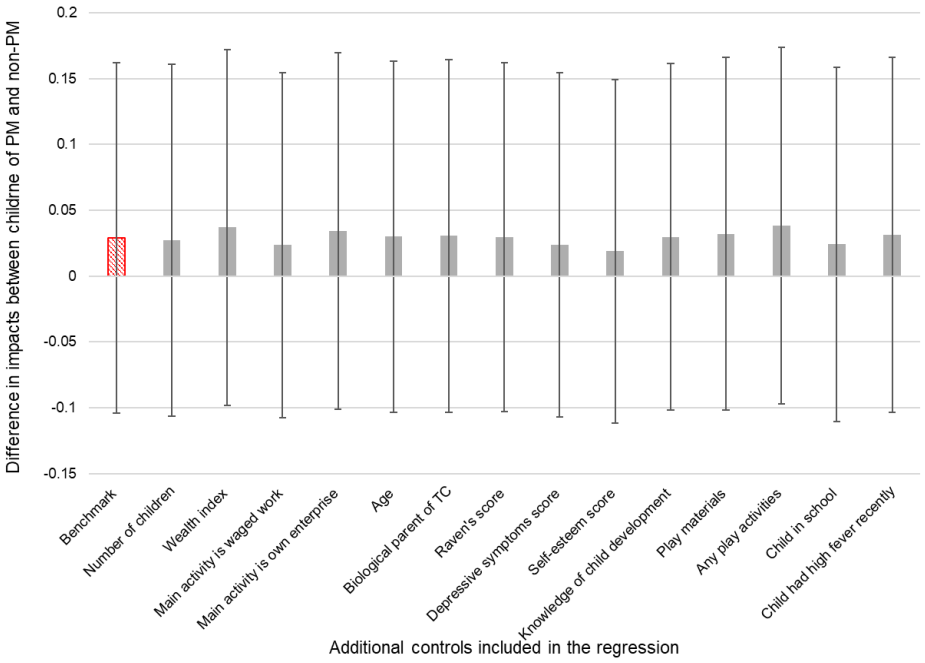
Notes: Outcomes are standardised to have mean 0, std 1 in the control group. See Table 6 note for more details.

Table B.13: Impacts of intervention on primary outcomes: Controls selected using post-double selection Lasso

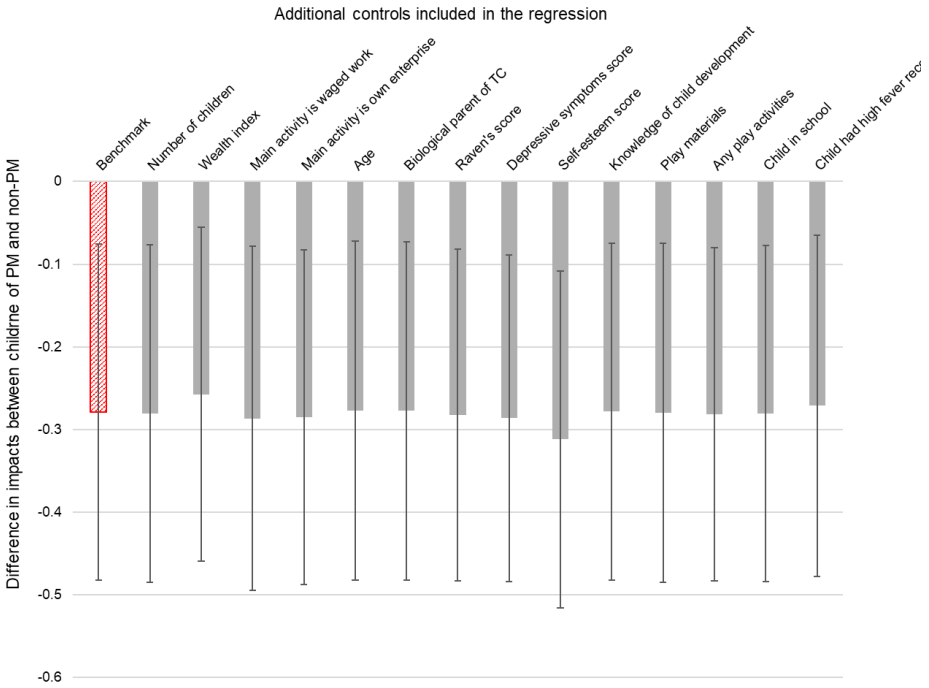
	Cognition	**Emotional Awareness**	**Emotional and behavioural problems**
	(1)	(2)	(3)
Treatment	0.1006**	0.0383	-0.0993*
	(0.0505)	(0.0704)	(0.0589)
p-val	*0.046*	*0.587*	*0.092*
Control mean endline	0.000	0.000	0.000
Strata Fixed Effects	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes
Observations	2,151	2,151	2,151

Notes: ***, **, and * indicate significance at the 1, 5, and 10 percent level respectively. OLS estimates are reported with standard errors clustered at the village level in parentheses; p-values are reported in italics. Controls are selected using the post-double selection Lasso method (Belloni et al., 2014).

Figure B.1: Difference in intervention impact on children's outcomes between children of PM and non-PM: Robustness check



(a) **Cognition (overall score)**



(b) **Emotional and behavioural problems**

Notes: The bars shown in each graph plot the coefficient measuring the difference in the treatment effect of the LM intervention on cognitive skill (subfigure a) and emotional and behavioural problems (subfigure b) between children of PM and non-PM in different specifications. The first bar in red plots this coefficient in the benchmark specification (shown in Table 5 and Table 7 respectively). Each of the grey bars to the right of the red bar plots the same coefficient in specifications also controlling for the variable shown at the bottom or top of the bar and its interaction with the PM indicator. The error bars plot the 95% confidence intervals for the coefficient estimate.

C Structured Ethics Appendix following Asiedu et al. (2021)

This appendix follows the nine-part structure proposed by Asiedu et al. (2021) for social science studies involving primary data collection.

C.1 Policy equipoise and scarcity

At the start of the study, there was genuine uncertainty about whether the Lively Minds model would improve child development when delivered through the public preschool system in rural Ghana, and whether any benefits would justify the demands it placed on schools and on volunteering mothers. This uncertainty was especially important because the model relied on low-cost implementation through existing government structures rather than on adding salaried staff or large material inputs.

The study was conducted in two districts where the program had not yet been introduced. Capacity and implementation constraints meant that the program could not be rolled out to all eligible schools at once. School-level randomization, therefore, served as a fair way to phase in access while generating evidence on effectiveness. Control schools continued to receive the standard public preschool provision during the study period, and the program was rolled out to control schools after endline data collection in November 2018.

C.2 Role of the researchers with respect to implementation

The researchers were active in the design of the evaluation, including the randomized rollout and the measurement strategy. They were not, however, the implementers of the program. Program delivery was carried out through the Ghana Education Service (part of the Ministry of Education, Ghana), preschool teachers, and Lively Minds NGO. Data collection was conducted by Innovations for Poverty Action (IPA).

This separation matters ethically. The researchers helped design the evaluation and, therefore, shared responsibility for the choice to study the program through a randomized phase-in. At the same time, routine delivery to schools, recruitment and training of mothers, and field data collection were handled by institutions other than the academic research team.

C.3 Potential harms to participants or nonparticipants from the intervention or policy

The main ex ante ethical concern was that the program depended on unpaid volunteer labor from low-income rural women who already had substantial domestic, care, and farm responsibilities. A second concern was that a greater focus on the target child could create strains within the household or reduce attention to siblings. A third concern was that adding a new classroom model might impose burdens on teachers or create tension between participating and non-participating mothers.

Several features of the program reduced these risks. Participation by mothers was voluntary. The intended time commitment was about one to two hours per week. Training was adapted for low-literacy participants. Participating mothers also received monthly parenting and well-being workshops, and classroom activities were supervised by preschool teachers. In addition, control schools continued to receive standard preschool services during the trial rather than losing access to an existing service.

The main paper also examines some of these risks directly. It finds no evidence of adverse effects on participating mothers' mental health or self-esteem, and no evidence of adverse spillovers to other children in the household.

C.4 Potential harms from data collection or research protocols

The main risks from research participation came from household surveys, interviews with caregivers and teachers, and direct child assessments. These risks included loss of privacy, discomfort when answering personal questions, fatigue, and the handling of information about young children and caregiver well-being. Because the study involved young children and many respondents with limited formal education, it was important that research procedures be understandable and low burden.

Data collection was conducted by IPA under protocols approved by the Ghana Health Service Ethics Review Committee (GHSERC012/07/17), IPA (14340), and University College London (10167/001). Adult respondents were surveyed under approved consent procedures, and child assessments were conducted under the study procedures approved by those ethics bodies. These approvals also covered confidentiality and data-management procedures. The study reports

results in aggregate form rather than in a way that identifies individual respondents.

The randomization itself also had ethical implications. Assignment was at the school level rather than at the individual child level, which reduced the scope for differential treatment of children within the same preschool. Control schools later received the program, which also reduced the ethical cost of temporary nonreceipt during the evaluation period.

C.5 Financial and reputational conflicts of interest

This project was funded by the Global Innovations Fund, the Jacobs Foundation, and the ESRC, and funders had no role in study design, data collection, data analysis, data interpretation, or the writing of results. There were no direct financial conflict of interest tied to the study findings.

The more relevant concern here is reputational rather than financial. The study was conducted in collaboration with Lively Minds NGO and with government partners who had a clear interest in understanding whether the program worked and whether it could be scaled. That creates a reason to be especially transparent about design and analysis choices. Important safeguards in this study include multiple ethics approvals, trial registration, a pre-analysis plan, and data collection by IPA rather than by the implementing NGO.

C.6 Intellectual freedom

We are not aware of any contractual restrictions that limited the researchers' ability to report the results. The paper names the implementing NGO and the government partners directly, and it states clearly that funders had no role in the design, collection, analysis, interpretation, or writing of the study. The trial was also registered before the main results were reported.

C.7 Feedback to participants or communities

Because this intervention was embedded in the public preschool system, an important ethical obligation was to ensure that results reached the institutions responsible for policy and implementation. Our findings informed later expansion of the program through the Ghana Education Service, which indicates that results were communicated to the main policy and implementation stakeholders.

We, however, did not have a formal household-level dissemination exercise for all study participants. In this setting, that is understandable: the study covered many rural communities, literacy levels were low, and the most policy-relevant decisions were at the school, district, and government levels. The most appropriate feedback channel was therefore likely through the Ghana Education Service, schools, and Lively Minds rather than through individualized written feedback to households.

C.8 Foreseeable misuse of research results

The clearest foreseeable misuse of these results would be to treat them as a general justification for shifting the burden of preschool quality improvement onto unpaid women, or for assuming that any volunteer-based model is low risk and scalable. That would be an overreading of the evidence. The results come from a specific model with a limited weekly time commitment, simple training adapted to low-literacy participants, teacher supervision, and explicit attention to mothers' well-being.

Another possible misuse would be to generalize the results too broadly beyond rural Northern Ghana, or to ignore the paper's own discussion of implementation quality, hidden costs, and the conditions required for scale-up. For this reason, it is important that the findings be communicated together with the program's implementation details, evidence on mothers' well-being, and the limits to external validity.

C.9 Other ethics issues

Two further points are worth noting. First, the study involved young children and socially disadvantaged mothers in poor rural communities. That makes respectful field procedures, low-burden participation, and plain communication especially important. Second, the evaluation was registered in both the ISRCTN registry (ISRCTN21215509) and the AEA RCT Registry (AEARCTR-0002777), which helped strengthen transparency around the study design and analysis.