IFS

Institute for Fiscal Studies

Bet Caeyers
Lina Cardona-Rosa
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[‖]

August 4, 2026

Abstract

Preschool enrollment has expanded rapidly in low-income countries, but quality has lagged, and evidence on how governments can improve it remains limited. We evaluate *Lively Minds*, a scalable early childhood program adopted and implemented by Ghana's Ministry of Education that trains community mothers to deliver structured, play-based activities in public preschools and support children's learning at home. Using a cluster randomized controlled trial in 80 rural preschools, we find that the program increases cognitive development by 0.12 standard deviations (SD) and reduces problem behaviors by 0.10 SD on average. Cognitive gains are similar for children of participating and non-participating mothers, whereas behavioral improvements are concentrated among children of participating mothers. Consistent with this pattern, only participating mothers improve their knowledge of child development and engage in more developmentally supportive interactions with their children at home. These findings suggest that engaging parents - even those with limited formal education — can generate additional benefits beyond those of approaches that rely solely on teachers and paraprofessionals. A cost-benefit analysis indicates that the program's private benefits substantially exceed its costs, with no adverse effects on mothers' well-being.

JEL Codes: J13, I10, I20
Keywords: Early childhood development, preschool quality, parenting, scalability.

*We are very grateful to the 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 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 Alison Andrew, Pedro Carneiro, Juanna Joensen, Costas Meghir, Imran Rasul, and seminar and conference participants at Boston College, Ghana Education Services, Institute for Fiscal Studies, Leeds, Liverpool, Paris School of Economics, Royal Holloway, Yale, Zurich, INSEAD RCT Days, IFS-Oxfam Workshop, Society for Research in Child Development, Thrive Conference, London Behavioral and Experimental Economics Workshop, Lancaster University Economics Workshop, Nordic Conference in Development Economics 2026, IFS Development Economics Conference, and Research Workshop on Family and Health Economics in London 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. Email: Bet.Caeyers@cmi.no

[‡]World Bank. Email: lcardonasosa@worldbank.org

[§]IFS; HCEO; IZA@LISER; CESifo. Email: sarah_c@ifs.org.uk

[¶]University of Manchester; IFS. Email: sonya.krutikova@manchester.ac.uk

[‖]Royal Holloway; IFS; IZA@LISER. Email: Abu.Siddique@rhul.ac.uk

1 Introduction

Children growing up in low- and lower-middle-income countries (LLMIC) face a host of barriers to reaching their developmental potential, including poverty, malnutrition, low parental education, and inadequate stimulation at home. Intervening early can yield high private and social returns, while early skill deficits are difficult and costly to remediate later in life (Currie and Almond, 2011; Heckman and Mosso, 2014; Attanasio et al., 2022; Andrew et al., 2024). In light of this evidence, access to pre-primary education in LLMIC has expanded substantially over the past two decades, with enrollment rising from 13% in 2000 to over 30% in 2018 (UNICEF, 2019b). However, governments are struggling to deliver services of adequate quality, owing to shortages of trained teachers, high teacher turnover and absenteeism, large class sizes, weak pedagogical practices, limited funding, and weak implementation systems (The World Bank, 2024). Scaling preschool access without improving service quality risks generating enrollment without learning, with children completing preschool without mastering basic competencies (World Bank, 2018a). A critical challenge for many LLMIC is therefore 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, Ghana, 2025). Yet, learning outcomes remain poor, particularly in rural and high-poverty areas; a recent study found that fewer than half of preschool-aged children are developmentally on track in literacy and numeracy (Lim et al., 2023). Challenges in delivering high-quality preschool education are compounded by constraints in the home environment, where children often receive insufficient care and stimulation because of low parental education, financial and time constraints, and the widespread use of harsh discipline (UNICEF, 2019a; Lim et al., 2023). For instance, 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.

Against this background, we experimentally test a novel approach to improving children's learning and development in public preschools. The approach is predicated on the idea that parents of young children are a key underutilized resource for supporting early learning and development in LLMICs, particularly in rural communities where needs are greatest. The

program was developed by the NGO Lively Minds and is currently being scaled up across rural Ghana by the Ghana Education Service.[1] It trains mothers to serve as regular volunteers in public preschools, where they deliver structured, play-based learning activities during school hours. By relying on community mothers rather than teaching assistants, the program has the potential to improve children's outcomes through two complementary channels. First, it aims to directly improve the preschool learning environment by increasing children's exposure to structured, play-based instruction and more individualized attention, which may strengthen cognitive skills (Grantham-McGregor et al., 1991; Currie, 2001; Hirsh-Pasek et al., 2009; Cascio and Schanzenbach, 2013). Structured, adult-supervised play and routines such as turn-taking may also improve socioemotional development by strengthening self-regulation and inhibitory control (Diamond and Lee, 2011; Healey and Healey, 2019). Second, implementation itself may increase participating mothers' parenting human capital. Through training and repeated classroom practice, mothers acquire knowledge, skills, and habits that may improve interactions with their own children at home, generating additional benefits through a more cognitively stimulating and responsive home environment.

We study the impact of this approach through a cluster-randomized controlled trial (cRCT) conducted across a random sample of 80 public 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 Ghana Education Service, with support from Lively Minds NGO, 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. Trained mothers were then 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, these mothers attended monthly parenting and well-being workshops run by the same preschool teachers.

While random assignment of the *Lively Minds* program across preschools identifies its average treatment effect, a unique feature of our design allows us to also estimate treatment effects

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

among mothers who participated in the program. Specifically, we conducted identical mobilization and sign-up processes in treatment and control communities, except that mothers in control communities were invited to enroll for participation in the following year, after the RCT, when the program was scheduled to be rolled out there. Hereafter, we refer to mothers who enrolled to participate in the program as *participating mothers*. Provided that participating mothers in treatment and control communities are comparable, a condition we validate through extensive balance checks, this design allows us to estimate the program's impacts on participating mothers separately from the impacts on non-participating mothers, thereby shedding light on the channels through which the program operates.

Our primary outcome of interest is children's cognitive and socioemotional development. Cognitive development is measured through 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 emotional and behavioral problems.

Our analysis leads to three main findings. First, despite being relatively low-intensity and implemented within a highly constrained education system, the program is effective at improving the early-childhood development knowledge and skills of participating mothers, the great majority of whom have no formal education and little prior knowledge of child development. These improvements are an important prerequisite for achieving improvements in children's outcomes. Participating mothers' knowledge of child development increases by nearly half a standard deviation (SD), and enumerator observations of mother-child play suggest some improvement in the quality of their interactions. These effects are distinct from the program's broader effects on non-participating parents in the community, who primarily update their beliefs about readily observable aspects of preschool quality rather than broader child development and do not change the way that they engage with their children.

Second, we find an economically meaningful improvement in the cognitive skills of children who, at baseline, were due to start preschool in the treatment group. The aggregate cognitive development index significantly increases by 0.12 SD, with gains concentrated in emergent numeracy and executive function. These gains are similar across gender, robust to alternative ways of constructing the cognitive development index that reduce measurement error, and are

concentrated in the middle of the distribution.

The program also reduces children's emotional and behavioral difficulties, with a marginally significant decline of nearly 10% of a SD overall. This effect lies within the 0.1–0.15 SD range reported in a meta-analysis of the effects of preschool interventions on socioemotional skills (Holla et al., 2021). It is driven by a reduction in externalizing behaviors (disruptive, hyperactive behaviors) among children with higher baseline difficulties. In line with the wider literature, it is concentrated among boys, who are also more likely to exhibit externalizing behaviors. These patterns align with the program's emphasis on structured, adult-supervised small-group activities and routines, an approach shown in the literature to reduce externalizing behaviors effectively.

Third, we find evidence that the program generates additional benefits for the children of participating mothers. While cognitive gains are similar across children of participating and non-participating mothers, socioemotional improvements are concentrated among children of participating mothers. Moreover, participating mothers in the treatment group report investing substantially more time in play and learning activities with their child at home than the participating mothers in the control group. These changes are not observed among non-participating mothers. Indeed, among this group we find a *reduction* in investment in play materials, perhaps reflecting a substitution response to perceived improvements in preschool quality. Together, these findings suggest that mothers' participation likely generates additional impacts on children through the home environment, in line with the aims of the program. However, this interpretation should be treated with caution. Although selection into participation is comparable across treatment and control groups, participation itself is not randomly assigned. Consequently, differences in the causal program impacts between participating and non-participating mothers may reflect both the causal effect of participation and unobserved selection on gains.

We estimate the program's cost-effectiveness by projecting the earnings gains implied by improvements in child development and comparing their present discounted value with program costs, following the approach of Ganimian et al. (2024). Under our most conservative assumptions, the projected present discounted value of private earnings gains is $295 per child, compared with a program cost of $37 per child (both in implementation year – 2017 - US dol-

lars), implying that the benefit is nearly 8 times as high as the cost. [2] From a public-finance perspective, the program would pay for itself fiscally if the government recovered 12.5% of the projected earnings gains through future tax revenues.

A key driver of low costs is the program's reliance on mothers volunteering their time rather than additional salaried staff. However, this design feature may have hidden costs if it has adverse impacts on the welfare of the often already vulnerable group of women who participate. Reassuringly, we find no evidence of such adverse effects on the two dimension of well-being that we measure: mental health and self-esteem. This is consistent with the program's modest time demands and its emphasis on well-being in the parenting workshops.

Our paper contributes 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 existing evidence suggests some promising approaches, it is mostly based on NGO-provided programs (e.g., see Dean and Jayachandran (2019); Martinez et al. (2017)). Two recent studies find that teaching assistant (TA) models can effectively improve the quality of government early childhood education provision, as long as teachers receive adequate training on deploying TAs appropriately in the classroom (Ganimian et al., 2024; Andrew et al., 2024). In line with these studies, we find that bringing additional support into the classroom can be a cost-effective way of improving children's outcomes in government preschools. However, we show that this support can be provided by low-educated, mostly illiterate parents rather than additional salaried staff, without adverse effects on their observed well-being and with the potential for additional benefits through improvements in the home learning environment. The effect of the *Lively Minds* program on an aggregate cognitive index is comparable to the ITT effect of introducing TAs in pre-schools in India (0.11 SD gain on a composite learning measure in Ganimian et al., 2024).

By focusing on parents, our study also contributes to the parenting-intervention literature. A large literature shows that parenting programs can effectively improve parenting and child development in the first 1,000 days (Jeong et al., 2021; Carneiro et al., 2024). Much less is known about parenting interventions for preschool-age children, despite the fact that parents continue

[2] The calculation assumes a 13% increase in earnings for a one-SD increase in cognitive skills (Chetty et al., 2011; Ganimian et al., 2024), applied to the estimated treatment effect on cognitive skills of 0.12 SD. This return is applied to average annual rural per-capita income in Ghana from the 2016-17 Ghana Living Standards Survey (Ghana Statistical Service, 2019). Earnings are assumed to accrue between ages 22 and 60 with zero real earnings growth and are discounted to the trial year (2017) at 3%. Section 7 provides further details

to shape children's learning and development well beyond the first 1,000 days. This question is particularly important in poor, low-education settings, where parents may find it more difficult to support children's development during the preschool years than to provide the nurturing care required in early infancy. Existing evidence in such contexts is mixed: Angrist et al. (2025) finds no impacts on parental knowledge or behavior from one-to-one NGO-led training, while Özler et al. (2018) reports more encouraging results from a group-based parenting program. Our study adds new evidence that it is possible to improve ECD knowledge and practices among parents of preschool-age children in poor, rural, low-education settings. Moreover, we show that this can be achieved through a relatively light-touch intervention, embedded within existing government education infrastructure, rather than NGO led, intensive parent training that characterizes many successful parenting programs.

Our findings also have implications for the design of *scalable* government programs. A growing literature documents that interventions often become less effective as they expand beyond the conditions of the original trial —a phenomenon referred to as "voltage drop". Contributing factors include replacing highly motivated early implementers with a more heterogeneous workforce, differences between marginal intervention recipients and participants in the original trial, shortages of skilled personnel, weak incentives among frontline staff, and growing administrative and coordination challenges which result in higher marginal delivery costs at scale (Banerjee et al., 2017; Al-Ubaydli et al., 2017; List et al., 2021). The *Lively Minds* model is explicitly designed to mitigate several of these constraints. Rather than relying on a small and difficult-to-expand pool of trained professionals, it recruits motivated mothers from local communities, while embedding program delivery within the existing public preschool system with only limited additional demands on existing personnel (List, 2026).

While our experiment was not designed to evaluate scalability directly, subsequent government implementation provides suggestive evidence that these design features may indeed help sustain effectiveness at scale. Following the positive findings from this trial, the Government of Ghana scaled the program to reach more than 500,000 children - over 40% of all children enrolled in public preschools - and is continuing to scale it across rural Ghana. Administrative records indicate that the program has largely preserved its core implementation features during scale-up: sufficient numbers of mothers continue to volunteer and remain engaged over time and

play schemes and parenting workshops continue to operate as intended. At the same time, implementation costs have fallen by around 60%. Together, these patterns are consistent with the view that the program's design mitigates several of the implementation challenges commonly associated with large-scale expansion. If the impacts estimated in this study are sustained at scale, the implied benefit-cost ratio would increase to around 21. This estimate is itself conservative, as a substantial share of the remaining implementation costs reflects one-off set-up expenditures rather than recurrent delivery costs. Taken together, these findings suggest that the already high cost-effectiveness observed in the trial could be maintained, or even enhanced, under large-scale government implementation.

The rest of the paper is organized as follows. Section 2 presents the Ghanaian context and the *Lively Minds* program in detail. Section 3 describes the experimental design and intervention implementation, while Section 4 describes the data. Section 5 presents our empirical framework and identification strategy. Section 6 presents our findings, and Section 7 discusses the program's cost effectiveness. Section 8 concludes.

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, 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 learning materials (The World Bank, 2024). At the same time, many children grow up in home environments that provide little cognitive stimulation, due to 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, 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 (Ministry of Education, Ghana, 2021). 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 preschool interventions 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, supervision, 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 the 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 and support children's learning and development at home. 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.

Program description. The *Lively Minds* program, developed by the Lively Minds NGO, seeks to improve the early childhood development (ECD) outcomes of preschool children. To achieve this it mobilizes and trains women in the community to run weekly, play-based learning activities in classrooms alongside teachers and to improve their care and education practices at home. We refer to these women as participating mothers (PMs, henceforth) since they tend to be mothers of preschool-age children. During each play scheme session, PMs facilitate discovery-based activities at simple indoor "play stations", working with small groups of no more than five children. A complete session comprises five stations operating simultaneously - matching/sorting, numeracy, sizes/colors/senses, books, and building - with children rotating

through all five, spending 10 minutes at each station. Preschool teachers supervise the sessions, while additional children participate in structured outdoor play also led by a PM (Figure B.1 and Figure B.2). 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, the PMs are offered monthly parenting and well-being workshops delivered by the preschool teachers which also reinforce the initial training and provide ongoing support for their work in the play scheme.[3]

Implementation model. The program is embedded within the 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 PMs. 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 have to complete a teacher-led training course consisting of eight sessions, two hours each, over the course of three weeks. 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, the PMs are divided into four groups; each group is given a different day on which they run the play schemes at the preschool for one hour (with extra time for preparation and debrief).

GES officials monitor the play schemes and parenting workshops 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 program 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 preschools 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

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

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 low: each PM volunteers for 1-2 hours per week, with different groups of mothers taking turns to run the play schemes on different days. Participation is encouraged through free monthly parenting and well-being workshops rather than financial payments. Finally, the training and play curricula as well as the implementation model were developed with and in rural communities of Northern Ghana and refined over nearly a decade of prior implementation, ensuring that they are context-appropriate, accessible to women with limited formal education and feasible in predominantly rural settings where women often have flexible farm-based work schedules.

3 Study Design and Implementation

3.1 Experimental Design

In order to evaluate the *Lively Minds* program, we conducted a cluster-RCT between 2017 and 2019. The randomization was conducted at the preschool level: 80 preschools 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. However, implementation in the RCT districts was restricted to the sub-set of preschools randomly selected for treatment. The remaining preschools received the program the following year.

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 that education levels among women in northern Ghana were significantly below the national average; for example while nationally 65% of women in the 15-49 age-group were literate, in the Northern region it was only 29% (Ghana Statistical Service and UNICEF, 2018). The two study regions differ along several dimensions, including language,

religion and culture, providing an opportunity to test the program in a heterogeneous set of communities that reflect some of the diversity of rural Ghana.

The randomization was conducted using two levels of stratification - circuit and school size.[4] The final sample includes 38 preschools in the Bongo District, and 42 preschools 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 included children intending to start preschool in the academic year 2017/18 (age 4-5).

3.2 Participating Mothers

A distinguishing feature of the *Lively Minds* program is its reliance on parents rather than paraprofessionals, such as teaching assistants, as the primary source of additional classroom support. To better understand the role of parental participation in the program, we would like to examine impacts separately for PMs and non-PMs and their children.

A key challenge is that participation in the program is voluntary. To estimate impacts on PMs and their children, we therefore require a comparison group of mothers who would have participated had the program been available to them. To construct such a comparison group, we conducted the same mother-mobilization activities in communities served by the treatment and control preschools, allowing us to identify mothers in control communities who would have participated had the program been available to them. In communities served by treatment preschools, mothers were invited to sign up to participate in the school year that was about to start. In communities served by control preschools, mothers were invited to sign up to participate in the following year, when the program was scheduled to be rolled out.

We were able to match lists of mothers identified as intending to participate in the program in the treatment and control preschools to the mothers of children in our sample. In total, 437 sample children had mothers who expressed an intention to participate: 263 in the treatment group and 174 in the control group. Our measure of intention to participate is based on signing up during the mobilization event in control communities and attending program training in treatment communities. As noted in the Introduction, throughout the paper we refer to these women as *Participating* Mothers, although they are, more precisely, mothers who signaled an intention to participate. At endline, 34 of 263 PMs in the treatment group reported not having

[4] A circuit is a geographical cluster of around 10 preschools that falls under one supervisor from the GES.

participated in the program, suggesting some attrition between training and program implementation. We, nevertheless, retain these mothers in the analysis because the corresponding mothers in control communities are observed only at the intention-to-participate stage. Section 5 formally sets out the conditions under which impacts on PMs and non-PMs can be identified separately and provides evidence in support of these assumptions.

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 preschool teachers, headteachers and PTA leads in treatment preschools. The teachers conducted recruitment and training of PMs in October 2017. Play schemes started to run in treatment preschools 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 the 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 two or three times per week. Over the course of the academic year, the play schemes ran over three terms: 6 weeks in November-December in the first term, 13 weeks in January-April in the second term, and 11 weeks in May-July in the third term. 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. 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.

Table 1: Teacher reports related to program compliance

	Parents Contribute (1)	Parents support at least once a month (2)	Any games for free play (3)	Any play in small groups (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 standardized 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.

Each school also received an average of between 4 and 5 monitoring visits at the monthly parent workshops organized by the preschool teachers. In 88% of the visits, the workshop was found to be taking place. In most cases, two preschool teachers were present with an average attendance of 87% of PMs per session (30 out of 35).

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 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 Uni-

versity College London IRB (ref: 10167/001). A structured ethical appendix, following Asiedu et al. (2021), can be found in Appendix C. It covers policy equipoise, role of the researcher, potential harms to participants and nonparticipants, conflicts of interest, intellectual freedom, feedback to participants, and foreseeable misuse of research results.

4 Data, Sample and Outcomes

4.1 Data collection and sampling procedures

To ensure that baseline data collection was completed before the intervention began, sampling and baseline surveys were conducted in July 2017, before the start of the school year. This meant that it was not possible to use school enrollment lists to identify the population of interest for the sampling. We, therefore, conducted a census of households located in the vicinity of the sampled schools to identify children who were due to start preschool. The households closest were enumerated first, gradually moving to those further away to reach 150 households.[5] 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.)

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

4.2 Balance and sample characteristics

Table 2 and Table 3 present key baseline characteristics of primary caregiver and target children, respectively, for the sample of 2,151 non-attritors, split by treatment status.

Table 2: Baseline balance: households and primary caregiver 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*	2,151
	(6.369)	(6.756)	(5.967)			
Number of children 16 or under	4.417	4.573	4.264	*0.045*	*0.228*	2,151
	(3.568)	(3.800)	(3.321)			
Number of children 6 or under	1.617	1.676	1.560	*0.140*	*0.561*	2,151
	(1.825)	(1.893)	(1.755)			
Wealth index (z-score)	0.025	-0.025	0.074	*0.283*	*0.729*	2,139
	(2.114)	(2.067)	(2.159)			
Main income source:						
Farming own land	0.446	0.447	0.445	*0.933*	*0.990*	2,151
Wage work	0.274	0.274	0.274	*0.998*	*0.994*	2,151
Livestock	0.033	0.034	0.031	*0.721*	*0.970*	2,151
Profits from small enterprise	0.183	0.189	0.177	*0.458*	*0.882*	2,151
B. Primary carers						
Age	36.435	36.948	35.935	*0.050*	*0.315*	2,151
	(11.996)	(12.156)	(11.823)			
Target child's biological mother	0.790	0.774	0.806	*0.063*	*0.365*	2,151
Some education	0.196	0.192	0.200	*0.652*	*0.896*	2,151
Illiterate	0.926	0.930	0.922	*0.466*	*0.882*	2,151
Raven's score (%)	43.178	43.473	42.890	*0.413*	*0.882*	2,151
	(16.524)	(17.090)	(15.956)			
Depression score (0-20)	0.028	0.000	0.056	*0.193*	*0.677*	2,151
	(1.002)	(1.000)	(1.003)			
Self-esteem score (0-30)	0.009	-0.000	0.017	*0.684*	*0.896*	2,151
	(0.980)	(1.000)	(0.961)			
Knowledge of child development (z-score)	-0.021	0.000	-0.042	*0.315*	*0.826*	2,151
	(0.965)	(1.000)	(0.931)			
Play materials (z-score)	0.088	-0.000	0.173	*0.000*	*0.004*	2,151
	(1.074)	(1.000)	(1.136)			
Any play activities with child (0/1)	0.132	0.120	0.143	*0.108*	*0.511*	2,151

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.

The sample is well balanced across the treatment and control groups on most key baseline household, caregiver and child characteristics. A notable exception is an imbalance in availabil-

Table 3: 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*	2,151
	(9.499)	(9.481)	(9.521)			
Gender	0.492	0.496	0.489	*0.754*	*0.996*	2,151
Currently in school	0.781	0.783	0.779	*0.809*	*0.996*	2,151
In the last 30 days:						
Any cough	0.549	0.525	0.572	*0.025*	*0.144*	2,150
Any stomach pain	0.508	0.481	0.535	*0.011*	*0.068*	2,150
Any high fever	0.340	0.319	0.361	*0.036*	*0.154*	2,151
Any diarrhea	0.320	0.321	0.320	*0.975*	*0.996*	2,144
B. Cognitive outcomes (z-scores)						
Aggregate score	-0.002	-0.000	-0.004	*0.929*	*0.998*	2,151
	(1.052)	(1.000)	(1.100)			
Emergent literacy	-0.005	0.000	-0.011	*0.808*	*0.988*	2,151
	(1.031)	(1.000)	(1.061)			
Emergent numeracy	0.017	0.000	0.033	*0.462*	*0.910*	2,151
	(1.045)	(1.000)	(1.088)			
Executive function	-0.029	0.000	-0.057	*0.202*	*0.681*	2,151
	(1.034)	(1.000)	(1.065)			
C. Socioemotional outcomes (z-scores)						
Emotional awareness	0.018	-0.000	0.036	*0.415*	*0.910*	2,151
	(1.038)	(1.000)	(1.073)			
Strengths and Difficulties:						
Emotional and behavioural problems	0.003	0.000	0.005	*0.907*	*0.998*	2,151
	(0.987)	(1.000)	(0.975)			
Externalising behaviours	0.012	-0.000	0.024	*0.568*	*0.936*	2,151
	(0.983)	(1.000)	(0.967)			
Internalising behaviours	-0.006	-0.000	-0.013	*0.767*	*0.988*	2,151
	(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.

ity of play materials, which we control for in the analysis. At baseline, children in the sample were, on average, around 56 months (4.7 years) old, evenly split by gender, with 78% enrolled to go to preschool in the upcoming academic year. They lived in large households, with an average household size of more than 10 members, of whom just over 7 were aged 16 or younger. Farming on own land was the most common main source of household income (45%), followed by wage

employment (24%). Those identified as target children's primary caregivers and interviewed at baseline and endline were on average 36.4 years of age ; only 7% reported being literate and 20% having some formal education (for many this is incomplete primary). The great majority of primary caregivers (79%) are the target child's mother; their average age was 32.8. Levels of engagement in developmentally appropriate activities with children were strikingly low in these families: only 13% of primary caregivers report that anyone in the household played with the target child in the last three days. Together, these patterns of low caregiver education and very limited engagement with children underscore the challenging context in which the *Lively Minds* program was implemented.

4.3 Outcomes

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

4.3.1 Child development

At baseline and endline, we administered the International Development and Early Learning Assessment (IDELA) to measure cognitive and socioemotional development (Pisani et al., 2015), alongside selected items from the Strengths and Difficulties Questionnaire (SDQ) to capture emotional and behavioral difficulties (Goodman, 1997). The IDELA was administered directly to children by trained interviewers in one-to-one sessions and covers emergent literacy, emergent numeracy, executive functioning, and emotional awareness; the behavioral difficulties instrument was administered to the child's primary caregiver (Hoosen et al., 2018a). At endline, we additionally administered child tasks developed at Harvard's Laboratory for Developmental Studies and piloted in Ghana (Coffey and Spelke, 2024). These tasks were designed to provide more precise measurement of child development in domains captured by the IDELA and, like the IDELA, were administered directly to children by trained interviewers. Scores used in the main analysis follow developer-prescribed scoring rules and are standardized to have mean 0 and SD 1 in the control group distribution. Appendix A provides further details on the assessments, score construction, and robustness to alternative scoring approaches.

We also collected child development data for one older and one younger sibling of each

target child, where applicable, to study intra-household spillovers. Younger siblings' development was captured using the Caregiver-Reported Early Development Index (CREDI) short form — a caregiver-report instrument for population-level measurement of early development from birth to age three in low-resource settings. Scores were constructed following developers' age-standardization guidance and then standardized relative to the control group (McCoy et al., 2017). Older siblings completed direct, one-to-one assessments comprising 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's Laboratory of Developmental Studies (see Amadu et al. (2020) for further details).

4.3.2 ECD knowledge, practices and the home environment

The *Lively Minds* program aims to improve child development by strengthening parents' ECD knowledge and skills and engaging them in supporting children's development in preschool and at home. Our secondary outcomes, therefore, include a rich set of measures of ECD knowledge and practices among parents in our sample, collected through primary caregiver surveys. ECD knowledge is assessed using a short subset of the Knowledge of Infant Development Inventory (KIDI), a widely used instrument asking about caregivers' beliefs about children's developmental capabilities and appropriate expectations (MacPhee, 1981a). We measure caregivers' understanding of preschool quality using a contextually tailored instrument. Caregivers were shown pairs of illustrated vignettes depicting contrasting classroom environments and teaching practices and asked to identify the option that provided the better learning environment for young children.

We capture ECD practices by combining caregiver reports with direct observation. Our first measure was collected by observing mothers' interactions with their own children during structured play tasks, coding a range of supportive and less supportive instructional behaviors, drawing on items from UNICEF's Multiple Indicator Cluster Surveys (MICS) (Bornstein et al., 2010). We also measure home investments in child development using the Family Care Indicators Questionnaire (FCI), which records both the availability of play and learning materials in the home and the play and learning activities caregivers and other adults in the household engaged in with the child during the previous three days (Hamadani et al., 2010). Total scores are

standardized to have mean 0 and SD 1 in the control group.

5 Empirical strategy

5.1 Average Treatment Effect of the *Lively Minds* program

Our primary objective is to estimate the effect of the *Lively Minds* program on the outcomes of children and their primary caregivers. Because some mothers self-select into program participation, we set out a potential outcome framework to define the treatment effects of interest and how they are identified.

Assuming a linear and additively separable specification, let $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes of child i when their preschool is not assigned, and is assigned, to the *Lively Minds* program, respectively:

$$Y_i(0) = \mu_0 + \alpha_0 P_i + U_{0i}$$

$$Y_i(1) = \mu_1 + \alpha_1 P_i + U_{1i}$$

where μ_0 and μ_1 are intercepts, P_i is an indicator equal to 1 if child i's mother participates in the program and 0 otherwise, and U_{0i} and U_{1i} are unobservables. We assume, without loss of generality, that $\mathbb{E}(U_{1i}) = \mathbb{E}(U_{0i}) = 0$.

We allow the effect of P_i on potential outcomes to differ by treatment status: α_0 captures selection into participation (e.g. engaged mothers being more likely to volunteer), while α_1 captures this same selection effect *plus* any indirect impact of *Lively Minds* on the child via the mother's training and role in delivering the play schemes – for instance, if mothers acquire skills that change how they interact with their children at home, or if children learn more when the play activities are led by their own mothers.

Our primary interest is in the Average Treatment Effect (ATE) of the program, defined as:

$$ATE = \mathbb{E}[Y_i(1) - Y_i(0)] = \underbrace{(\mu_1 - \mu_0)}_{\text{Direct channel}} + \underbrace{(\alpha_1 - \alpha_0)}_{\text{Indirect channel}} \mathbb{E}[P_i] \tag{5.1}$$

The ATE thus reflects two channels: a direct channel, $(\mu_1 - \mu_0)$, from the exposure of all children to the play schemes, and an indirect channel, $(\alpha_1 - \alpha_0)$, from maternal participation, weighted

by the share of children with a PM, $\mathbb{E}[P_i]$.

To identify the ATE, we leverage the experimental design and compare the average outcomes of children in the treatment and control groups. We estimate this difference via the following OLS regression:

$$Y_{i,s,t} = \alpha + \beta T_{s(i)} + \sigma Y_{i,t-1} + \gamma X_{i,t-1} + \rho_{g(s(i))} + \epsilon_i \tag{5.2}$$

where $Y_{i,t}$ is the outcome of interest for child i measured at endline (t); $T_{s(i)}$ is a dummy equal to 1 if their preschool s received the *Lively Minds* program; $Y_{i,t-1}$ is the same outcome measured at baseline $(t-1)$[6]; as pre-specified, X_i 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; $\rho_{g(s(i))}$ is a fixed effect for the randomization strata g of child i's preschool s; finally ϵ_i is an individual error term.

The coefficient β identifies the ATE as long as unobservables are balanced across treatment groups. The randomization justifies this assumption, while the balance tests in Table 2 and Table 3 confirm that observed characteristics are well balanced across treatment groups. In taking this model to the data, two points are worth noting. First, because the sample comprises children who intended to attend pre-school at baseline, before the start of the school year, rather than those who actually attended, the estimated ATE should be interpreted as an intention-to-treat (ITT) effect. Second, given the number of outcomes examined across cognitive, socioemotional, and parental-input domains, we adjust for multiple inference using Romano-Wolf stepdown p-values (Romano and Wolf, 2005) within pre-specified outcome families, reporting both unadjusted and (where relevant) adjusted p-values. Standard errors are clustered at the preschool level, the unit of randomization.

5.2 Treatment effects by mothers' participation status

A novel feature of the *Lively Minds* program is its reliance on a sub-group of mothers as volunteers. This motivates examining impacts separately for PMs and non-PMs and their children. Using the notation above, we can define the average effect of preschool assignment to

[6]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.

Lively Minds program among children whose mothers self-select into participating (ATPM):

$$ATPM = \mathbb{E}[Y_i(1) - Y_i(0)|P_i = 1] = \underbrace{(\mu_1 - \mu_0)}_{\text{Direct channel}} + \underbrace{(\alpha_1 - \alpha_0)}_{\text{Indirect channel}} + \underbrace{E(U_{i1} - U_{i0}|P_i = 1)}_{\text{Sorting on gains}} \quad (5.3)$$

The ATPM captures both the direct and indirect channels through which the program may work, as well as a third term capturing potential unobservable gains from the program among the children whose mothers participate.[7]

To identify the ATPM, we leverage the prospective PMs in the control preschools, namely the women who signed up to deliver the play schemes in the following year. We estimate Equation 5.2 separately for the samples of PMs and non-PMs, where the estimand for the coefficient on T_i in the PM sample, β^{PM}, can be written as:

$$\beta^{PM} = \mu_1 - \mu_0 + \alpha_1 - \alpha_0 + \mathbb{E}[U_{i1}|T_i = 1, P_i = 1] - \mathbb{E}[U_{i0}|T_i = 0, P_i = 1]$$

This expression shows that β^{PM} simplifies to ATPM if and only if $\mathbb{E}[U_{i1}|T_i = 1, P_i = 1] - \mathbb{E}[U_{i0}|T_i = 0, P_i = 1] = \mathbb{E}[U_{i1} - U_{i0}|P_i = 1]$; i.e., if on average the unobservable characteristics of PMs are independent of treatment arm. We view this assumption as plausible for three reasons. First, the mobilization and sign-up activities were conducted identically across arms, differing only in the timing of program roll-out. Second, because children spend two years in preschool and our sample consists of children entering their first year of preschool, mothers in both treatment and control communities were presented with a similar opportunity to participate while their own child was enrolled. Third, and most importantly, we verify that PMs in treatment and control communities are well balanced on a wide range of baseline household, caregiver, and child characteristics: Table B.3 and Table B.4 show no systematic differences between the two groups. Taken together, this evidence supports the assumption that mothers who intended to participate in control communities constitute a credible comparison group for mothers who intended to participate in treatment communities, enabling the estimation of treatment effects among PMs. Importantly, because not all mothers who intended to participate ultimately did so, β^{PM} should be interpreted as an intention-to-treat (ITT) effect among

[7] These gains could be positive if the children of PMs are especially responsive to intervention activities due to, for example, deficits in their development not captured by our baseline measures — or negative if, for example, PMs already invest more in their children, leaving less room for the program to improve outcomes.

mothers who intended to participate.

Applying the same logic to the identification and estimation of the ATE among Non-participating Mothers (ATNM), we can then examine how treatment effects differ between PMs and non-PMs. We return to how to interpret this difference in subsection 6.4.

6 Results

6.1 Knowledge and practices of parents

We begin by estimating the program's impacts on mothers' ECD knowledge, investments in child development at home, and the quality of their interactions with children (see Appendix A for measurement details). Examining these outcomes allows us to assess a central premise of the program: that low-educated mothers can be equipped to support the development of pre-school age children. As discussed in Section 4, our main estimates use mean-of-raw score indices standardized to have mean 0 and SD 1 in the control group distribution; we show that our results are robust to alternative scoring approaches using factor analysis in Appendix B. We present impacts separately for PMs and non-PMs. Impacts on the latter may capture spillover impacts of the program on mothers in the community. Differences in impacts and p-values for their statistical significance are reported at the bottom of the table.

ECD Knowledge. Table 4 shows clear improvements in ECD knowledge among PMs: knowledge about child development increases by just under half a SD (column 1, $p = 0.012$).[8]

As discussed in Section 2, the *Lively Minds* approach differs markedly from business-as-usual methods in Ghanaian public preschools, emphasizing small-group play, simple educational games with props, and classroom organization that supports children's learning. It is thus plausible that even non-PMs update their beliefs about child development and preschool quality. Table 4 provides some evidence of this. While we observe no effect on non-PMs' knowledge of child development (column 1, $p = 0.854$), knowledge of preschool quality increases by 0.14 SD (column 2, $p = 0.096$). This estimate is similar in magnitude to that for PMs, although the latter is estimated less precisely, likely reflecting the smaller PM sample. This pattern is consistent with differences in exposure to program content by PM status; non-PMs appear

[8]Note that these estimates use mean-of-raw score indices; corresponding specifications using factor scores show very similar effects (Table A.9).

to update beliefs about more easily observable changes at preschools, while PMs also acquire better overall knowledge about child development [9].

ECD practices. Better ECD knowledge 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). While training, parenting workshops, and experience running the play schemes may have influenced PMs' ECD practices, their low education levels and the light-touch nature of the intervention make the impact less certain.

We examine *Lively Minds*' impacts on these behaviors in Table 4. Results in columns 3 and 4 of Panel A show evidence of improvements in ECD practices among PMs. There is a large and significant increase in play and engagement in developmentally supportive activities in PM households. The share of PMs reporting that someone in the household had engaged in play activities with the sample child during the previous three days increases by 18 percentage points in the treatment group (column 4, $p = 0.012$), equivalent to almost 50% of the PM control group mean. 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, $p = 0.974$).

This large increase in play activities without a corresponding increase in material investment is consistent with PMs adopting the low-cost, routine-based ways of engaging children emphasized in training and parenting workshops, embedding learning opportunities into everyday household chores. Disaggregating the results by specific play activities in Table B.7, 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 of Table 4 further provide suggestive evidence of improved quality of PMs' engagement with children, measured through enumerators' observations of the quality of PMs' interactions with their own children (see Section 4 for details on this measure). We see a 0.20 SD increase in positive parenting behaviors (supportive, responsive, developmentally appropriate), though it is not statistically significant (column 5, Panel A, $p = 0.145$), particularly after adjusting for multiple hypothesis testing ($p = 0.453$).

[9] Estimates of impacts on individual items of the pre-school quality measure reveal that the overall effect is driven primarily by an increase in the proportion of mothers who thought that play-based learning in small groups was better for kids than business-as-usual rote learning methods (Table B.5).

Table 4: 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	0.460	0.123	-0.034	0.182	0.198	-0.025
	(0.132)	(0.107)	(0.135)	(0.049)	(0.134)	(0.141)
p-val	*0.001*	*0.252*	*0.800*	*0.000*	*0.145*	*0.859*
Adj. p-val	*0.012*	*0.537*	*0.974*	*0.012*	*0.453*	*0.974*
Treatment among non-PMs	-0.046	0.138	-0.201	-0.032	0.048	-0.011
	(0.105)	(0.060)	(0.104)	(0.036)	(0.076)	(0.095)
p-val	*0.663*	*0.023*	*0.056*	*0.385*	*0.532*	*0.905*
Adj. p-val	*0.854*	*0.096*	*0.172*	*0.760*	*0.820*	*0.872*
Difference in treatment	0.506	-0.015	0.167	0.214	0.150	-0.014
	(0.120)	(0.114)	(0.130)	(0.042)	(0.104)	(0.138)
p-val	*0.000*	*0.893*	*0.202*	*0.000*	*0.152*	*0.921*
Adj. p-val	*0.002*	*0.984*	*0.367*	*0.002*	*0.363*	*0.984*
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 standardized 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.

The evidence of changes in PMs' parenting behavior, together with impacts on non-PMs' beliefs about preschool quality, suggest that the *Lively Minds* program may have indirectly influenced the parenting practices of non-PMs. Columns 3–6 of Table 4 show some suggestive evidence of this, though in a pattern very different from PMs. There is no impact on time spent with children among non-PMs. This is a marked contrast to the large positive effects among PMs, with the difference in treatment effects on time investment statistically significant (column 4, 'Difference in treatment' row). There is, however, suggestive evidence of a *reduction* in investment in play materials at home (column 3), which we do not observe for PMs: investment among non-PMs fell by close to a fifth of a SD, though not significantly after adjusting for multiple hypothesis testing ($p = 0.172$). This reduction is driven by negative impacts on several items, including household objects used as toys, books, and toys to help with moving (Table B.6). Taken together, these patterns suggest that non-PMs were aware of the improvements in preschool quality and may have responded by reducing investments at home, perhaps viewing preschool inputs as a substitute for home-based investments. A related substitution mechanism has been suggested by Walters (2015) as a possible explanation for the negative cognitive impacts estimated for some individual Head Start centers.

6.2 Child development

Cognitive development. Consistent with the program's success in strengthening PM's ECD knowledge and practices, and with high compliance in the delivery of classroom play schemes (Section 3), we find that *Lively Minds* generated economically meaningful gains in children's cognitive development. Table 5 shows an ITT effect of 0.12 SD (column 1, $p = 0.028$) on the aggregate cognitive development measure, combining emergent numeracy, emergent literacy, and executive functions. Disaggregating by subdomain, the gains are concentrated in emergent numeracy (column 3, $p = 0.024$) and executive functions (column 4, $p = 0.024$). [10]

The "Cognition" panel of Table 6 shows that the program was not equally effective across the skill distribution: it significantly increased the probability of scoring above the 25^{th} and 75^{th}

[10] Our pre-analysis plan (AEA RCT Registry No. AEARCTR-0002777) was finalized before we decided to include the Harvard Laboratory for Developmental Studies measures in the endline survey. The pre-specified primary child development outcome therefore comprised only the IDELA-based measures. In line with the pre-analysis plan, Table B.11 reports treatment effects on this pre-specified cognition index (including emotional awareness but excluding the Harvard tasks). The estimated impacts are very similar in magnitude to our benchmark results, although somewhat less precisely estimated, and remain statistically significant.

Table 5: Treatment effects on cognitive outcomes

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
A. Full sample				
Treatment	0.120	0.063	0.123	0.139
	(0.054)	(0.057)	(0.051)	(0.052)
p-val	*0.028*	*0.272*	*0.019*	*0.010*
Adj. *p*-val	-	*0.176*	*0.024*	*0.024*
Baseline outcome	0.451	0.370	0.382	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	0.116	0.068	0.106	0.150
	(0.060)	(0.070)	(0.060)	(0.062)
p-val	*0.058*	*0.334*	*0.083*	*0.017*
Adj. *p*-val	-	*0.250*	*0.090*	*0.046*
Treatment among girls	0.136	0.072	0.144	0.145
	(0.060)	(0.060)	(0.057)	(0.065)
p-val	*0.027*	*0.232*	*0.014*	*0.030*
Adj. *p*-val	-	*0.196*	*0.022*	*0.034*
Difference	0.020	0.005	0.038	-0.005
	(0.055)	(0.063)	(0.058)	(0.074)
p-val	*0.720*	*0.942*	*0.516*	*0.947*
Adj. *p*-val	-	*0.990*	*0.852*	*0.990*
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 randomized 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 standardized to have mean 0 and standard deviation 1 in the control group.

Table 6: 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	0.011	0.053	0.035	0.067	0.002
	(0.013)	(0.016)	(0.025)	(0.027)	(0.018)
p-val	*0.385*	*0.001*	*0.169*	*0.014*	*0.928*
Emergent Literacy	0.005	0.004	0.042	0.023	0.008
	(0.013)	(0.020)	(0.021)	(0.026)	(0.015)
p-val	*0.726*	*0.845*	*0.051*	*0.379*	*0.613*
Emergent Numeracy	0.010	0.039	0.045	0.049	0.018
	(0.009)	(0.016)	(0.024)	(0.025)	(0.016)
p-val	*0.270*	*0.018*	*0.061*	*0.051*	*0.276*
Executive Function	0.026	0.055	0.022	0.035	0.027
	(0.013)	(0.018)	(0.023)	(0.028)	(0.018)
p-val	*0.050*	*0.002*	*0.341*	*0.220*	*0.138*
Emotional Awareness	-0.006	0.019	0.034	0.018	0.009
	(0.013)	(0.020)	(0.026)	(0.032)	(0.021)
p-val	*0.622*	*0.340*	*0.196*	*0.575*	*0.659*
Emotional and Behavioural Pbms	0.006	-0.021	0.002	-0.070	-0.055
	(0.016)	(0.029)	(0.027)	(0.023)	(0.016)
p-val	*0.715*	*0.464*	*0.938*	*0.004*	*0.001*
Externalising Behaviours	0.006	-0.011	-0.058	-0.046	-0.033
	(0.009)	(0.025)	(0.031)	(0.025)	(0.015)
p-val	*0.534*	*0.671*	*0.059*	*0.068*	*0.030*
Internalising Behaviours	0.009	0.019	0.019	-0.032	-0.053
	(0.010)	(0.024)	(0.031)	(0.027)	(0.022)
p-val	*0.396*	*0.438*	*0.544*	*0.243*	*0.018*
Strata fixed effects	Yes	Yes	Yes	Yes	Yes
Baseline outcome	Yes	Yes	Yes	Yes	Yes
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.

percentiles of the control distribution, with larger effects at higher skill levels, but had no impact at the tails (10^{th} or 90^{th} percentile). There is some indication that the distributional pattern of treatment effects is different from this for the executive functions domain: the intervention was most effective at increasing executive functions among those at the lower end of the executive functions distribution.

The program was equally effective at improving the cognitive skills of boys and girls (column 1, Panel B of Table 5). Estimated treatment effects are similar in magnitude for boys ('boy' row, $p = 0.058$) and girls ('girl' row, $p = 0.027$), and we cannot reject equality of effects across genders ('difference' row, $p = 0.720$). It was also equally effective among children of PMs and non-PMs (Table 7). This pattern suggests that gains in children's cognitive skills were driven primarily by changes in the pre-schools environment - likely the introduction of the play schemes – rather than by the additional improvements in the home learning environment experienced by the children of PMs.

The estimated effect of 0.12 SD on cognitive skills is comparable in magnitude to those typically achieved by successful education interventions. It is close to the median effect reported across randomized evaluations of education interventions in both high-income and low- and middle-income countries (Kraft, 2020; Evans and Yuan, 2022). It is also similar to the 0.16 SD gain reported by Bassi et al. (2025) for a large-scale preschool construction program in rural Mozambique, suggesting that, in low-capacity systems, low-cost improvements in the quality of existing preschools may generate learning gains comparable to those from large-scale expansions in access.

Socioemotional development. Table 8 reports effects of *Lively Minds* on children's socioemotional development. A positive coefficient in column 1 indicates improved emotional awareness; negative coefficients in columns 2–4 indicate reduced emotional and behavioral problems.

We find little evidence that the program affected emotional awareness (column 1, $p = 0.458$). This absence of effects is evident across the score distribution (Table 6) and holds for both boys and girls (Panel B, column 1, Table 8). There is, however, suggestive evidence that the program reduced behavioral difficulties: column 2 shows an effect of -0.099 SD ($p = 0.096$). Distributional analysis shows the program significantly improved emotional and behavioral problems among children with above-average problem load, at the 75^{th} and 90^{th} percentiles of the outcome

Table 7: Impacts on cognitive outcomes, by PM status

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.137	0.028	0.160	0.144
	(0.070)	(0.071)	(0.080)	(0.092)
p-val	*0.056*	*0.689*	*0.050*	*0.123*
Adj. *p*-val		*0.697*	*0.112*	*0.184*
Treatment, non-PM children	0.103	0.058	0.097	0.137
	(0.055)	(0.059)	(0.054)	(0.050)
p-val	*0.064*	*0.332*	*0.074*	*0.008*
Adj. *p*-val		*0.230*	*0.092*	*0.018*
Difference in treatment	0.034	-0.029	0.063	0.007
	(0.061)	(0.062)	(0.082)	(0.081)
p-val	*0.581*	*0.638*	*0.449*	*0.929*
Adj. *p*-val		*0.848*	*0.770*	*0.922*
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: The sample includes children attending pre-schools randomized 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 program. 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 program in the following year.

distribution (Table 6).

This reduction appears to be driven by lower externalizing behaviors, such as hyperactivity, conduct problems and inattention (column 3, 0.11 SD reduction, $p = 0.102$). Impacts are concentrated in the upper half of the distribution, indicating larger benefits for children with relatively high levels of externalizing behaviors (Table 6). This is consistent with the program's emphasis on structured, adult-supervised small-group play, which promotes rule-following and turn-taking, aligning with evidence that play-based activities strengthen self-regulation (Dia-

mond and Lee, 2011; Healey and Healey, 2019), and that weak self-regulation is closely linked to externalizing difficulties (Sher-Censor et al., 2016; Coelho et al., 2023).

These reductions are concentrated among boys, for whom *Lively Minds* reduces problem behaviors by just over a fifth of a SD (see Panel B of Table 8, column 2, $p = 0.003$). This is consistent with evidence that boys exhibit higher externalizing behavior than girls in early childhood (Keenan and Shaw, 1997) (in our control group, the problem behavior score is 0.11 SD lower for girls than for boys on average) and that the effect of *Lively Minds* is concentrated at the higher end of the problem score distribution.

In contrast to the impacts on cognitive development, Table 9 shows that reductions in behavioral difficulties are concentrated among the children of PMs. These children experience large and statistically significant gains, in sharp contrast to the small and insignificant impacts on children of non-PMs. Among PMs' children, the program reduces behavioral difficulties by 0.265 SD (column 2, $p = 0.022$), an effect that exceeds the 0.1–0.15 SD range reported in a meta-analysis of preschool interventions (Holla et al., 2021). The overall reduction reflects improvements in both externalizing and internalizing behaviors, although the effects are larger for externalizing behaviors.

This finding suggests that the program's benefits for children's socioemotional development are likely linked to improvements in the home environment and caregiving practices among PMs, stemming from gains in their child development knowledge and behaviors. The absence of corresponding effects among the children of non-PMs suggests that changes within preschools, while beneficial for cognitive outcomes, were not enough to improve children's socioemotional development in the absence of complementary improvements in the home environment.

6.3 Spillovers on siblings' development

Evidence of spillovers to other children in the household could shed further light on the mechanisms underlying the impacts of *Lively Minds*. In line with evidence that parent-focused interventions often generate broader familial impacts (e.g., Carneiro et al., 2023), spillovers could be positive if target children transmit what they learn at preschool to siblings, and/or PMs apply new knowledge and practices to all children in the household. Conversely, spillovers could be negative if mothers reallocate attention toward the target child at siblings' expense.

Table 8: Impacts of intervention on children's socioemotional outcomes

	Emotional Awareness	**Emotional Behavioral Problems**		
		Overall	**Externalizing Behaviors**	**Internalizing Behaviors**
	(1)	(2)	(3)	(4)
Panel A - Full sample				
Treatment	0.054	-0.099	-0.112	-0.061
	(0.072)	(0.059)	(0.061)	(0.060)
p-val	*0.458*	*0.096*	*0.068*	*0.310*
Adj. *p*-val			*0.102*	*0.234*
Baseline outcome	0.228	0.066	0.093	0.038
	(0.024)	(0.023)	(0.022)	(0.028)
p-val	*0.000*	*0.006*	*0.000*	*0.172*
Panel B - Heterogeneity by gender				
Treatment among boys	0.083	-0.214	-0.190	-0.180
	(0.080)	(0.070)	(0.071)	(0.073)
p-val	*0.302*	*0.003*	*0.010*	*0.015*
Adj. *p*-val			*0.024*	*0.024*
Treatment among girls	0.034	0.015	-0.038	0.061
	(0.082)	(0.077)	(0.077)	(0.076)
p-val	*0.677*	*0.842*	*0.621*	*0.428*
Adj. *p*-val			*0.577*	*0.577*
Difference	-0.049	0.229	0.151	0.241
	(0.071)	(0.089)	(0.087)	(0.089)
p-val	*0.497*	*0.012*	*0.084*	*0.008*
Adj. *p*-val			*0.062*	*0.008*
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: The sample includes children attending pre-schools randomized 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 externalizing (conduct and hyperactivity) and internalizing (emotional and peer problems) behaviors, 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 standardized 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 9: Impacts on children's socioemotional outcomes, by PM status

	Emotional Awareness	**Emotional Behavioral Problems**		
		Overall	**Externalizing Behaviors**	**Internalizing Behaviors**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.020	-0.265	-0.280	-0.181
	(0.087)	(0.113)	(0.111)	(0.113)
p-val	*0.819*	*0.022*	*0.014*	*0.113*
Adj. *p*-val			*0.036*	*0.102*
Treatment, non-PM children	0.045	-0.065	-0.083	-0.031
	(0.074)	(0.056)	(0.055)	(0.063)
p-val	*0.543*	*0.245*	*0.137*	*0.626*
Adj. *p*-val			*0.192*	*0.591*
Difference in treatment	-0.025	-0.199	-0.197	-0.150
	(0.074)	(0.103)	(0.096)	(0.112)
p-val	*0.736*	*0.055*	*0.045*	*0.184*
Adj. *p*-val			*0.070*	*0.134*
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: The sample includes children attending pre-schools randomized 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 externalizing (conduct and hyperactivity) and internalizing (emotional and peer problems) behaviors, 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* program. Control means correspond to children in the control group whose mothers did or did not sign up to participate in the program in the following year.

We estimate impacts on one younger and one older sibling of all target child who have siblings (Section 4), reported in Table 10 for a sample of 749 younger and 1,083 older siblings. We find no evidence of negative spillover effects on either the older or the younger siblings. For younger siblings in PM households, the estimated impact is positive, of an economically meaningful magnitude but imprecisely estimates (0.21 SD, *p*-val=0.244). The effect among

Table 10: 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: The sample consists of the next older and younger siblings of children who attended preschools randomized 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 standardized 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* program. 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 program in the following year.

younger siblings in non-PM households is close to zero (0.03 SD, p-val=0.770). Although these estimates are imprecise, the contrast between them suggests that any spillovers are unlikely to arise from children's direct exposure to the preschool intervention. Instead, they are more plausibly explained by changes in the home environment, such as PMs applying improved ECD knowledge and practices more broadly across their younger children.

6.4 Interpreting differences by mother participation status

While our findings demonstrate that the *Lively Minds* program improves the home environment of PMs in addition to enhancing preschool provision, they also raise the question of whether these gains would extend to any mother induced to participate in the program or whether they are specific to mothers who self-select into participation and their children.

Using the notation introduced in Section 5, the difference in program impact between PMs and non-PMs can be written as:

$$\gamma = \beta^{PM} - \beta^{NM} = \underbrace{\alpha_1 - \alpha_0}_{\text{Indirect effect}} + \underbrace{\mathbb{E}[U_{i1} - U_{i0}|P_i = 1] - \mathbb{E}[U_{i1} - U_{i0}|P_i = 0]}_{\text{Difference in unobservable gains b/w PMs and non-PMs}} .$$

This expression shows that γ captures both the indirect effect of *Lively Minds* program operating through maternal participation and any difference in unobservable gains from the program between the children of PMs and non-PMs. Separately identifying $\alpha_1 - \alpha_0$ would require a source of exogenous variation in P_i, which we do not have. Unless differences in unobservable gains are either zero or fully accounted for by the observable characteristics used to compare PMs and non-PMs, γ may either under- or overstate the average causal effect of participation on outcomes ($\alpha_1 - \alpha_0$).

In what follows, we provide an exploration of the likely importance of sorting-on-gains relative to the causal indirect effect of participation by re-estimating γ in a version of our benchmark model that controls for observable characteristics predictive of participation and their interactions with the PM indicator. If sorting on gains is driving γ, mothers who select into participation should differ systematically from those who do not in ways that also predict differential gains from the program. Controlling for observable characteristics that predict participation, and their interactions with T_i, should therefore absorb this source of variation to the extent that it operates through observables. If estimates of γ remain stable across specifications, this would suggest that observable selection into participation is unlikely to account for the gap in impacts. Such stability cannot, however, rule out sorting on gains operating through unobserved characteristics that are uncorrelated with the controls we include.

The results are presented in Figure B.3. Each of the plots report the coefficient estimating the difference in the impact of the program on cognitive (top) and socio-emotional development

(bottom) between children of PM and non-PMs. In each of these figures, the first bar plots the benchmark coefficient (shown in Table 7 and Table 9); each of the following bars plots this coefficient in a specification where we also control for the variable indicated below the x-axis and the interaction between that variable and the PM indicator. These variables have been selected using a LASSO regression of the PM indicator on all the observable characteristics of the PM, their households and their children measured at baseline.

The coefficient is highly stable across all specifications, indicating that observable differences between PMs and non-PMs — and their differential relationship with outcomes — account for little of the gap in program impacts. This is consistent with γ primarily reflecting the causal indirect effect of the program, though we cannot rule out a residual contribution from sorting on gains driven by unobservables uncorrelated with the covariates that we observe.

6.5 Robustness of findings

Selection of controls. Our main specifications follow the pre-specified approach to selecting controls described in Section 5. As a robustness check, we re-estimate impacts on the cognitive aggregate, emotional awareness, and behavioral difficulties scores, selecting controls using post-double selection LASSO (Belloni et al., 2014). Estimates, shown in Table B.8, are very similar in magnitude and significance to the main results above.

Measurement error. We re-estimate the main specifications using cognitive scores constructed using latent factor analysis to assess the sensitivity of our results to measurement error in the assessment data. As discussed in Appendix A, factor analysis explicitly models measurement error and allows individual questions to differ in difficulty and discrimination, rather than imposing equal weights as in simple averages. Table B.9 shows that the estimated effects on the aggregate cognitive score, emergent numeracy, and executive functions are similar in magnitude and significance to those obtained using raw scores. The estimated effect on emergent literacy is somewhat larger, increasing from 0.063 to 0.085 SD ($p = 0.064$). Table B.10 presents the corresponding socioemotional results. Here too, theestimated effects are very similar to those obtained using raw scores.

Measurement invariance. A concern in interpreting treatment effects is that the intervention may affect how assessment items function rather than the underlying skills themselves, for example if treated mothers (especially treated PMs) become more self-aware of how they interact with their children as a result of the program (Heckman and Zhou, 2022). If measurement invariance does not hold across treatment and control groups and across PMs and non-PMs, comparisons across these four groups may conflate the impact of the intervention on the underlying constructs with its impact on their measurement.

We test for measurement invariance across all aggregate latent constructs, using standard multi-group factor models to check whether items measure the same underlying construct in the same way, for PM and non-PMs in the treatment and control groups (4 groups). We test invariance at the two levels necessary to establish meaningful mean comparisons between these groups: *metric* invariance (whether items relate to the underlying skill with the same strength in both groups) and *scalar* invariance (whether items have the same baseline difficulty in both groups). We test each level by starting from a fully unconstrained model and imposing equality restrictions on the relevant item parameters (factor loadings for metric invariance, and item intercepts for scalar invariance) and comparing model fit at each step. Since full scalar or strict invariance rarely holds in applied early-childhood settings, we allow for *partial* invariance: most items are held equal across groups, but a small number are allowed to differ where they show the clearest evidence of functioning differently by treatment status. 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.10, 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. See Appendix A for a more in-depth discussion and analysis of this issue.

7 Cost-Effectiveness

7.1 Benefit-cost ratio

We estimate the cost-effectiveness of the *Lively Minds* program by projecting the future labour-market benefits implied by its effects on children's cognitive skills, following the cost-benefit framework recently applied by Ganimian et al. (2024).

We begin by computing the present discounted value (PDV) of lifetime earnings for the control group:

$$PDV = \sum_{a=a_0}^{a_1} \frac{Y(1+g)^{a-a_T}}{(1+r)^{a-a_T}} \tag{7.1}$$

where Y_s is annual baseline income in USD, a_T is age at treatment (which is 4.5 years at baseline), a_0 and a_1 are the first and last working ages, g is real earnings growth, and r is the discount rate.

The projected earnings gains implied by the program is then:

$$\Delta E = PDV \times 0.12 \times \rho \tag{7.2}$$

where 0.12 is the estimated treatment effect of *Lively Minds* on cognition (in SD units) and ρ_s denotes the percentage earnings return to a one-SD increase in cognitive skills. The corresponding benefit-cost ratio is ΔE divided by the program's implementation cost of \$37 per child.

A key parameter is ρ_s, the earnings return to a one-SD increase in cognitive skills. Because study participants are currently in their teens, this parameter cannot be estimated directly from our data. We therefore draw on two existing estimates. The first comes from Chetty et al. (2011), an experimental study of improvements in kindergarten classroom quality in the US which finds a 13% earnings gain from a one-SD increase in cognitive skills. This is also the estimate used by Ganimian et al. (2024). The second is based on a recent meta-analysis by Patrinos and Psacharopoulos (2026), which covers 73 studies from 41 countries. Averaging their estimates for the sub-set of studies conducted in LLMICs implies a 16.4% earnings gain from a one-SD increase in test scores. The advantage of the former estimate is that it is based on experimental evidence from skill improvements at a similar stage of childhood to our study. However, it is drawn from a high-income context that differs substantially from ours. The advantage of the latter estimate is that it is based on evidence from LLMICs. However, these estimates are for improvements in skills among older (school-age) children and are drawn largely from observational rather than experimental studies. We therefore view these two estimates as providing plausible lower and upper bounds for the returns to cognitive skills in our setting.

Table 11 reports results from three specifications. All assume a 3% discount rate (in line

with Ganimian et al. (2024)), working age 22 - 60 (where 60 is the compulsory retirement age in Ghana) and rural annual per-capita income from the 2016-1017 Ghana Living Standard Survey Ghana Statistical Service (2019, GLSS7) [11]. The *Rural* specification assumes zero real earnings growth (in line with evidence of stagnating per capita income in Ghana over the last 10 years (World Bank Group, 2025)) and a 13% earnings return per SD increase in cognition following Chetty et al. (2011). The *Rural + Growth* specification keeps the same assumptions but allows annual real earnings to grow by 2%. The *PP* specification retains these assumptions but replaces the earnings return with the LMIC estimate of 16.4% reported by Patrinos and Psacharopoulos (2026). Panel A reports the assumptions and the resulting PDV of lifetime earnings for the control group. Panel B shows the PDV of the projects earnings gain from the *Lively Minds* program, the implied BCR and the components used to calculate these.

These estimates are based on four main assumptions. First, the cognition index in our study is treated as comparable to the test-score measures used in the long-run earnings literature. Second, the short-run cognition gain is assumed to translate into adult earnings according to the selected earnings-return parameters. Third, the main Ghana-specific earnings base uses rural annual per-capita income from GLSS7. Fourth, the estimates include only private earnings benefits; they exclude fiscal externalities, maternal opportunity costs and benefits, socioemotional benefits, and any intrinsic value of improved child development or general-equilibrium effects.

Estimates presented in Table 11 show that the projected private earnings benefits of the *Lively Minds* program exceed program costs in all three specifications. Under the most conservative assumptions (specification 'Rural'), the PDV of lifetime earnings gains is $295 per child. Relative to a program cost of $37 per child, this implies that each $1 spent on a child is projected to generate about $8 in discounted future private earnings, or a net private earnings gain of about $258 for that child. Allowing rural earnings to grow by 2% per year increases the projected PDV of lifetime earnings gains to $579 per child and raises the BCR from 8 to 15.7. The PP specification, which combines the same 2% earnings growth assumption with the higher test-score-to-earnings return reported by Patrinos and Psacharopoulos (2026), produces a projected lifetime earnings gain of $731 per child and a BCR of 19.8. [12]

[11] Income reported in GLSS7 is converted to USD using the 2017 World Bank official exchange rate(World Bank, 2026)]

[12] If we apply the exact set of assumptions used in Ganimian et al. (2024), the estimates are much larger suggesting a projected lifetime earnings gain of $1,855 per child and a BCR of 50.1. This large value is driven by the combination of assuming a higher annual real earnings growth of 5% and a longer working-life (22-65).

From a public-finance perspective, the projected earnings gains imply that the program would pay for itself fiscally if the government recovered between 5% and 13% of beneficiaries' future earnings gains through taxes.

Table 11: Benefit-cost analysis

		Specifications		
Input	Source	Rural	Rural + Growth	PP
Panel A. Projecting future earnings				
Income base	GLSS7 Table 10.17	Rural per capita	Rural per capita	Rural per capita
Annual income, GHS	GLSS7 Table 10.17	5,880	5,880	5,880
Exchange rate, GHS/USD	World Bank WDI, 2017	4.351	4.351	4.351
Annual income, USD	Calculation	1,351	1,351	1,351
Age at treatment	Baseline mean age	4.5	4.5	4.5
Working ages (22-60)	Assumption	39 years	39 years	39 years
Annual real earnings growth	Assumption	0%	2%	2%
Discount rate	From Ganimian et al. (2024)	3%	3%	3%
Average PDV lifetime earnings at 4.5	Calculation	$18,926	$37,138	$37,138
Panel B. Projecting benefits and costs				
Cognitive gain in SD	Lively Minds RCT	0.12	0.12	0.12
Earnings gain per one SD	Literature assumption	13%	13%	16.4%
PDV of lifetime earnings gain per child, USD	$\Delta E_s = PDV_s \times 0.12 \times \rho_s$	$295	$579	$731
Program cost per child, USD	Lively Minds estimate	$37	$37	$37
Benefit-cost ratio (BCR)	**Earnings gain / cost**	**8**	**15.7**	**19.8**

Notes: The preferred headline estimate is the 'Rural' column, with 0% real wage growth, which is the most conservative estimate. The 'Rural + Growth' column adds 2% annual real earnings growth. The PP column uses a higher earnings return per standard deviation. All monetary values are treated as real USD. Only the 'Earnings gain per one SD' assumption in column PP is taken from Patrinos and Psacharopoulos (2026). The remaining 0.13 in that row comes from Chetty et al. (2011)/Ganimian et al. (2024).

Following Ganimian et al. (2024), we also examine how sensitive the benefit-cost estimates are to alternative assumptions about the parameters used to monetize the treatment effect. Figure B.4 reports the results of a simulation exercise in which we repeatedly draw key parameters from plausible ranges and recompute the BCR. Annual rural income is allowed to vary by ±20% around the GLSS7 value, annual real earnings growth ranges from 0% to 2%, the earnings return to a one SD gain in cognition ranges from 7% to 21%, and the discount rate ranges from 1% to 7%.[13] Panel A draws each parameter independently from a uniform distribution over its range. Panel B instead uses independent truncated-normal draws, centered on the preferred value of each parameter and bounded by the same ranges. The figure plots the resulting distribution of simulated BCRs. The dashed vertical lines mark the 5^{th} and 95^{th} percentiles, while the solid black line marks the preferred *lower-bound* estimate from Table 11. The simulated BCR is between 3.5 and 16 for the 5^{th} to 95^{th} percentile range in both panels, suggesting that the conclusion that benefits exceed costs is robust to substantial variation in the calibrated parameters.

7.2 Hidden costs

A common feature of early childhood and related social programs in low-resource LLMIC 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 range 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).[14] The *Lively Minds* program depends on unpaid, low-educated mothers who already shoulder substantial domestic and care responsibilities.

While such models can be low cost, a key welfare and scalability question is whether they impose hidden costs on the participating women. 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.,

[13] Because our earnings measure comes directly from annual rural per-capita income in GLSS7, rather than being constructed from daily wages and days worked as in Ganimian et al. (2024), we vary the annual earnings base directly instead of varying days worked per year.

[14] 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).

Table 12: Psychological impacts of intervention on PMs

	Depressive symptoms	**Self-esteem**
	(1)	(2)
Treatment	0.118	0.027
	(0.128)	(0.114)
p-val	*0.360*	*0.813*
Baseline outcome	0.145	0.002
	(0.054)	(0.051)
p-val	*0.009*	*0.972*
Control mean endline	0.018	0.055
Strata fixed effects	Yes	Yes
Characteristics	Yes	Yes
Observations	437	437

Notes: The sample includes primary caregivers of children attending pre-schools randomized 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 standardized 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.

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 is especially 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, it could be reflected in reduced well-being among the PMs.

Reassuringly, we find no evidence of adverse effects of the program on two measures of PMs' well-being - mental health or self-esteem (Table 12).[15] This aligns with the fact that 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

[15] See Section 4 for details on how these are measured.

activities in this rural setting are informal, seasonal, and flexible, which likely reduced the risk of large crowd-out effects. 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). The *Lively Minds* program training and monthly parenting workshops also explicitly emphasized maternal well-being, which may have further mitigated potential stress.

8 Conclusion

Many LLMICs have expanded access to pre-primary schooling in recent years but are struggling to ensure that children experience high quality learning environments once they enroll. Similar difficulties accompanied the earlier rapid expansion of primary education, where gains in access often outpaced gains in learning and many children continue to progress through school without acquiring even foundational skills (World Bank, 2018b). Preventing a similar disconnect between access and learning at preschool level is becoming increasingly urgent as enrollment continues to grow, particularly as shrinking donor funding, rising debt burdens, and growing student populations put pressure on education budgets across Sub-Saharan Africa (World Bank and UNESCO, 2024).

This paper studies a government-led response to this challenge in rural Ghana. Although Ghana was a regional trailblazer in introducing universal access to public preschool nearly two decades ago, it still faces significant challenges in ensuring that children learn and thrive in these settings. We partnered with the Ghana Ministry of Education and the NGO Lively Minds to evaluate a program designed to leverage parents as a resource for improving the quality of public preschool provision. The program recruits mothers from local communities, 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. A key innovation is that the program uses mothers, rather than hired paraprofessionals or additional teachers, as the marginal classroom input. As a result, implementation itself serves as an investment in parents' knowledge and skills, with the potential to improve the home environment alongside the preschool environment.

We find that the program successfully improves the ECD knowledge and practices of participating mothers, despite the fact that most have little or no formal education. These changes demonstrate that low-educated parents can be effectively trained to support the development of pre-school age children through a relatively light-touch intervention embedded within the public education system. At the child level, we observe economically meaningful gains in cognitive development: children who, at baseline, planned to attend preschools where the play schemes were implemented score 0.12 SD higher on an aggregate cognitive index, irrespective of whether their mothers participate. This pattern suggests that trained mothers can effectively augment preschool capacity and improve learning opportunities for all children in the classroom. The program also reduces behavioral difficulties, but these gains are concentrated among the children of participating mothers. This suggests that engaging parents as implementers can generate additional benefits beyond those arising from improvements in preschool quality alone, creating a second channel through the home environment that would not be present in models relying on salaried paraprofessionals.

These impacts are achieved at relatively low cost. A key reason is that the program relies on volunteer mothers rather than additional salaried staff. During the trial, the program required an additional $37 per child (in implementation-year USD) beyond existing public education resources. This figure overstates the cost of routine implementation because it includes substantial one-off set-up costs, as well as additional expenses arising from the within-district randomization required for the evaluation. Even at this higher cost, our estimates imply that the present discounted value of future earnings gains is around eight times the cost of the program. Importantly, we find no evidence that the program adversely affected the mental health or self-esteem of participating mothers, alleviating concerns about the potential hidden well-being costs of such volunteer-based models in settings where women already shoulder substantial caregiving responsibilities.

At the time of the RCT, the *Lively Minds* program was also operating in two additional districts beyond the two included in this trial, serving around 5,400 children. Following the positive findings from the trial, the Government of Ghana scaled the program to 64 high-need rural districts between 2021 and 2024, reaching 532,413 children — more than 40% of all children enrolled in public preschools nationaly. Administrative records from 46 of these districts, cov-

ering 2,543 preschools, provide encouraging evidence that the program's core implementation model has been sustained at this scale.

The program continues to attract and retain large numbers of participating mothers. During the scale-up, more than 96,000 mothers (37.9 per preschool) enrolled, 95% completed the initial training, and 86% re-enrolled two years later. Unannounced GES monitoring visits also indicate high implementation fidelity: in most observed schools, both parenting workshops and classroom play schemes were operating as scheduled, and more than 90% of observed sessions were rated satisfactory throughout the monitoring period. These patterns are particularly encouraging given evidence that implementation quality often deteriorates as programs move into routine delivery at scale (Banerjee et al., 2017; List et al., 2021).

At the same time, implementation costs have fallen as the program has become more fully integrated into the public preschool system, declining from approximately $37 per child during the trial in 2017 to around $17 during the national roll-out ($14.24 in comparable 2017 dollars). If the impacts observed in the trial are sustained at this scale, this would imply a benefit-cost ratio of at least 20.9. This estimate is itself conservative because it does not account for the further reductions in recurrent implementation costs expected once the initial set-up phase has been completed.

In sum, our findings suggest that engaging parents directly in the delivery of early childhood education can be a promising, scalable and cost-effective way to strengthen both preschool and home learning environments in low-capacity education systems. Whether these results generalize to other contexts facing similar constraints remains an open question. The model's reliance on volunteer mothers may be more effective in contexts where women have flexible work schedules, strong community ties, and strong intrinsic motivation to support children's education - conditions that may not hold uniformly across LLMICs. More research is also needed on equilibrium and welfare implications of such models, including their effects on women's labor supply and intra-household resource allocation, as well as the extent to which volunteer fatigue or changes in the composition of volunteers emerge over the longer-term.

References

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

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review 107*(5), 282–286.

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.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives 31*(4), 73–102.

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.

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.

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.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of economics 126*(4), 1593–1660.

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.

Evans, D. K. and F. Yuan (2022). What we learn about girls' education from interventions that do not focus on girls. *Educational Evaluation and Policy Analysis 44*(2), 244–267.

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.

Ghana Statistical Service (2019). Ghana living standards survey round 7 (glss7): Main report. Technical report, Ghana Statistical Service, Accra, Ghana. Survey data collected October 22, 2016 to October 17, 2017.

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

Heckman, J. J. and J. Zhou (2022, November). Measuring knowledge and learning. NBER Working Paper 29990, National Bureau of Economic Research.

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).

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.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher 49*(4), 241–253.

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.

List, J. A. (2026). The economics of scaling early childhood programs: Lessons from the Chicago school. *Journal of Political Economy 134*(1), 1–48.

List, J. A., D. Suskind, and L. H. Supplee (Eds.) (2021). *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It*. London and New York: Routledge.

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.

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, Ghana (2021). Early childhood education policy framework. https://www.ece-accelerator.org/sites/default/files/2020-11/1.%20ECE%20Policy%20Framework.pdf.

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

National Council for Curriculum and Assessment (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.

Patrinos, H. A. and G. Psacharopoulos (2026). The effect of raising school quality on earnings. *Prospects*.

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.

Pisani, L., I. Borisova, and A. J. Dowd (2015, August). International development and early learning assessment (idela): Technical working paper. Technical report, Save the Children.

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.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from head start. *American Economic Journal: Applied Economics* 7(4), 76–102.

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 (2026). World Development Indicators: Official Exchange Rate (LCU per USD, Period Average), Ghana. https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=GH. Accessed June 30, 2026.

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

World Bank Group (2025). 2025 policy notes: Transforming ghana in a generation. Technical report, World Bank Group, Washington, DC.

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

ONLINE APPENDIX

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 behaviors. 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 aggregate 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 recognize letters
- *fine motor skills*, i.e. the mental control and self-regulation, for example the ability to memorize, or plan and control impulses

- *socioemotional skills*, specifically around the child's expression and management of their own emotion, their ability to empathize 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.[16]

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

[16]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).

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 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 behavioral assessment informed by the structure and content of the Strengths and Difficulties Questionnaire (Goodman, 2001) [17] The instrument includes 25 items describing child behaviors, with responses recorded on a 3-point Likert scale: (0) not true, (1) somewhat true, and (2) certainly true.[18] These items map onto key domains of socioemotional development: conduct problems, hyperactivity, emotional problems, peer problems, and prosocial behavior. Following this framework, conduct problems and hyperactivity are combined as indicators of externalizing behaviors, while emotional and peer problems represent internalizing behaviors. Our analysis focuses on 20 items capturing internalizing and externalizing problem behaviors (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

[17] 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).

[18] Instrument available on request

of population-level measures designed to assess early childhood development for children from 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 socioemotional 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.[19]

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 caregiver'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).

[19] Instrument available on request

In its classic form, the KIDI consists of 75 items, covering a broad range of content relevant to 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.

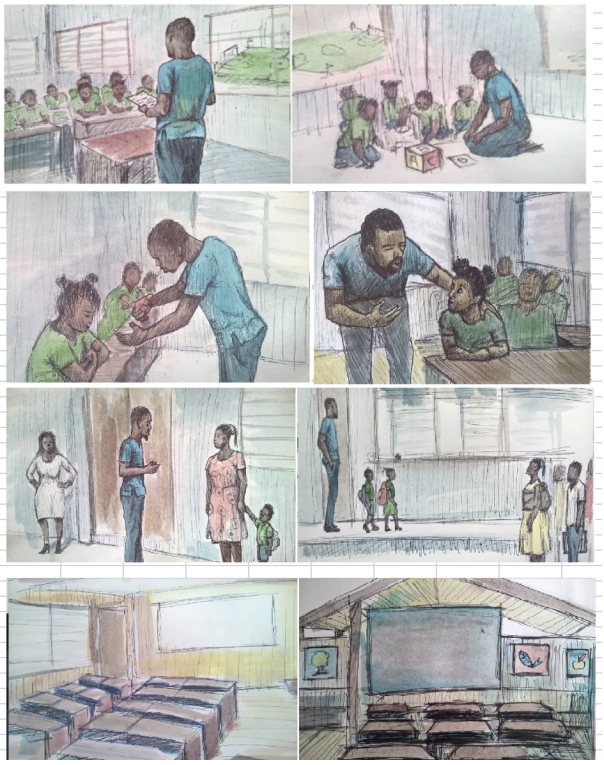


Figure A.1: Vignettes used to measure parental knowledge about pre-school quality

A.1.6 Instrument to measure parenting behaviors

Parenting behaviors 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 behaviors and approaches into two groups: one group of more supportive or positive parenting behaviors and another group of less supporting or more negative parenting behaviors.

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' well being 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 20 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 summarizes 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 behavioral 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 behavioral 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: externalizing behaviors (combining conduct problems and hyperactivity) and internalizing behaviors (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 behaviors 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 behaviors.

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 behavioral problems	**Externalizing behaviors**	SDQ tempers, obeys, fights, lies, steals
		SDQ restless, fidgety, distracted, thinks before acting, good attention
	Internalizing behaviors	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 color, 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, colors, count objects, or say numbers

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

Dimensions	Sub-dimensions	Measurements
Parenting behaviors	**Positive/supporting parenting behaviors from instrument constructed for study**	***Surveyor's observations while the primary carer plays a game with the child***
		Ways the mother encouraged good behavior
		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 behaviors the primary caregiver 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 caregiver 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 vocalizations

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

Dimensions	Sub-dimensions	Measurements
Parenting behaviors (cont.)	**Negative/harsh parenting behaviors from instrument constructed for study**	***Surveyor's observations while the primary carer plays a game with the child***
		Ways the mother dealt with bad behavior
		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 behaviors 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 construct. Specifically, we analyze 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
Emotional awareness	0.61	0.66
Emotional and behavioral problems	0.81	0.75
Younger sibling outcomes		
CREDI	-	-
Older sibling outcomes		
Aggregate score	0.71	0.67
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 behaviors		0.81
Negative parenting behaviors		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. Among socioemotional measures, emotional awareness shows moderate reliability, while the SDQ-based measure of emotional and behavioral problems displays good internal consistency overall.

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.

Taken together, the results indicate that the main aggregate constructs—particularly cognitive skills and emotional and behavioral 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 characterized 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 potentially stronger for cognitive outcomes and emotional awareness than for emotional and behavioral 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 behaviors, which will have been potentially been differently affected by the intervention and whether a mother was participating in the *Lively Minds* program.

Table A.5 reports the results of these tests for child outcomes. For the overall cognitive construct and emergent literacy, metric invariance is initially rejected; however, partial metric invariance can be achieved by freeing a small number of factor loadings. Scalar invariance is initially rejected, but partial scalar invariance is achieved by freeing up a few intercepts.

Turning to socioemotional outcomes, metric and scalar invariance is retained from emotional awareness. For the overall measure of emotional and behavioral problems, metric invariance is initially rejected but partial metric invariance is achieved by freeing a single factor loading;

scalar invariance is then rejected, though partial scalar invariance can be restored with a small number of freed intercepts.

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 positive parenting behaviors, metric invariance is rejected, but partial metric invariance is achieved by freeing a small number of factor loadings; scalar invariance is likewise rejected, with partial scalar invariance restored by freeing a limited set of intercepts and one factor loading. For negative parenting behaviors, metric invariance is rejected, but partial metric and partial scalar invariance can both 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 rejected, but partial metric invariance is achieved by freeing a small number of factor loadings, after which scalar invariance is not rejected. 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 limited deviations from the baseline model, the main exception being externalizing behaviors, for which no level of invariance could be established. This broadly supports the use of a common latent framework for comparing outcomes across groups, while suggesting that group comparisons on externalizing behaviors specifically should be treated with caution.

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 (scores A). Second, for constructs where invariance is largely supported, we estimate factor scores imposing equality of parameters across groups (scores B). Third, for transparency and comparability with existing work, we construct indices based on simple averages of item scores (scores C).

In the main analysis, we report results based on averaged indices (scores C), reflecting both

their simplicity and their alignment with standard practice. 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. 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. See Table A.7, Table A.8, and Table A.9 for analysis using factor scores imposing measurement invariance across groups (method B) and Table A.10 for analyzing using factor scores allowing for partial measurement invariance (method C).

Table A.5: Measurement invariance tests: Child outcomes

Construct	Test	p-value	Conclusion	Parameters freed
Cognitive skill (overall)	Metric vs. Configural	0.019	Metric invariance rejected	
	Partial metric vs. Configural	0.186	Partial metric invariance retained	Factor loadings: Extra number, IDELA 17
	Scalar vs. Configural	0.037	Scalar invariance rejected	
	Partial Scalar vs Configural	0.378	Partial scalar invariance retained	Intercepts: IDELA 8, 3
Emotional awareness	Metric vs. Configural	0.900	Metric invariance not rejected	
	Scalar vs. Configural	0.389	Scalar invariance not rejected	
Emotional and behavioral problems	Metric vs. Configural	0.004	Metric invariance rejected	
	Partial metric vs. Configural	0.297	Partial metric invariance retained	Factor loading: SDQ steals
	Scalar vs. Configural	0.013	Scalar invariance rejected	
	Partial scalar vs. Configural	0.123	Partial scalar invariance retained	Intercepts: SDQ popular, obeys, attends, oldbest

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 summarizes 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
Knowledge of child development	Metric vs. Configural	0.223	Metric invariance not rejected	
	Scalar vs. Configural	0.123	Scalar invariance not rejected	
Positive parenting behaviours	Metric vs. Configural	0.003	Metric invariance rejected	
	Partial metric vs. Configural	0.218	Partial metric invariance retained	Factor loadings: reward, spontaneous, kiss, strategy
	Scalar vs. Configural	0.000	Scalar invariance rejected	
	Partial scalar vs. Configural	0.151	Partial scalar invariance retained	Intercepts: spontaneous, focus, praise, reward, visual, kiss; Factor loading: voice
Negative parenting behaviours	Metric vs. Configural	0.000	Metric invariance rejected	
	Partial metric vs. Configural	0.097	Partial metric invariance retained	Factor loadings: hit (interview), hit (obs), shook (interview), shout (interview), spank (interview), shook (obs), shout (obs), gave (interview), exp (obs), spank (obs), gave (obs), called (interview)
	Scalar vs. Configural	0.025	Scalar invariance rejected	
	Partial scalar vs. Configural	0.615	Partial scalar invariance retained	Intercepts: gave (interview), gave (obs)
Depressive symptoms	Metric vs. Configural	0.301	Metric invariance not rejected	
	Scalar vs. Configural	0.001	Scalar invariance rejected	
	Partial scalar vs. Configural	0.121	Partial scalar invariance retained	Intercepts: lost interest, tired easily, frightened, tired always, sleep
Self-esteem	Metric vs. Configural	0.000	Metric invariance rejected	
	Partial metric vs. Configural	0.408	Partial metric invariance retained	Factor loadings: proud, self-respect, useless, failure, good enough
	Scalar vs. Configural	0.217	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.7: Impacts on cognitive outcomes measured using factor scores (assuming measurement invariance), by PM status

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.147	0.045	0.186	0.133
	(0.068)	(0.071)	(0.079)	(0.085)
p-val	*0.035*	*0.530*	*0.021*	*0.120*
Adj. *p*-val		*0.521*	*0.048*	*0.196*
Treatment, non-PM children	0.080	0.082	0.075	0.130
	(0.054)	(0.057)	(0.054)	(0.052)
p-val	*0.145*	*0.156*	*0.165*	*0.014*
Adj. *p*-val		*0.170*	*0.170*	*0.030*
Difference in treatment	0.066	-0.037	0.111	0.004
	(0.064)	(0.065)	(0.085)	(0.078)
p-val	*0.306*	*0.570*	*0.194*	*0.964*
Adj. *p*-val		*0.768*	*0.373*	*0.964*
Control mean non-PM children	0.015	0.002	0.021	0.004
Control mean PM children	0.141	0.090	0.155	0.098
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 factor score obtained from a confirmatory factor analysis of the IDELA and Spelke tasks measuring emergent literacy, numeracy, and executive function. Columns (2)–(4) report the corresponding subdomains measured using a confirmatory factor analysis of the relevant tasks. All specifications estimate fully interacted models controlling for baseline outcomes (measured as factor scores), 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 A.8: Impacts on children's socio-emotional outcomes measured as factor scores (assuming measurement invariance), by PM status

	Emotional Awareness	**Emotional Behavioural Problems**		
		Overall	**Externalising Behaviours**	**Internalising Behaviours**
	(1)	(2)	(3)	(4)
Treatment, PM children	0.016	-0.205	-0.248	-0.136
	(0.088)	(0.124)	(0.119)	(0.121)
p-val	*0.854*	*0.102*	*0.040*	*0.266*
Adj. *p*-val			*0.060*	*0.228*
Treatment, non-PM children	0.042	-0.079	-0.078	-0.060
	(0.075)	(0.059)	(0.048)	(0.073)
p-val	*0.572*	*0.184*	*0.105*	*0.411*
Adj. *p*-val			*0.116*	*0.341*
Difference in treatment	-0.026	-0.126	-0.170	-0.076
	(0.073)	(0.107)	(0.110)	(0.107)
p-val	*0.722*	*0.245*	*0.128*	*0.482*
Adj. *p*-val			*0.190*	*0.461*
Control mean non-PM children	0.011	-0.058	-0.057	-0.044
Control mean PM children	0.060	0.002	-0.007	0.011
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 a factor score obtained from a confirmatory factor analysis (CFA) of IDELA and Spelke tasks measuring emotional awareness. Column (2) is a factor score obtained from a CFA of SDQ items capturing emotional, peer, conduct, and hyperactivity problems. Columns (3) and (4) report factor scores 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 A.9: 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	0.424	0.200	-0.077	0.182	0.218	0.042
	(0.131)	(0.120)	(0.143)	(0.049)	(0.124)	(0.141)
p-val	*0.002*	*0.101*	*0.590*	*0.000*	*0.084*	*0.764*
Adj. p-val	*0.026*	*0.299*	*0.808*	*0.010*	*0.299*	*0.808*
Treatment among non-PMs	-0.034	0.184	-0.246	-0.032	0.040	0.022
	(0.102)	(0.059)	(0.108)	(0.036)	(0.072)	(0.103)
p-val	*0.739*	*0.003*	*0.026*	*0.385*	*0.582*	*0.830*
Adj. p-val	*0.912*	*0.014*	*0.086*	*0.770*	*0.882*	*0.912*
Difference in treatment	0.458	0.016	0.169	0.214	0.178	0.020
	(0.118)	(0.120)	(0.126)	(0.042)	(0.099)	(0.141)
p-val	*0.000*	*0.895*	*0.184*	*0.000*	*0.075*	*0.887*
Adj. p-val	*0.004*	*0.976*	*0.367*	*0.002*	*0.210*	*0.976*
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 standardized 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 A.10: Treatment effects by participation status: Factors allowing for partial measurement invariance

	I. Child Outcomes		**II. Parenting Behaviors**		**III. Maternal Wellbeing**	
	Cognitive	Behavioural Problems	Positive Behaviours	Negative Behaviours	Depressive Symptoms	Self-esteem
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment among non-PMs	0.082	-0.082	0.049	0.044	0.069	-0.075
	(0.058)	(0.056)	(0.072)	(0.098)	(0.075)	(0.094)
p-val	*0.162*	*0.146*	*0.500*	*0.657*	*0.360*	*0.428*
Treatment among PMs	0.127	-0.262	0.284	-0.028	0.160	0.174
	(0.069)	(0.121)	(0.129)	(0.087)	(0.122)	(0.097)
p-val	*0.069*	*0.033*	*0.031*	*0.747*	*0.195*	*0.078*
Difference in treatment	0.045	-0.180	0.235	-0.072	0.091	0.249
	(0.064)	(0.104)	(0.103)	(0.098)	(0.113)	(0.099)
p-val	*0.482*	*0.088*	*0.025*	*0.468*	*0.425*	*0.014*
Control Mean Endline non-PMs	0.006	-0.063	0.009	0.051	0.008	-0.017
Control Mean Endline PMs	0.147	0.001	0.236	-0.066	0.023	0.162
Baseline outcome	Yes	Yes	No	No	Yes	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,151	2,151	2,151	2,121	2,151	2,151

Notes: All outcomes are raw scores standardised to have mean 0, std 1 in the control group. The sample includes children successfully tracked at endline. OLS estimates are presented where standard errors are clustered at the [cluster level – confirm], and p-values are reported in italics. Throughout we estimate fully interacted models that control for the baseline outcome (where available), strata fixed effects, and child/household characteristics, and interactions of all these variables and the treatment indicator with an indicator for whether the child's mother participated in the play/parenting scheme (, participating mother). and report the treatment coefficient for each subgroup from a single fully-interacted regression; reports the coefficient on the treatment × PM interaction, testing equality of treatment effects across subgroups. Control means are reported for each subgroup at the bottom of the table.

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	2,407
Number of children 16 or under	3.676	4.417	4.205	4.474	0.009	0.602	2,407
Number of children 6 or under	1.332	1.617	1.536	1.639	0.045	0.667	2,407
Wealth index (z-score)	-0.212	0.025	0.025	-0.025	0.067	0.361	2,394
Main income source:							
Farming own land	0.359	0.446	0.435	0.438	0.044	0.842	2,407
Wage work	0.328	0.274	0.276	0.284	0.665	0.221	2,407
Livestock	0.047	0.033	0.035	0.033	0.036	0.075	2,407
Profits from small enterprise	0.219	0.183	0.183	0.191	0.254	0.815	2,407
B. Primary carers							
Age	36.648	36.435	36.092	36.833	0.687	0.744	2,407
Target child's biological mother	0.656	0.790	0.789	0.763	0.000	0.949	2,407
Some education	0.289	0.196	0.211	0.201	0.004	0.575	2,407
Illiterate	0.859	0.926	0.915	0.923	0.003	0.589	2,407
Raven's score	45.280	43.178	43.094	43.717	0.385	0.480	2,407
Depression score	-0.211	0.052	0.051	-0.003	0.019	0.492	2,407
Self-esteem score	0.155	-0.029	0.014	-0.033	0.000	0.014	2,407
Knowledge of child development	-0.166	-0.008	-0.030	-0.020	0.354	0.225	2,407
Play materials	0.305	0.053	0.143	0.014	0.009	0.869	2,407
Any play activities with child	0.129	0.132	0.141	0.121	0.914	0.796	2,407

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	2,405
Gender	0.551	0.492	0.498	0.499	0.081	0.483	2,407
Currently in school	0.766	0.781	0.779	0.780	0.870	0.734	2,407
In the last 30 days:							
Any cough	0.512	0.549	0.567	0.522	0.975	0.281	2,406
Any stomach pain	0.477	0.508	0.532	0.478	0.677	0.122	2,406
Any high fever	0.340	0.340	0.359	0.321	0.651	0.522	2,407
Any diarrhea	0.287	0.320	0.313	0.321	0.116	0.251	2,395

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 using sample of PMs: households and primary caregivers 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*	*0.840*	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*	*0.866*	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)	-0.009	-0.070	0.031	*0.285*	*0.870*	437
	(0.971)	(1.039)	(0.923)			
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*	*0.295*	437
	(1.140)	(1.009)	(1.212)			
Any play activities with child (0/1)	0.121	0.098	0.137	*0.220*	*0.866*	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.4: 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.5: Parental knowledge about preschool quality

	Supportive teaching	**Play-based learning**	**Parents' involvement**	**Child-friendly buildings**
Treatment among PMs	0.006	0.105	-0.053	0.058
	(0.042)	(0.061)	(0.047)	(0.035)
p-val	*0.886*	*0.089*	*0.265*	*0.106*
Adj. *p*-val	*0.860*	*0.293*	*0.431*	*0.293*
Treatment among non-PMs	0.042	0.089	-0.023	0.022
	(0.030)	(0.027)	(0.028)	(0.016)
p-val	*0.158*	*0.001*	*0.423*	*0.191*
Adj. *p*-val	*0.365*	*0.006*	*0.365*	*0.365*
Difference in treatment	-0.036	0.016	-0.030	0.036
	(0.040)	(0.059)	(0.049)	(0.040)
p-val	*0.371*	*0.787*	*0.543*	*0.372*
Adj. *p*-val	*0.786*	*0.786*	*0.786*	*0.786*
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.6: 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	-0.023	0.014	-0.032	-0.072	-0.029	0.003	0.000	0.016	-0.001	0.007	0.022	0.010	-0.015	0.014	0.030
	(0.075)	(0.007)	(0.037)	(0.060)	(0.036)	(0.003)		(0.016)	(0.036)	(0.062)	(0.058)	(0.025)	(0.012)	(0.008)	(0.013)
p-val	*0.762*	*0.044*	*0.388*	*0.236*	*0.424*	*0.297*		*0.311*	*0.987*	*0.908*	*0.707*	*0.673*	*0.213*	*0.074*	*0.024*
Treatment non-PMs	0.011	-0.007	-0.009	-0.080	-0.034	-0.001	-0.002	-0.018	-0.032	-0.065	-0.035	-0.035	-0.007	-0.003	0.017
	(0.045)	(0.005)	(0.027)	(0.030)	(0.026)	(0.002)	(0.004)	(0.011)	(0.032)	(0.039)	(0.035)	(0.019)	(0.004)	(0.004)	(0.010)
p-val	*0.809*	*0.178*	*0.734*	*0.010*	*0.196*	*0.400*	*0.721*	*0.128*	*0.321*	*0.099*	*0.324*	*0.067*	*0.141*	*0.522*	*0.101*
Difference in treatment	-0.034	0.021	-0.023	0.008	0.004	0.005	0.002	0.034	0.031	0.072	0.057	0.045	-0.008	0.017	0.013
	(0.063)	(0.008)	(0.045)	(0.062)	(0.031)	(0.003)	(0.004)	(0.019)	(0.037)	(0.056)	(0.052)	(0.022)	(0.011)	(0.008)	(0.015)
p-val	*0.597*	*0.012*	*0.612*	*0.895*	*0.886*	*0.185*	*0.721*	*0.080*	*0.405*	*0.201*	*0.275*	*0.046*	*0.490*	*0.033*	*0.391*
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-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 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.7: 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	0.061	0.111	0.026	0.016	0.029	0.031	0.020
	(0.020)	(0.032)	(0.025)	(0.036)	(0.011)	(0.013)	(0.016)
p-val	*0.003*	*0.001*	*0.308*	*0.651*	*0.009*	*0.024*	*0.192*
Treatment among non-PMs	-0.002	0.000	-0.024	-0.011	0.003	-0.010	-0.004
	(0.018)	(0.019)	(0.019)	(0.031)	(0.005)	(0.008)	(0.007)
p-val	*0.926*	*0.987*	*0.201*	*0.722*	*0.589*	*0.210*	*0.527*
Difference in treatment	0.063	0.111	0.049	0.027	0.026	0.041	0.025
	(0.023)	(0.029)	(0.025)	(0.031)	(0.013)	(0.014)	(0.018)
p-val	*0.009*	*0.000*	*0.055*	*0.375*	*0.040*	*0.006*	*0.160*
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*-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 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.8: Impacts of intervention on primary outcomes: Controls selected using post-double selection LASSO

	Cognition	**Emotional Awareness**	**Emotional and behavioral problems**
	(1)	(2)	(3)
Treatment	0.1125**	0.0487	-0.1007*
	(0.0506)	(0.0703)	(0.0587)
p-val	*0.026*	*0.489*	*0.087*
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).

Table B.9: Impacts on children's cognitive outcomes using factor scores

	Cognition (overall score)	**Emergent Literacy**	**Emergent Numeracy**	**Executive Function**
	(1)	(2)	(3)	(4)
Treatment	0.103	0.085	0.108	0.133
	(0.053)	(0.055)	(0.050)	(0.051)
p-val	*0.053*	*0.122*	*0.034*	*0.011*
Adj. *p*-val		*0.064*	*0.042*	*0.020*
Baseline outcome	0.489	0.361	0.440	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 5 note for more details.

Table B.10: Impacts of intervention on children's socioemotional outcomes using factor scores

	Emotional Awareness	**Emotional Behavioral Problems**		
		Overall	**Externalizing Behaviors**	**Internalizing Behaviors**
	(1)	(2)	(3)	(4)
Treatment	0.052	-0.096	-0.100	-0.072
	(0.074)	(0.063)	(0.053)	(0.072)
p-val	*0.481*	*0.133*	*0.063*	*0.315*
Adj. *p*-val			*0.080*	*0.224*
Baseline outcome	0.232	0.048	0.064	0.044
	(0.024)	(0.026)	(0.022)	(0.029)
p-val	*0.000*	*0.067*	*0.005*	*0.134*
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 standardized to have mean 0, std 1 in the control group. See Table 8 note for more details.

Table B.11: Treatment effects on cognitive outcomes

	IDELA (raw score)	**IDELA (factor score)**
	(1)	(2)
Treatment	0.102	0.105
	(0.059)	(0.058)
p-val	*0.089*	*0.073*
Baseline outcome	0.482	0.518
	(0.029)	(0.029)
p-val	*0.000*	*0.000*

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 preschools randomized 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 all IDELA tasks (across emergent literacy, numeracy, executive function, fine motor, gross motor and emotional awareness). The outcome in Column (2) is a factor score measured with all IDELA tasks. Outcomes are standardized to have mean 0 and standard deviation 1 in the control group.

Figure B.1: The *Lively Minds* program structure for the children

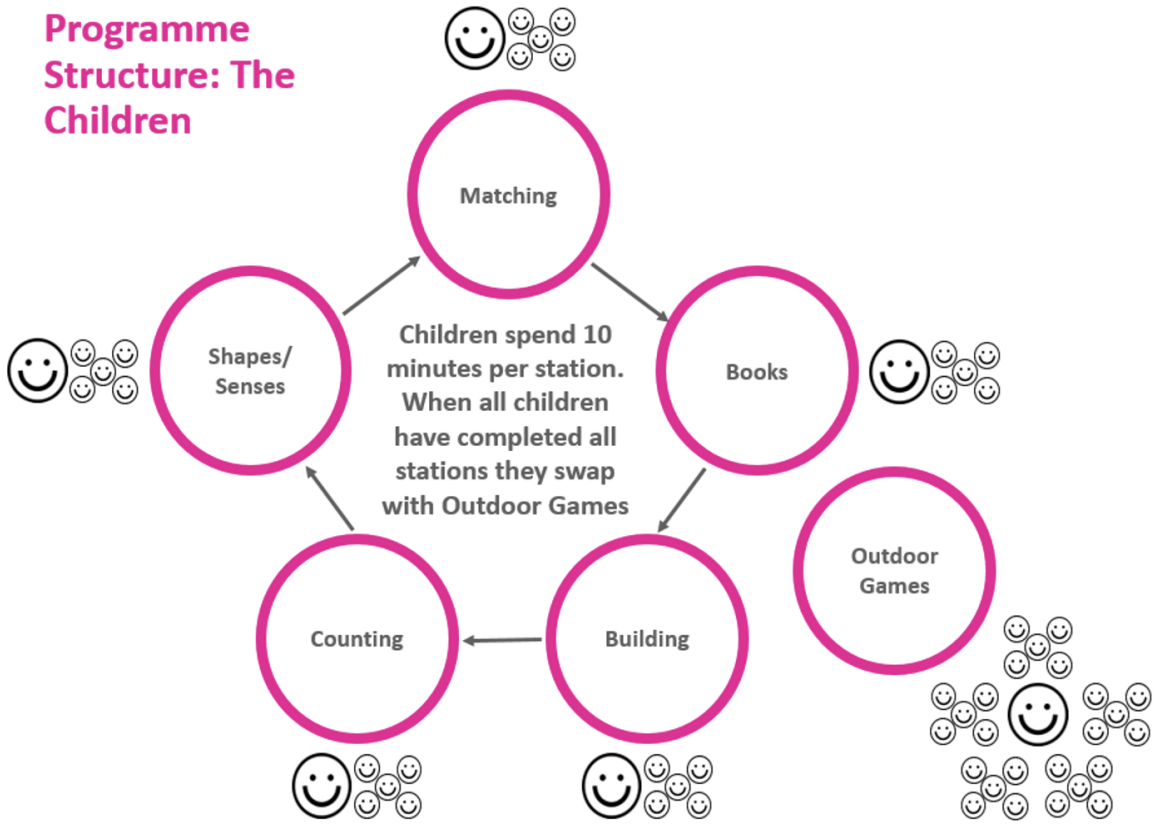
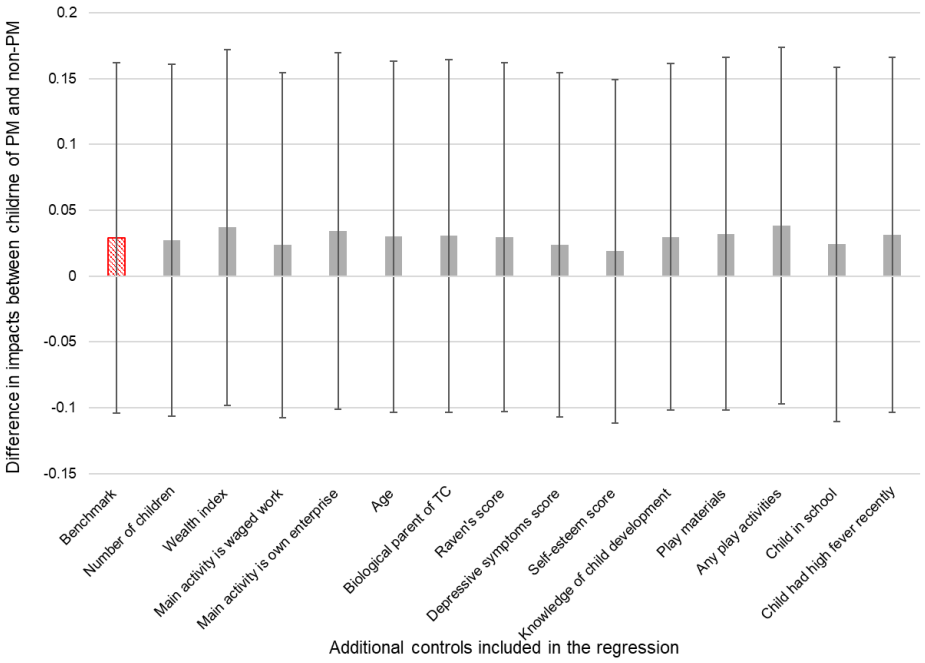


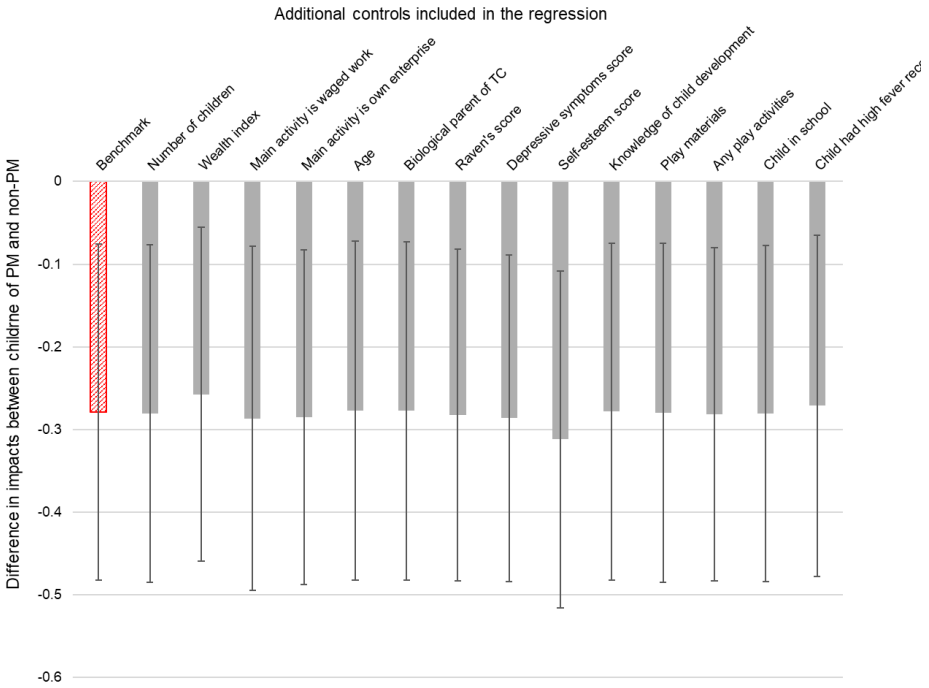
Figure B.2: The *Lively Minds* program indoor activities for the children

Note: The figure shows the main indoor activities of the *Lively Minds* program, such as matching/sorting, numeracy tasks, sizes/colors/senses, solving puzzles, etc.

Figure B.3: Difference in intervention impact on children's outcomes between children of PM and non-PM: Robustness check



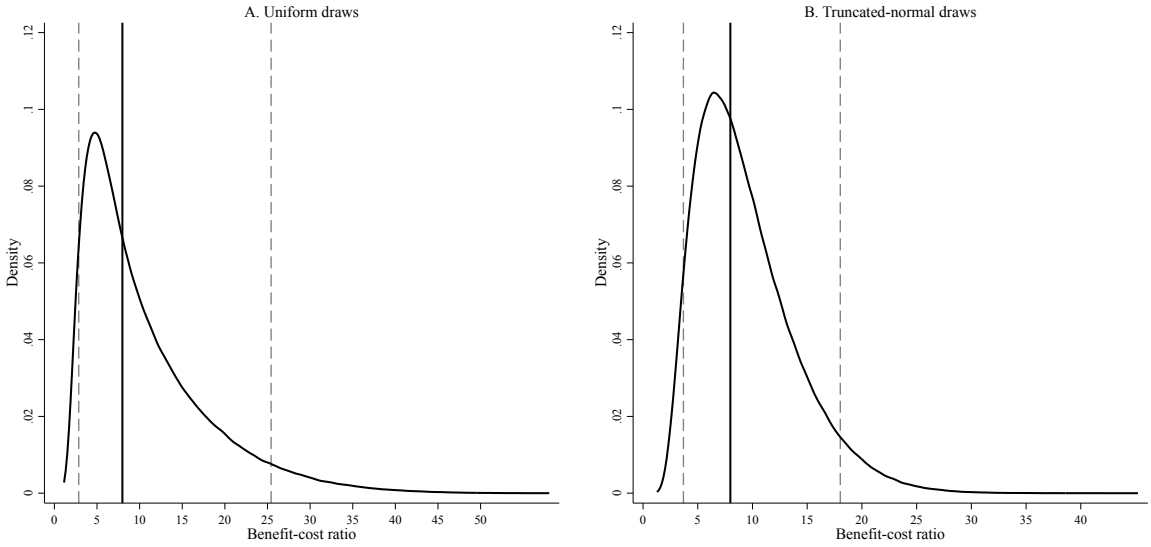
(a) **Cognition (overall score)**



(b) **Emotional and behavioral 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 behavioral 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 7 and Table 9 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.

Figure B.4: Benefit-cost sensitivity analysis



Notes: This figure reports a sensitivity analysis for the *Lively Minds* benefit-cost ratio. Each panel plots the distribution of simulated benefit-cost ratios from 1 million draws of the calibrated parameters. Panel A draws each parameter independently from a uniform distribution over its plausible range. Panel B draws each parameter independently from a truncated-normal distribution centered on the preferred value. The solid black vertical line marks the lower-bound Rural BCR, and the dashed gray vertical lines mark the 5^{th} and 95^{th} percentiles of the simulated BCR distribution.

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 over reading 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.