

**Improving Early Childhood Development in Rural Ghana
Through Scalable Community-Run Play Schemes:
Programme Impact Evaluation Report**



Improving Early Childhood Development in Rural Ghana Through Scalable Community-Run Play Schemes: Programme Impact Evaluation Report

Institute for Fiscal Studies (London, UK)

Orazio Attanasio, Bet Caeyers, Sarah Cattan, Lina Cardona Sosa, Sonya Krutikova,
Peter Leighton

Innovation for Poverty Action (Ghana)

Salifu Amadu, Mubarik Yakubu

July 2019

Acknowledgements: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 695300 - HKADeC - ERC-2015-AdG/ERC-2015-AdG). We also gratefully acknowledge support from the Jacobs Foundation and the Global Innovation Fund.

Address correspondence to: Sonya Krutikova, sonya_k@ifs.org.uk

Contents

Acknowledgements.....	6
1 Introduction.....	7
2 The <i>Lively Minds</i> Programme	10
2.1 Programme Components	10
2.2 Programme Theory of Change	11
3 Research Design	13
3.1 Evaluation Design	13
3.2 Evaluation Implementation.....	13
3.3 Outcome measures	17
3.4 Data collection procedures	21
4 Study context.....	24
4.1 Households' demographic and economic profile.....	24
4.2 Child development	25
4.3 Hand washing practice and health	28
4.4 Schools and teachers.....	29
4.5 Parents and the home environment	30
4.6 Volunteer mothers and their children	32
5 Programme Implementation.....	33
5.1 Compliance	33
5.2 Implementation challenges.....	36
6 Empirical strategy.....	38
6.1 Estimating equation	38
6.2 Balance tests.....	38
6.3 Construction of primary outcome measures	39
Measures of Child Development	39
Measures of Child Health.....	41
7 Impacts on primary outcomes	43
7.1 Cognitive domains.....	43
7.2 Socio-emotional domains.....	43
7.3 Health	45
7.4 Discussion	45
8 Mechanisms	47
8.1 Impacts on intermediary outcomes	47

Impacts on Primary Caregivers	47
Impacts on teachers.....	51
8.2 Mediating analysis.....	54
8.3 Spillover Effects	61
8.4 Impact Heterogeneity	62
9 Discussion and conclusions	66
References	69
Appendix A Study protocol.....	72
Appendix B Census sampling strategy.....	79
Appendix C Data collection procedures	80
Appendix D Study context descriptive	84
Appendix E Baseline balance between control and treatment groups	104

Acknowledgements

This report presents the main findings of the project “Improving early childhood development in rural Ghana through scalable low-cost community-run play schemes”. The project is run in collaboration between The Institute for Fiscal Studies (UK), Lively Minds (UK, Ghana) and Innovations for Poverty Action (IPA, Ghana), and is funded by the Jacobs Foundation and Global Innovation Fund (GIF).

The views expressed in this report are, however, those of the authors and do not necessarily reflect the views of the funders or of other individuals or institutions mentioned here, including the Institute for Fiscal Studies (IFS), which has no corporate view. The funders had no role in the study design, data collection, analysis or interpretation of results. Data was collected in collaboration with IPA. IPA bears no responsibility for the interpretation of the data in this report. All respondents agreed to participate in the surveys, and were assured of the confidentiality of any identifying information gathered. This research was approved by the ethics board of the University College London, UK, Number: 10167/001, as well as the ethics review of IPA, USA, Number: 14340, and the ethics review committee of the Ghana Health Services (GHSER), Ghana, Number: GHSERC012/07/17.

The authors would like to thank Madeleen Husselman, Ishmail Azindoo Baako, Joyce Jumpah and Lise Masselus of IPA, as well as countless field staff, for their valuable contributions during the design and implementation of the data collection. We would also like to thank Alison Naftalin, Chris Lewis, Sheena Lahren and all of the staff at Lively Minds Ghana, for the fruitful cooperation from the very beginning of this project, and their commitment to the implementation of the intervention and to the integrity of the study design. We are grateful for useful suggestions and collaboration from Marc Bornstein, Elizabeth Spelke, Sharon Wolf and Jere Behrman, and for excellent research assistance from Harry Gosling. Any errors and all views expressed are those of the authors.

1 Introduction

There is now general agreement among researchers and practitioners that early childhood care and education (ECCE) is critical to children's development and their success in adult life. The World Bank recently produced a report arguing “Investing in early years is one of the smartest things a country can do for equity and long-term prosperity” (World Bank, 2016). Good quality ECCE can help children be healthier, stay at school longer and have better trajectories in adult life (Cawley, 2001; Heckman, 2006; van Der Gaag, 2010). Children living in remote rural communities of northern Ghana do not receive these vital opportunities. Although Ghana has relatively advanced ECCE policies and has introduced two compulsory years of Kindergarten (KG) into the primary education system (for ages 4-5), two barriers to ECCE persist. First, the quality of KG is low and marred by a lack of trained teachers, large class sizes, lack of play-based resources, teacher absenteeism and rote-based teaching. Second, levels of maternal education and knowledge about best practice in ECCE in deprived rural communities, where most families live on less than US\$2 per day, are very low. Median educational attainment for women is 0 years in Northern Region and 2.9 years in Upper East Region, compared to a national median of 7.2 years (2014 Ghana Demographic and Health Survey).

The *Lively Minds* programme aims to reduce some of these barriers through training uneducated marginalised volunteer mothers and KG teachers to have the knowledge, skills and confidence to run free educational Play Schemes (PS) in KG classes and to provide better care and stimulation at home, using educational games made from cheap local materials. Having refined programme content through carefully controlled small-scale implementation over eight years in rural Ghana and Uganda, *Lively Minds* have begun integration of programme implementation into government education service provision in order to scale the programme to reach more families in a sustainable way. A critical first step is transition to a training of trainers approach in which rather than directly implementing training of the volunteer mothers and overseeing implementation of the programme, *Lively Minds* started working with Ghana Education Service to train KG teachers, who, in turn, train the volunteer mothers and oversee the implementation of the programme in their pre-schools.

In this study we present the results of a Randomised Controlled Trial (RCT) of the training of trainers model of the *Lively Minds* programme. We implemented the trial in Tolon and Bongo districts of Northern and Upper East Regions, randomly allocating 80 KGs, which had expressed an interest in adopting the programme, to either a group that received the programme immediately or to a group that was waitlisted for a year. The waitlisted KGs make up our control group; the focus of the evaluation is on comparing developmental indicators of a sample of 1,191 children who attended these control KGs to those of a sample of 1,216 children who attended KGs that were implementing the *Lively Minds* programme. We study impacts after one year on key developmental domains related to school readiness (including literacy, numeracy, executive functioning, fine motor skills and socio-emotional development), as well as health. We also consider impacts on the Volunteer Mothers (VMs) and KG teachers who run the Play Schemes and check for spillover effects by assessing siblings of the children targeted by the programme.

As is set out in the study protocol (Appendix A), the aim of the evaluation is to go beyond “black-box” estimates of impacts of *Lively Minds* on child development to shed light on impact pathways and identify features of the programme that are especially important for its effectiveness. To this end we test the following hypotheses:

1. *Lively Minds* has a positive impact on participating and non-participating children’s physical, cognitive and socio-emotional development and hygiene practices;
2. This is achieved through both a direct effect of the participation in the Play Schemes (participating children) as well as indirect effects (participating and non-participating children);
3. *Lively Minds* indirectly affects children’s outcomes through improving mothers’ stimulation and care practices (through acquiring knowledge in the training sessions and interacting with Volunteer Mothers);
4. *Lively Minds* indirectly affects children’s outcomes through improving Volunteer Mothers’ psychological wellbeing;
5. *Lively Minds* indirectly affects children’s outcomes through increasing teacher knowledge and awareness of good stimulation and hygiene practices (through training) and reduced teacher absenteeism (through higher motivation and improved accountability as a result of mothers’ enhanced pre-school engagement);
6. *Lively Minds* has indirect impact on children who have not participated in the Play Schemes through interaction with siblings who have;

In order to test these hypotheses we collected two rounds of data between September 2017 and September 2018. In designing the survey and child assessment instruments we placed a lot of emphasis on introducing innovative measurement tools in order to capture more subtle dimensions of children’s development, as well as features of their home and pre-school environments likely to be important for their development.

We find that the programme has significant positive impacts on children’s cognitive and socio-emotional development as well as health. The impacts on cognitive development are driven by improvements in numeracy, executive functioning and fine motor skills. Impacts on socio-emotional skills are mainly due to a reduction in negative externalising behaviours such as hyperactivity and conduct problems. There is an overall significant reduction in acute malnutrition, as measured by upper-arm circumference, but no evidence of an improvement in child health based on incidence of chronic and short-term illness as reported by children’s primary caregivers. While improvements in cognitive skills are similar across genders but greater for children from lower socio-economic groups, effects on socio-emotional development are confined to boys. We find significant improvements in the knowledge and child related practices of the Volunteer Mothers. There is strong suggestive evidence that these improvements are an important mechanism for programme impacts on socio-emotional development but not cognitive development or health, which appear to mainly reflect direct impacts of the Play Schemes.

The findings of this study provide promising new evidence on the potential of scalable models of ECCE programmes to be integrated into existing government services in remote deprived rural contexts in Low Income Countries and continue to be effective.

The rest of the report is organised as follows: Section 2 describes the *Lively Minds* programme in more detail and sets out the Theory of Change. The research design is then presented in Section 3 including details of the evaluation design and implementation, measurement of key outcomes and data collection procedures. We describe the study context in Section 4 and present evidence on programme implementation and compliance during the trial in Section 5. Our empirical strategy is explained in Section 6. Sections 7 and 8 present the main findings: we discuss impacts on primary outcomes in Section 7 and explore potential mechanisms in Section 8. Section 9 discusses the broader relevance of the results and concludes.

2 The *Lively Minds* Programme

The intervention is being carried out by *Lively Minds*, an award-winning organisation that has been running the programme in rural Ghana (as well as Uganda) since 2008. The intervention focuses on unlocking the potential of caregivers, both Volunteer Mothers (VMs) and teachers, training and empowering them with the knowledge, skills and confidence to run daily, one-hour long educational Play Schemes (PS) in KG classes (targeting children aged 4 and 5) and provide better care and stimulation at home, using cheap local materials. At the PS, children rotate around 6 play stations (5 in-door and 1 out-door) lead by VMs and play fun and interactive games that strengthen cognitive, language and socio-emotional skills. The PS also have a strong WASH component.

2.1 Programme Components

District on boarding & engagement: Ghana Education Service (“GES”) Districts are selected through a competitive process, and are given a series of on boarding activities. These include an orientation workshop for all district staff, negotiation meetings to agree a Memorandum of Understanding (MoU) setting out roles and responsibilities, an introduction workshop for head teachers and PTA representative from each school to invite them to participate in the programme, and a training workshop for key District staff.

Teacher training: Two KG teachers from each school receive a five-day training course. This covers the importance of education and play, classroom management, how to use and make games, and how to train Mothers. The Head teacher and PTA representative attend two of these days. The training is facilitated by *Lively Minds* staff and supported by GES officials.

Training of Volunteer Mothers (VMs): The trained KG Teachers then train 30-40 VMs in their community using a scripted curriculum. To maintain quality, teachers are supervised and supported by high performing KG Teachers from schools with existing PS. District officials and *Lively Minds* staff also monitor some of the PS activities. Training includes two community meetings and eight participatory workshops, each lasting two hours. It is designed for women who are illiterate and have never been to school. Content includes the importance of education and play, how to make and play games with children of KG age, child-friendly teaching. In addition, VMs are taught how to install simple hand washing devices (tippy-taps) at home.

Play Schemes (PS): VMs are divided into four groups, and each group is given a different day where they teach at the KG for one hour. 25 KG children are arranged in small groups (maximum of 5) and rotate around the following 5 in-door play stations: matching/sorting; numeracy; sizes, colours, senses; books; building. One VM runs each play station and they teach using discovery-based teaching methods, rather than rote learning which is the norm in formal settings in Ghana. The remaining KG children play outdoor games, led by VMs. The KG teachers supervise the sessions. Children have to hand wash with soap before using the PS, sensitising them to this vital WASH practice.

On-going support: GES officials and *Lively Minds* staff have monthly meetings to track progress of the PS and identify corrective measures. GES officials monitor the PS as part of their normal supervisory functions and *Lively Minds* conducts some unannounced monitoring visits. Once a month “top-up” training workshops are held for KG Teachers where they discuss

problems, share successes and are trained to provide the VMs with a monthly Parenting Workshop.

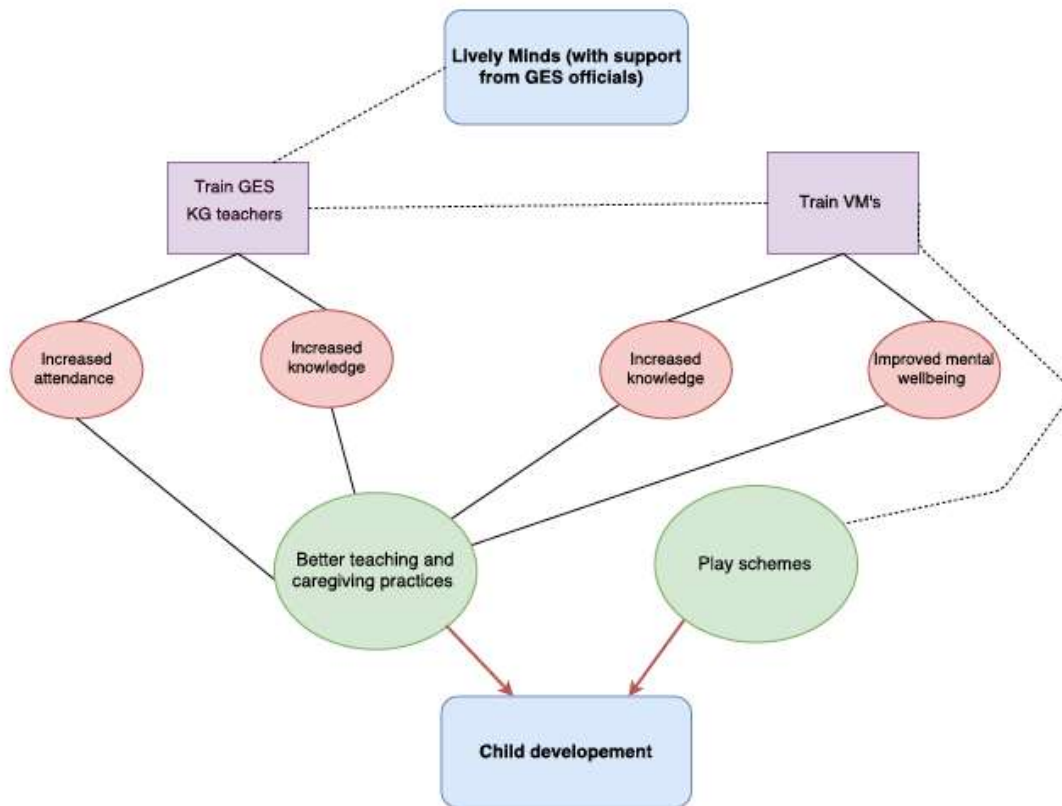
Parenting workshops: VMs are given monthly parenting workshops by the KG Teachers to increase their awareness on a variety of childcare issues, reinforce new behaviours, and to incentivise the VMs to keep them committed to volunteering. Topics include nutrition, hygiene, child rights, play, communication, malaria prevention, financial awareness, self-esteem and inclusive education. Over time, PS Committees are established in each school to manage their PS, made up of 4 VMs, a PTA representative and KG Teachers and Head Teacher.

2.2 Programme Theory of Change

An overview of the intervention's theory of change is outlined in Figure 1. Our central hypothesis is that the intervention can address three major constraints to improving children's developmental outcomes in this setting: (1) low teacher training, knowledge and motivation; (2) a lack of play based learning as opposed to rote learning; and (3) risk of mental health problems for VMs as well as lack of information on early childhood development and practice. We hypothesise that the intervention will address these constraints through three key channels; (i) Teachers will be equipped with greater knowledge of child development, which should improve the quality of their general teaching outside of the PS. In addition, being held more accountable for their attendance should incentivise teachers to come to work, and the presence of VMs should make large class sizes more manageable for them; (ii) Children will benefit directly from the PS through exposure to play based learning, which is expected to improve their cognitive, motor, and socio-emotional skills; and (iii) VMs will have better knowledge of best practices for their child's development, and improved self-esteem from being valued as part of the programme. This will motivate them to invest more effectively in their child.

This theory of change encapsulates the direct effects of the intervention. We hypothesise however that there could be a range of additional spillover effects on people not directly involved in the programme. Firstly, although only VMs directly participate in the intervention, there could be beneficial impacts upon other women in the community. These could set in through interaction with VMs and teachers, or with their children who are in the PS. Moreover there could be benefits for siblings of children attending a PS even if they do not attend themselves. This could happen through interaction with the TC as well as improvement in parenting practices by the mothers (particularly VMs) in ways that benefit all children in the household and not just the TC.

Figure 1. Theory of Change



3 Research Design

3.1 Evaluation Design

To evaluate the impacts of the programme on children and their families, we conducted a randomised controlled trial (RCT). We randomised 80 schools in 2 districts of Ghana – Bongo and Tolon – into two equally sized control and treatment groups. These include 38 schools in the Bongo District (Upper East Region), and 42 Schools in the Tolon District (Northern Region). Among the 40 schools allocated to the treatment group, 21 schools were located in Tolon and 19 schools in Bongo. Despite neighbouring each other, these two regions differ in a number of aspects, including language, religion, culture and schooling quality.

The randomisation was conducted using two levels of stratification: circuit and school size. A circuit is a geographical cluster of around 10 schools that falls under one supervisor from the Ghana Education Service (GES). This level of stratification was chosen in order to ensure strong geographical balance between treatment and control groups. The two strata used for the second level were (i) “small schools” defined as having below the median total number of KG children within the strata, (ii) “large schools” defined as having above the median number of total KG children.¹ School size is likely to be highly relevant for the efficacy of the intervention; hence this stratification was used to ensure strong balance on this variable. In total, we obtained 14 strata.

Power calculations

Power calculations were originally done at the project design stage, to inform the sample size that would be required for the study. These were done under conservative assumptions, which suggested that with 30 individuals per community, 80 communities (40 treatment and 40 control) and power to reject a wrong hypothesis with 80% probability, we had a minimum detectable effect size on the primary outcome (child test score) of between 19% and 30% of a standard deviation. Given that we now have actual data available, these power calculations have been updated using more accurate information. The new calculations, done by simulation, show a minimal detectable effect size of between 11% and 12% of a standard deviation, on the primary outcome (IDELA school readiness score described below). This is well within the acceptable range given the measured impact of previous early childhood interventions.

3.2 Evaluation Implementation

Timeline

Recruitment and training of VMs took place in October 2017. PS started to run in treatment schools in November 2017. The academic cycle covered by this study runs from October 2017 to July 2018. The data collection for the evaluation started in July 2017 with a census of target communities. In August 2017 the baseline sample was selected and in September 2017–November 2017 the baseline data collection took place.² In March 2018 and between September

¹ On average across circuits, the median number of KG children was 44 in Bongo and 52 in Tolon.

² The original plan was for the baseline data to be completed before the start of the VM training sessions in October 2017, as the latter could possibly influence VM responses to the baseline survey. However, following unexpected delays to the census, we had to adjust the plans and prioritise completing data collection in treatment communities ahead of the start of *Lively Minds* activities. This meant that only control schools were surveyed after this point.

and beginning of November 2018 endline data collection took place. This is just before PS started to run also in control communities.

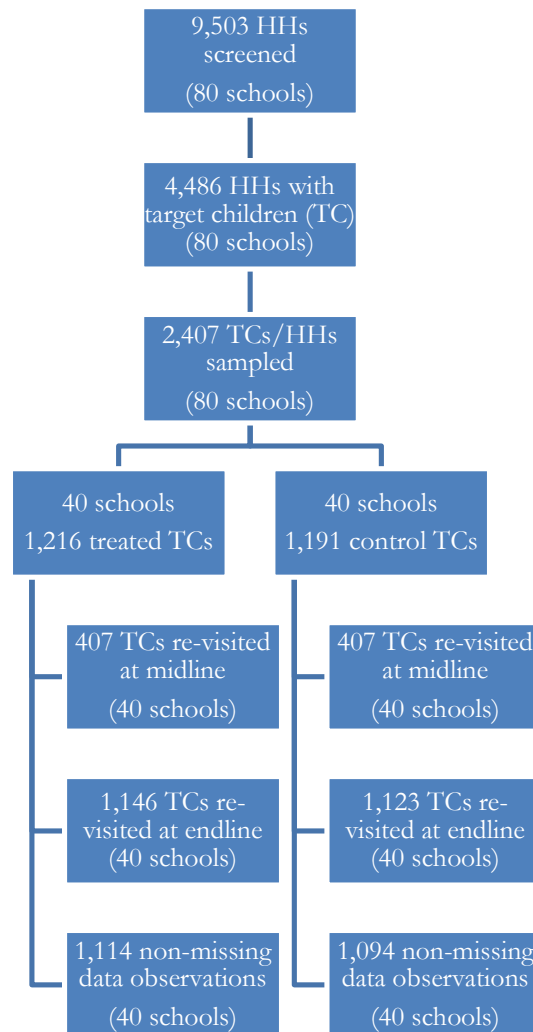
Sampling strategy and sample size

Figure 2 provides a summary of the sampling design and process. To frame the sample under study, we originally planned to conduct a census in 80 study communities, where we expected one school to serve the community and all children in the community to attend that one school. During preparations for census, however, we found that there were some schools that serve multiple communities and that children in a given community were spread out across multiple schools in the area. This, combined with the fact that the randomisation and intervention were conducted at the school level, meant that the census sampling needed to be conducted at the school level. In order to achieve a sample of TCs attending each school, households located closest to the school were enumerated first, gradually moving to those further away to reach 150 households. We detail the different approaches we used, depending on placement of a school within a community, in Appendix B.

Target households were those whose children of pre-school age were either attending school or intended to do so in September of 2017. Out of the 9503 surveys conducted, 6,446 of children living in 4,486 households were going to be attending school in the next academic period. From this set of households a random sample of 2407 TCs, their PCs and siblings (if any) was drawn for baseline survey. The PC was defined as the person that spends most of the time (more than half of the time) with the child when the child is not in school and who accompanies the child to the KG in case the child attends KG. If there was someone who accompanies the child to the school other than the one who spends the most time with the child in the household, then the one who spends the most time with the child was selected as the PC of the child. Baseline data show that 78% of PCs are the biological or adoption mothers of the TC, 1.10% are the biological fathers and the remaining 20% are other relatives and non-relatives. 86% of PCs have been PC of the child since the child's birth. Only one TC per household was selected. Even though the original plan was to select 30 children per community, in some cases the available number was as low as 11. In those cases all the TCs in that community were selected and the remaining number of children to be sampled was equally split among the rest of communities and randomly selected.

During the first two weeks of baseline data collection, an equal number of treatment and control schools were surveyed to make sure that initial interviewer learning effects are balanced across treatment and control schools. After that (from 1st October onwards) treatment schools were prioritised with the remaining control schools surveyed once treatment schools had been completed (and *Lively Minds* engagement commenced). Within the two treatment categories, schools were surveyed in a random order.

Figure 2. Sampling process



In addition to the child, PC and household sample, we also sampled KG teachers. The teacher sample was obtained by asking the circuit supervisors to invite all KG teachers that were working in the KG at the time of the survey to a central meeting. This same method was used both at baseline and at endline. If a baseline KG teacher was no longer working as a KG teacher at endline – for example, because of being transferred to another school, because of quitting or because of no longer being a teacher at the KG level - then that teacher was not captured in the endline sample, and vice versa.

As a result, our endline sample of KG teachers differs from our baseline sample. Specifically, 81 out of 160 KG teachers in our endline sample were new teachers for which we do not have baseline data available. Nearly all (97%) of the new KG teachers reported at endline to have been teaching in the KG school in the last term, suggesting that at least to some extent they participated in the *Lively Minds* activities. However, asked how long they had been teaching in the current KG school, 25% of the new teachers reported to have been teaching in the current KG

only for 4 months or less and 50% for 10 months or less. A significant sub-set of teachers in our endline sample therefore likely did not benefit from LM's initial training and/or did not participate in *Lively Minds*' programme for the full academic year. Our Intent-to-Treat (ITT) estimates of teacher impacts presented in Section 8.1 should therefore be considered lower bounds.

Attrition

Table 1. Attrition by treatment status

	Treatment	Control
Baseline	1216	1191
Endline	1146	1123
Attrition (number)	70	68
Attrition (%)	5.7%	5.8%
Reasons		
Death	6%	3%
Migration	16%	19%
Temporal migration	0	18%
Unknown	79%	60%

Out of the 2407 TCs sampled at baseline we were able to track 2269 children at endline. As shown in Table 1, this attrition rate of 6% one year after the intervention started does not differ between treatment and control groups or between districts. The main reasons for attrition include household reallocation (in more than half of the cases) and death of the TC. Furthermore, we were able to collect data on child cognitive and other development outcomes using the full battery of tests (SDQ, IDELA and Spelke tasks, see Section 3.3 for a discussion) only for 2208 of these TCs. For 2 out of those 2208 observations there was missing information for crucial regression controls (such as age). This yields a final sample for analysis of 2206 TCs and PCs (1114 in the treatment group and 1094 in the control group). We have complete PC data for all of these 2,206 TCs.

Additional Volunteer Mother sample

To look at mechanisms, it is important to identify households in our sample with PCs who ended up becoming VMs in the *Lively Minds* programme. For obvious reasons, this could only be done after the start of *Lively Minds* activities. To that end, the list of names of women that signed up to be VMs were matched with our sample through follow-up visits to the communities and manually matching lists of names (in control schools this was done one year later than the treated schools). Out of the 1,480 VMs in our study area, 437 turned out to be PCs that were captured in our sample.

The number of 437 sampled VMs was substantially smaller than the number on which our initial power calculations were based, which was 10 VMs per school or 800 in total. To improve statistical power, we therefore sampled at endline an additional 390 VMs from the list of mothers who had been matched to the census survey but whose households either (i) had not been sampled for baseline survey or (ii) whose households had been sampled for baseline survey but

who had not been surveyed as PCs. For this additional sample of VMs, a shortened version of the survey was conducted at endline, focusing on a core set of VM outcomes (see Section 3.3). Note that no baseline data are available for this additional VM sample nor do we have any child development outcome (baseline or endline) data available of their children.

3.3 Outcome measures

In this subsection, we detail the various assessments that we administered on children, PCs and teachers to measure the primary and secondary outcomes specified in the study protocol (available in Appendix A).

Primary outcomes

The primary outcomes of interest in this study are a set of five developmental outcomes for the TC, which constitute critical components of school readiness: **Emergent numeracy** (i.e. child's ability to do basic math which at the age of 4-5 is measured for example by recognition of numbers, shapes and sizes); **emergent literacy** (i.e. child's ability to read or to recognise numbers), **fine motor skills** (i.e. skills that use the small muscles, for example picking up small objects or drawing), **executive functioning** (i.e. skills involving mental control and self-regulation, for example ability to memorise, or plan and control impulses), and **socio-emotional skills** (i.e. child's expression and management of his or her own emotions, her ability to empathise with others and to establish positive rewarding relationships with them, and the child conduct and general behaviour when interacting with others).

When selecting measurement tools for these primary 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 etc. 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.

On the basis of the first two criteria, we selected an instrument developed by *Save the Children*, the IDELA (Pisani et al., 2015). This instrument was purposefully designed to be internationally applicable and comparable and has already been used and validated in various contexts, including in Ghana.

To measure children's socio-emotional development, we use a measure of maternal report, the Strengths and Difficulties Questionnaire or SDQ (Goodman, 2001). It comprises scales measuring the five following domains: conduct problems, hyperactivity, emotional problems, peer problems, and pro-sociality. The conduct problems and hyperactivity scales together make a measure of externalising problems, while the emotional problems and peer problems scales make a measure of internalising problems. The SDQ has also been validated in various contexts around the world, including in Ghana (Hoosen et al., 2018).

Based on our third consideration around the comprehensiveness of our child development assessment and the risk of “teaching to the test”, we complemented the IDELA tool with a battery of assessments developed for and used in the Ghanaian context by Harvard Cognitive Psychologist Professor Elizabeth Spelke (Spelke lab). With the exception of motor skills, the Spelke tasks measure similar developmental domains as the IDELA.

Finally, we also assess the impact of the intervention on child’s health, by using data on incidence of diarrhoea, fever and respiratory infections using the definitions of the WHO as measures of morbidity. We also measure middle-upper arm circumference (MUAC) to assess child’s malnutrition more objectively. MUAC is generally perceived to be a better indicator of mortality risk associated with malnutrition than, for example, weight-for-height (Briend et al, 1989; Vella et al, 1994; Briend et al, 2012).

Secondary outcomes

In addition to TC outcomes, we also estimate the effects of the intervention on PCs and siblings of children exposed to the programme. Although the study was not designed to evaluate this rigorously, we also assess changes in outcomes of teachers in treatment schools relative to those in control schools. The outcome measures used in each case are as follows:

Primary caregivers: We collect information on PC’s knowledge of ECCE and her beliefs regarding the importance of this for children’s development. To test knowledge, we rely on a selection of items from the Knowledge of Infant Development (KIDI). We also collect information on PC’s perception of what features are most important for the quality of a KG. We measure psychological outcomes of PCs through the use of two well-established scales; the SRQ-20 measure of depression (World Health Organisation, 1994), and the Rosenberg measure of self-esteem (Rosenberg, 1965). To understand parenting styles we also include a module for parental practices (e.g., dealing with bad behaviour, encouraging behaviours, games involvement and beliefs) not only as reported by PCs but also reported by interviewers after observing the PC and the child playing a game together.³

Home environment: The presence of toys and learning materials in the house are assessed together with parental involvement with the child, the child’s routines and organisation of the child’s time inside and outside the family house. This is assessed using the Family Care Indicators (FCI) developed and tested by UNICEF (Kariger et al., 2012).

Younger siblings: The main outcome measure for younger siblings is the Caregiver Reported Early Development Index, CREDI in its short form (McCoy et al., 2017). The CREDI is designed to serve as a population-level measure of early childhood development for children from birth to age three. It exclusively relies on caregiver reports and thus primarily focuses on milestones and behaviours that are easy for caregivers to understand, observe, and describe.

³ The scale was developed with the advice of psychologist Marc Bornstein based on existing measures. For instance, indicators for dealing with bad behaviour are based on the child discipline module of the Multiple Indicator Cluster Survey (MICS).

Older siblings: We measure literacy and numeracy 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. At endline we complement these literacy and numeracy tests with a set of tasks developed in the Spelke lab. These tasks are aimed to assess math, literacy, executive function and socio-emotional skills.

Teachers: We assess changes in outcomes of *teachers*, using an instrument developed by Aber, Berman and Wolf (2017) for use in Ghana to measure the effectiveness of the QP4G pre-school programme. This includes a variety of measures including teacher practices, burnout, and job satisfaction amongst other things. We also used the SRQ-20 to assess teachers' mental wellbeing.

In addition to these core sets of primary and secondary outcomes, additional contextual data was collected on children, PCs, households, teachers, schools and communities. Table 2 gives an overview of all survey instruments and their contents used in this study.

Table 2. The different instruments and their respondents

Survey	Main respondent	Contents	Rounds administered
Household (HH) Survey	The head of the household that the TC lives in. If the household head was not available, another knowledgeable household member was interviewed. The preferred order of respondents in absence of the head was: 1. The PC of the TC 2. The spouse of the head 3. Other knowledgeable Adult household members ($>16y/o$).	Gender, age and literacy of all household's members; household assets, dwelling, land and livestock ownership, income, consumption and expenditures, savings and credit, shocks, among others.	Baseline and endline Location: the compound/house where the household lives.
Poverty Probability Index (PPI) (part of HH Survey)	The head of the household that the TC lives in.	A Poverty Probability Index (PPI) developed by IPA was administered to all households at endline.	Endline Location: the compound/house where the household lives.
Primary Caregiver (PC) Survey	PC of the TC	PC demographic characteristics, health, maternal wellbeing, time use, cognitive skills, empowerment and school engagement.	Baseline and endline Location: the compound/house where the household lives.
Parenting styles (part of PC Survey)	PC of the TC and observed by the interviewer	Parenting practices such as how they deal with bad behaviour and how to praise good behaviour	Endline Location: home
IDELA	TC (between 4 and 6 years old)	School readiness assessment. It assesses emergent math, literacy, executive function, fine motor and socio-emotional skills.	Baseline and endline Location at baseline: home Location at endline: School (occasionally

			at home in cases where it was not possible to move the TC to the school).
TC Task Assessment	TC (between 4 and 6 years old)	Spelke's tasks to assess math, literacy and socio-emotional skills.	Endline Location: school (occasionally at home in cases where it was not possible to move the TC to the school).
Older Sibling (OS) cognitive assessment	The youngest Older Sibling of the TC (5-10 years old)	Raven's matrices to assess cognitive ability of the OS	Baseline Location: the compound/house where household lives
Older Sibling (OS) Assessment	The youngest Older Sibling of the TC (5-10 years old)	Literacy and numeracy assessment.	Baseline and endline Location: the compound/house where household lives
Older Sibling (OS) tasks	The youngest Older Sibling of the TC (5-10 years old)	Spelke's tasks to assess math, literacy and socioemotional skills.	Endline Location: the compound/house where household lives
Younger Sibling (YS) Survey	PC of the oldest younger sibling of the TC (who shares the same PC with the TC).	CREDI: a maternal report that measures overall development of the child.	Baseline and endline Location: the compound/house where household lives
Community Survey	Assembly of knowledgeable persons in the community (up to 6 maximum)	This includes community' general characteristics, education and childcare services, health services, local economy and wages, local prices, water and sanitation, shocks, social protection and migration.	Baseline and endline Location: Community
Teacher Survey	Kindergarten Teachers	It includes teachers' demographic characteristics such as age, gender, religion, education, experience and working conditions, measures for burnout, teaching practices and wellbeing.	Baseline and endline Location at endline: Central Location. Tolon: Ghana Education Service (GES) Office Bongo: Bongo World Vision Centre (managed by district assembly)
School Observation Survey	Conducted by the Field Supervisors who observed the schools. If needed, additional questions were asked to the headmaster or another (KG)	Observation of basic amenities, state of physical building, learning, materials and environment, toilet and sanitation.	Baseline and endline Location: school

	teacher in the school.		
Additional Volunteer Mother (aVM) Survey	Mothers registered as Volunteer Mother (VM) in the PS, but who were not in our sample (as PC).	The compound/house where the VM lives	Endline Location: the compound/house where the VM lives

3.4 Data collection procedures

This section summarises the main procedures followed by the IPA-Ghana project team to collect the endline survey data for this study. Further information about the process can be found in Appendix C and a description of the baseline survey data collection procedures is provided in the baseline report of this study (Attanasio et al., 2018).

Organisation survey modules

A total number of 10 different survey instruments was administered at endline: Household (HH), Primary Caregiver (PC), additional Volunteer Mother (aVM), Young Sibling (YS), Older Sibling (OS), IDELA, Task assessment, Community, Teacher, School observation. The HH, PC, teacher and aVM surveys were conducted by IPA-trained enumerators who were divided over 12 field teams (6 teams in the Northern Region and 6 teams in the Upper East Region). These field teams each consisted of four enumerators and were supervised by a team leader so that there were 5 people in each team. To conduct the teacher surveys, we invited all teachers to a central location. The TC assessments (IDELA and Tasks), YS and OS surveys were carried out by a total number of 34 assessors, 18 and 16 in the Northern and Upper East region, respectively (36 enumerators were trained with the aim of having backups). The community survey was administered by the Team Leaders and the Field Supervisors completed the School Observation.

Piloting

Since most survey instruments were already used for baseline data collection and had hardly changed, they did not have to be piloted again. Piloting – between the 23rd and the 25th of July of 2018 - therefore focused on the Spelke Lab assessments and the YS and OS assessments that had undergone some changes since baseline. After piloting, feedback from the field was incorporated to finalise the instruments. These include changes in the way of administering some of the tasks assessments (i.e., using one surveyor rather than two), reducing the number of trials, modifying the content (e.g., numbers) of exercises to improve variability.

Field team recruitment and training

Field Supervisors, TLs and Surveyors were all recruited through an online job advert, which specified that candidates needed to have obtained at least the Higher National Diploma (NHD). Suitable candidates were invited to training, where surveyors were divided into two groups: one group that focused on child assessments of the TC, the TC's Older- and Younger Siblings (OS and YS) and another group that focused on the Household (HH), Primary Caregiver (PC), Teacher and Additional Volunteer Mother (aVM) surveys.

Different training sessions were organised for each group of surveyors. Each training session consisted of a classroom training session, focusing on familiarising with the survey instruments, followed by a field training session, which focused on practicing the instrument in a real field context.

Fieldwork

Endline data were collected in the months of September and November 2018. The HH/PC survey team started first, followed by the child assessment team. The child assessment team started later after the government schools had reopened. Government schools reopened on the 12th of September and we allowed for a week's lag before starting the TC assessments. At endline it was decided that TC assessments were done in schools, in contrast to the assessment at home during the baseline data collection. The average duration of each survey instrument during actual fieldwork is reported in Appendix C.

Quality assurance

Monitoring happened by the TL, FS, FM and RA. Monitoring happened either through accompaniments or spot checks. During an accompaniment, the person monitoring sat through the entire survey, whereas spot checks were shorter. For spot checks, the person monitoring walked in at a random point during the assessment and witnessed a short part of the questionnaire. The goal for each team leader was to complete at least 5 monitoring activities (a combination of spot checks and accompaniments) per surveyor per week. In practice, this number was not always reached. Reasons for this include that team leaders were often called out to solve problems elsewhere and that the monitoring forms were not always filled out (even when monitoring was done).

Each district also had two auditors who were engaged in audio auditing and backchecks alternately. These auditors were employed to ensure data quality. Backcheck auditors revisited a randomly selected 10% of the interviewed households to re-ask a certain number of questions from the questionnaire. The questions that were re-asked are questions of which the answers are not expected to change much, such as the construction material of the house. The data collected by the backcheck auditor was then compared to the original survey data to see if there were discrepancies. In the case of non-matching answers, the underlying reason for this was investigated.

At the beginning of each interview, the respondent was asked for permission to record portions of the interview. The large majority (97% in the HH Survey and 95% in the PC Survey) of respondents consented to the recording. The audio auditor listened to these recordings and checked if all procedures were followed and whether questions were asked correctly. A small report was written for every audio file. In case of a suspicious recording, the report was shared with the FS, who then further investigated the matter and made recommendations.

Finally, daily high-frequency checks were run on the HH, PC, IDELA and TC Task Assessment instruments to maintain data quality and rectify errors when data came in.

Fieldwork challenges

Tracking Primary Caregivers. Enumerators initially faced problems tracking PCs. This was a result of the fact that our survey began during the farming season. Most of the PCs were usually engaged in farm activities when enumerators visited. Other reasons include permanent relocation of some households and temporary rural-urban migration. This problem occurred in both districts but was larger in the Upper East region. Since there is only one farming season in the northern part of Ghana and alternative livelihood opportunities are rare, migration to the south is very common. Especially for the PC survey, this significantly affected the number of respondents in our sample.

Tracking Target Children. Some of the TCs in our sample were not enrolled in school yet/anymore and child assessors had to trace them to their homes to bring them to the school compound to assess them. This was not always feasible. Moreover, some of the TCs ended up attending different schools, both within and outside the community, than those they had reported they would attend during the census (and which we therefore sampled). Child assessors travelled to these schools to bring these TCs to our sample schools to be assessed. To mitigate this problem, we recruited child mobilisers to mobilise TCs so that child assessors could continue with their normal duties. These child mobilisers were enumerators who had been part of the HH/PC team before. The introduction of child mobilisers was extremely useful, as it increased the productivity, as well as the percentage of children we were able to find and assess in each community.

Community accessibility. Most challenges we faced related to accessibility of communities. First of all, the large number of streams and rivers during the rainy season meant that some communities within the Bongo district were not accessible, especially on days of heavy downpour. These communities had to be revisited on different days. Second, some of the communities in the sample are very remote, leading to long travel times. Finally, transportation to the communities was a major problem in this study. About 90% of the Bongo district roads are feeder roads and rocky in nature. Due to the rough nature of the roads, motorbikes easily broke down and regular services were required.

Staff turnover. Halfway through fieldwork, a number of enumerators and child assessors dropped out because they had found a different job, wished to go back to school or found the surveying job too stressful. Even though some of these dropouts were replaced, it did to some extent affect survey productivity.

4 Study context

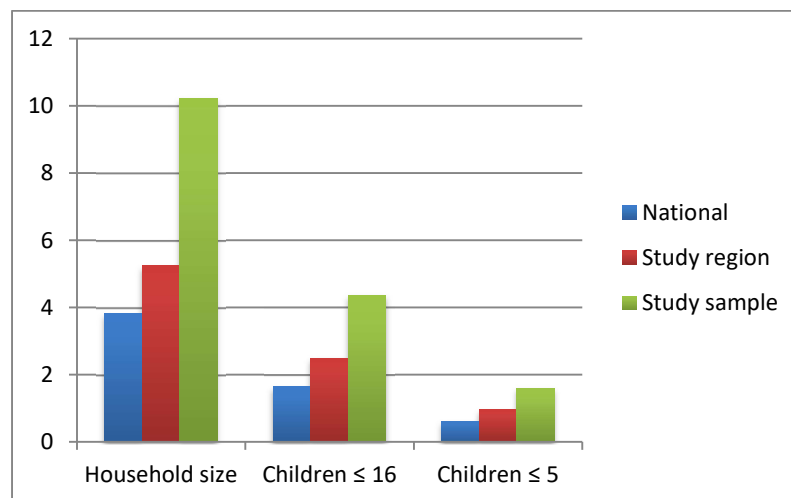
Before we discuss the main impact evaluation results, we describe in more detail our study area and the setting in which the *Lively Minds* intervention was introduced.⁴ This section discusses only a snapshot of the full list of descriptive baseline statistics provided in Appendix D.

4.1 Households' demographic and economic profile

We start off by describing our sample's demographic and economic profile and comparing it to the regional and national population. We draw statistics for the region or country either from the 2014 Ghana Demographic and Health Survey data (DHS)⁵ or from the 2017 Ghana Maternal Health Survey (GMHS).

Tables 1-2 in Appendix D show that our study communities have, on average, about 2000 inhabitants living in 300 households with an average of 10 members. This household size is large relative to the average Ghanaian household that has 4 members (GMHS) and large even relative to households in other parts of the same regions. Figure 3 highlights this specific demographic difference of our sample relative to the national and regional averages. This is (at least partially) a result of our sampling strategy targeting children aged 3-5, yielding a sample of households that have on average relatively more children (between 1 and 2 children aged 5 or below compared to the national average of less than one (GMHS)).

Figure 3. Sample household demographics relative to national and regional average



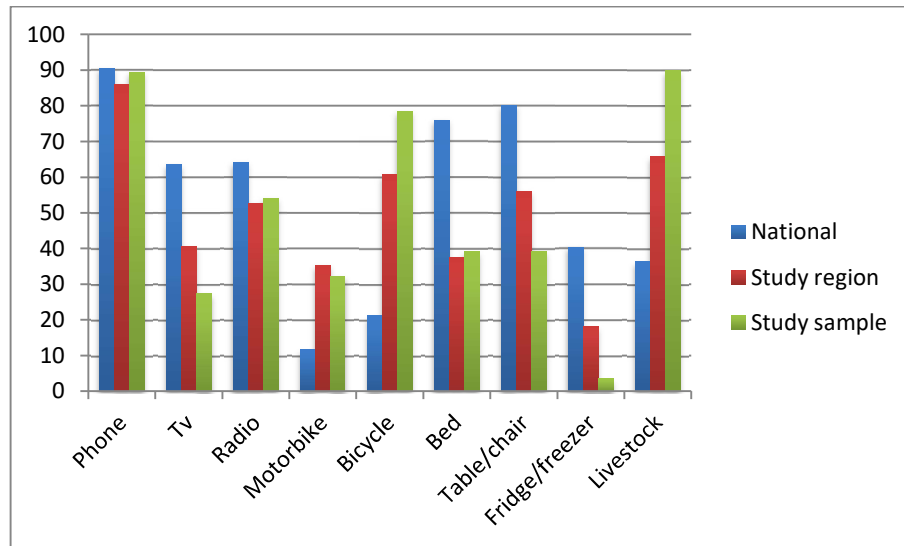
Another notable demographic difference is that, whereas most families in Ghana are Christian (DHS), slightly more than half of the households in our sample are Muslim.⁶

⁴ Our study sample was not designed to be representative of the two study districts, regions or Ghana as a whole, since it focuses on 80 schools belonging to almost as many communities located in districts Tolon (Northern Region) and Bongo (Upper East Region). The sample is more representative of the population typically targeted by *Lively Minds*.

⁵ The figures reported for the 2014 DHS are obtained in most cases from the Ghana Demographic and Health Survey Report (GSS, 2015) and in few cases from accessing directly the data.

⁶ Note that this is not necessarily what is observed in the typical areas where *Lively Minds* operates.

Figure 4. Sample asset ownership relative to national and regional average (%)



In terms of their economic status, our study sample is significantly more rural and poorer than the average household in Ghana. Whereas 45% of Ghana’s population nationwide is reportedly engaged in agriculture (DHS), more than 6 in 10 households in our study sample report to receive their main income either from the land used for crops or from a salaried job in that sector (Table 4 in Appendix D). As highlighted in Figure 4, most households in our sample own a mobile phone, as is the case in Ghana nationwide. However, the proportion of households in our sample that own other types of assets is below what is observed at the national level and in other parts of the same regions. On the other hand, in contrast to the average Ghanaian household, a larger proportion of households in the study sample own a bicycle or a motorbike, just like other households in the study regions. Similarly, livestock ownership is particularly high in the study sample relative to the rest of the country and other parts of the study regions.

The lack of non-agricultural related asset ownership in our study sample makes the households particularly vulnerable to unexpected natural shocks affecting their daily activities. As shown in Table 4 in Appendix D, among the main shocks reported by households in our sample, since the PC’s pregnancy with TC, are droughts, the death of livestock and crop failure.

Finally, Table 1 in Appendix D informs us that households in our study sample live in communities with generally poor infrastructure. One in three communities does not have access to electricity and more than half do not have access to paved roads making them completely inaccessible for 1 to 2 months every year during the rainy season. Overall, it takes households on average more than 1 hour one way to go to the country capital using their most commonly used mode of transport (which is on foot for 30% of households).

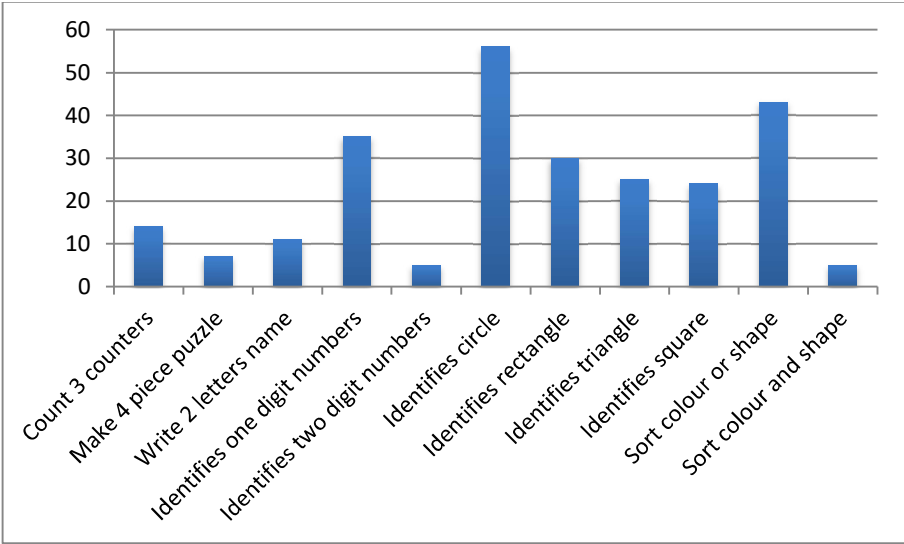
4.2 Child development

The development of children in our sample is low. At baseline (just as at endline), we administered the IDELA tool, which includes several tasks aimed at assessing children’s ability in different domains, such as emergent literacy and numeracy, motor skills, and socio-emotional development. Sample average baseline IDELA results are presented in the first column (‘All’) of

Tables 5-6 in Appendix D. The same format is used in subsequent tables. Below in Section 4.6 we compare IDELA and other results by whether or not the PC was a VM (Columns 3-6).

On individual items, Figure 5 shows that less than one in five children could count out 3 counters and about one in ten could complete a simple puzzle or write at least two letters when asked to write their name. Number identification is also poor, with less than 4 in 10 identifying at least one 1-digit number and almost none of them could identify at least one 2-digit number, on average. Less than one third of the 3-5 year old children could identify a shape such as a rectangle or a triangle and even less than that were able to sort figures based on both colour and shape.⁷

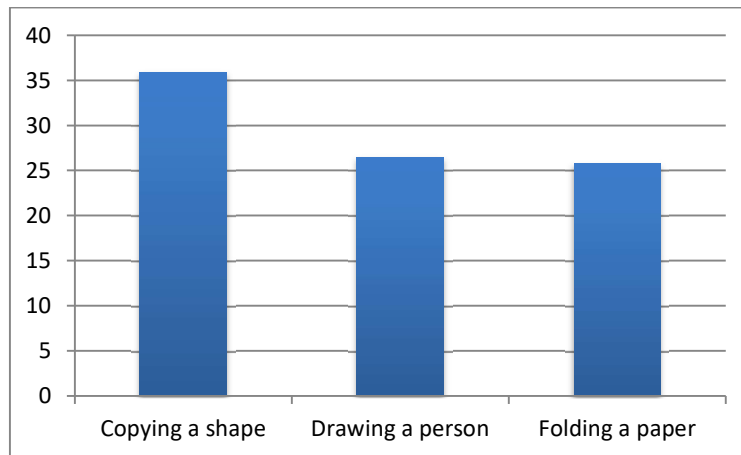
Figure 5. Baseline cognition Target Children (%)



As shown in Figure 6, fine motor skills, important for writing development, are also poor. Less than a third of the 3-5 year old sample was able to copy a particular shape, to draw a person, or to fold a paper in a specific way.

⁷ We also observe significant differences between districts. Total test scores for both target children and their older siblings were over 25% higher in Bongo compared to Tolon. The data provide some potential reasons why these gaps could exist. Despite households being richer on observed measures in Tolon, schools in Bongo have more educational resources, teachers work more hours, and parents are more highly educated. In addition there has been a wider proliferation of NGOs and religious organisations with a focus on education in Bongo.

Figure 6. Baseline fine motor skills Target Child (%)



The results of the socio-emotional component in the IDELA tool presented in Table 6 in Appendix D reveal that children correctly scored on average 6 in 10 questions around self-awareness (e.g., knowledge of own name, age, gender), around 30% were able to solve conflicts when playing with another child, 15% showed empathy towards others feelings and very few were able to identify what makes them feel sad/happy and deal with those feelings (emotional awareness).

Basic cognitive learning outcomes, designed from the school curriculum, were also assessed for older siblings of the TC, aged 6 to 10. This revealed relatively low rates of literacy and numeracy, presented in Figure 7 and Figure 8. Less than half were able to do simple addition such as $46 + 31$ and only one in four was able to solve a simple addition problem such as adding the price of a 7 GHS hat and a 4 GHS drum. In terms of literacy, only 4 in 10 was able to read the letter 'o' and less than one in six was able to read a simple word in English such as 'dog'.

Figure 7. Baseline math skills siblings 6-10 years old (%)

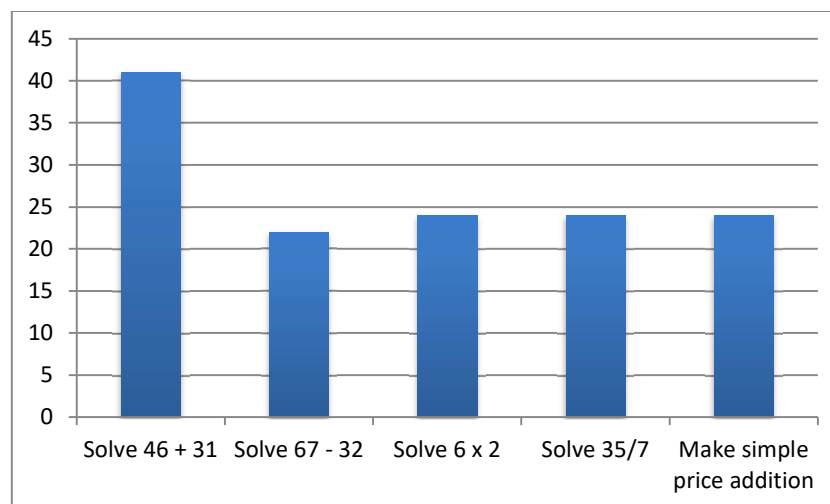
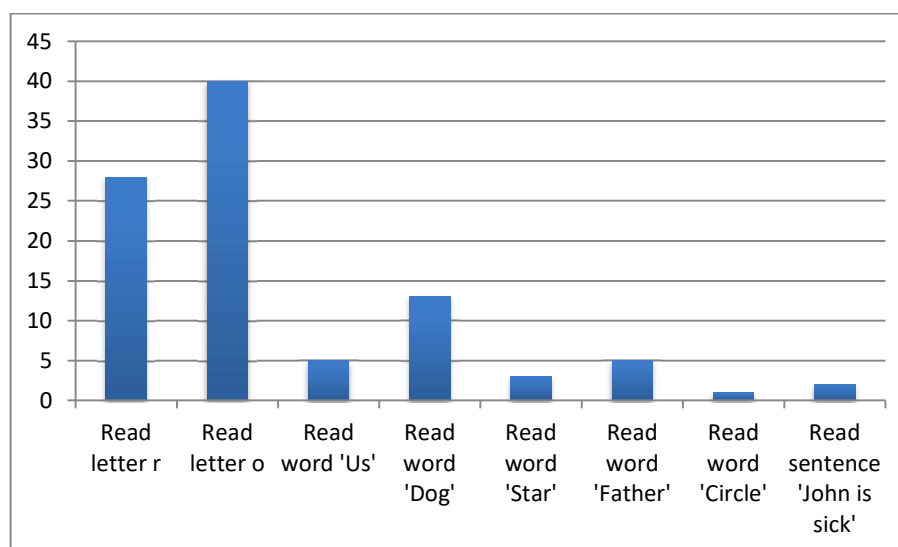


Figure 8. Baseline literacy skills siblings 6-10 years old (%)



The IDELA has been used in various other studies throughout the developing world. We used the IDELA data available on the Save the Children website to compare the performance of target children in our sample on the IDELA tools to that of children sampled in these studies. To do so, we constructed an aggregate score measuring the average percentage of correct answers on the test for children in each of the samples. This score can be broadly viewed as a measure of school readiness as it reflects ability in a variety of domains that are important to be school ready. On average, children in our sample rank 7th from the bottom (out of 43 samples), just above a sample collected in Lebanon and below a sample collected in Vietnam. Interestingly, it ranks much worse than a sample of children living in the urban areas of Accra (the full ranking is reported in Figure 1 in Appendix D).

4.3 Hand washing practice and health

WASH, and hand washing practice in particular, is another key focus of the *Lively Minds* intervention. The monthly parenting workshops put much emphasis on this and at the start of each daily PS children are required to wash their hands. The baseline data available confirm the lack of sufficient WASH awareness in Ghana, particularly amongst pre-school children. Even though the Ghanaian government has promoted hand washing (DHS, 2014) since the 2000s⁸, the practice is still not widespread in the country as the national statistics reveal. Only 40% of households in the country have water and soap in the designated place to wash their hands, another 20% only have water and 4% only have soap.

Knowledge on best hand washing practice is high among PCs in our sample however, as highlighted in Table 8 in Appendix D. At baseline, nearly all PCs were able to explain why hand

⁸ The Ghana Public-Private Partnership to Promote Hand washing with Soap (PPPHW) ran a campaign between 2003 and 2005. The Community Water and Sanitation Agency, run a campaign from 2001.

washing is important and could state a time when hand washing is needed. Knowledge among TCs was considerably lower (Table 9 in Appendix D). Only half of 3-5 year old children were able to state at least one good time to wash their hands, and only one in ten could explain why it is important. In terms of what is needed to wash their hands, less than half of the children reported that water is needed and a bit more than half stated the need of soap. As can be seen in Table 10 in Appendix D, related child health problems are also seen to be relatively prevalent in this population. For example, a third of TCs had experienced diarrhoea in the last 30 days, defined by 3 or more loose or watery stools. This could be linked to the fact that sanitation seems to be problematic in our study area, as shown in Table 1 in Appendix D. In our study sample, nearly 8 in ten households report to practice open defecation (OD). This is significantly larger than the OD prevalence reported in our study region (60%) or Ghana as a whole (15%) (GMHS).

4.4 Schools and teachers

The education system in Ghana provides two years of free and compulsory pre-school (KG) for those aged 4 and 5, with the aim of improving school readiness. Despite this, not all children end up attending: 68% of children aged 36–48 months attend preschool (McCoy et al., 2019). This figure is about 90% in urban areas like Accra (Wolf and McCoy, 2019). Among the sample of children aged 36 and 73 months in our study area, around 77% had ever attended pre-school before the intervention (Table 11 in Appendix D). The most common reason stated for non-attendance was the child being too young (59%), followed by the child not wanting to go to school (9%), too high school fees (5%) and too far distance to the school (4%).

Table 12 in Appendix D summarises the baseline profile of KG teachers in our sample and their baseline teaching practice. The majority of KG teachers in our sample are highly educated. Of the 151 teachers surveyed at baseline, the vast majority have a diploma/HND or higher and around three in four teachers have some kind of early childhood care and education (ECCE) training. Their education is high relative to KG teachers elsewhere in the country. Four in 10 KG teachers in Ghana have not received any formal training (SABRE, 2014) and a large proportion of them have only reached primary education (Wolf et al., 2017). This is possibly because *Lively Minds* only works with schools and teachers that are motivated enough to apply to participate in the *Lively Minds* programme and to sign a MoU with *Lively Minds* as part of the district onboarding & engagement activities (see Section 2.1). The flipside of this, however, is that teachers in our sample are young and relatively inexperienced, having taught only for four years on average.

Teachers in our sample work an average of 29 hours a week, with 22 hours spent in school and an additional 7 hours preparing for classes. In general teachers report being fairly highly motivated, with positive views of their colleagues, supervisors and the beneficial impact they are having on children. However, three quarters find their work monotonous and over a half believes that their work is not sufficiently challenging or stimulating (not reported in tables). This could at least partially explain the fact that KG teacher's turnover is a major problem in Ghana (Osei, 2006; Wolf, 2019). In the peri-urban region of Accra, for example, Wolf et al. (2019) report a midyear turnover rate of 40% in private schools and 16% in public schools. This is also revealed in our sample with teachers being in the actual KG only for about two years on average (Table 12 in Appendix D). One in five lives in the community where they teach but only one in

ten was born in that community. Half of the teachers report to have moved to the community for their KG job.

At baseline, teachers in our sample were asked about issues potentially preventing the delivery of high-quality education in their pre-schools. As summarised in Table 12 in Appendix D, the three most commonly cited issues were lack of financial resources, low parental involvement (either not actively involved or not sufficiently interested) and large class sizes. These answers are not surprising. KG Classrooms in Ghana are generally overcrowded with about 55 pupils per class (SABRE, 2014). A similar figure was observed in our survey sample (58), as shown in Table 13 in Appendix D. School infrastructure is largely inadequate. National statistics reveal that desks are available only for half of children in the KGs of the country. In our sample, there is one desk and chair for every 8 pupils and, in almost half of the schools, major safety hazards exists such as an unprotected hot stove or fireplace, tools not meant for children's use accessible, sharp or dangerous objects present, easy access to road, play equipment not well maintained posing a threat of injury, loose electric cords, play areas in front of doors, medicines, cleaning materials and other substances not locked away.

4.5 Parents and the home environment

The main innovative feature of the *Lively Minds* model is the fact that it trains and actively involves mothers in pre-school activities. Baseline statistics presented in Table 14 in Appendix D reveal that parental involvement in their child's education is otherwise indeed fairly low in our study area; as discussed above, this is consistent with reports from teachers that parents are not sufficiently interested in their child's education. Less than half of PCs of children attending KG report to know the KG teacher's name and to have visited the school in the last month. Only 15% of caregivers are member of a Parent Teacher Association (PTA) executive, i.e., chairman, treasurer, secretary, etc., who calls for PTA meetings and takes decisions on important matters related to school quality and infrastructure with inputs from the parents. Despite this, in the household survey the household heads report generally being quite satisfied with the quality of the KG school, with almost 80% of heads rating the school as good or excellent (Table 15 in Appendix D).

Lack of parental school involvement may at least partially be attributed to low educational levels of the parents themselves. Most of today's KG children in Ghana are first-generation students. One in five adults has never attended any form of education and amongst those who have, only a quarter reached junior secondary school and only one in ten continued with higher education (Table 2 in Appendix D). Our study sample exhibits even lower levels of education than the national average, with 8 in 10 adults reporting to have never attended school. It is therefore not surprising that almost all PCs in our study sample are illiterate (92%) with low cognitive ability as revealed by the score obtained in the Ravens' test with an average rate of correct answers of 43%. As a comparison, in a meta-analysis of 798 samples from 45 countries, Brouwers et al. (2009) found that Ghanaians scored on average 49% across a number of studies (including urban Ghana) while people in the UK scored on average 62.

As discussed in Section 2.2, an important channel through which *Lively Minds* expects to achieve impacts on child development is through improving maternal mental wellbeing. In the self-reported SRQ-20 questionnaire developed by the World Health Organisation to screen

depression and anxiety disorders, PCs are asked 20 questions, such as “is your appetite poor?” and “Do you find it difficult to enjoy your daily activities?”. Each of the 20 questions is scored 1 if the symptom was present during the past month and 0 otherwise (WHO, 1994). While there are no validated cut-offs to diagnose depression in the context of our study, a score of 8 or above is generally considered as an indication of depression. The average score amongst primary carers in our sample is just above 8 (Table 16 in Appendix D).

Another key feature of the *Lively Minds* model is the training of mothers and teachers in how to make low-cost toys and how to improve learning through play, as the home learning environment of children in our sample is very poor. As shown in Table 17 in Appendix D, over a third of households report having no household objects suitable for children to play with and households have on average only one bought toy for children to play with. Children’s books with illustrations or to paint are scarce to non-existent. When counting all the different types of homemade and bought play materials available at home (e.g., toys to play, toys to learn shapes, building blocks, vocabulary picture charts, book), households report having on average between 1 and 2.

This lack of play material in the dwelling could at least partially be explained by the fact that only 12.5% of communities have a shop nearby where toys can be purchased (Table 1 in Appendix D). Even putting material constraints aside, very few households (13%) report having conducted any form of play activities with the TC in the last 3 days. These figures are even lower for the younger children in our sample (siblings below the age of the 3-5 year old TC), with only 10% of the families conducting any play activity with them in the last three days. This is despite the fact that on average, PCs agree with most statements that reinforce the importance that parenting, maternal literacy, home support, games and parental engagement have for child’s development (Table 18 in Appendix D).

Table 19 in Appendix D presents statistics describing parental disciplinary behaviour. Note that since we did not collect such information at baseline, we use endline data and restrict the sample to the PCs in the control group to capture the trend in the absence of the LM intervention. Almost half of PCs report shouting to the child when dealing with bad behaviour whereas 4 in 10 report to explain to the child why the behaviour was wrong. One in 4 also reports taking away child’s privileges or forbidding something they like. In terms of physical punishment, 1 in 4 reports to hit the child on the bottom either with something or with bare hand. However, the interviewer observed the PC shouting to the child during interview only for 1 in 4 cases and the observation of physical punishment was very rare. To encourage good behaviours, 6 in 10 report to praise the child with words, 1 in 4 report to give sweets and 1 in 3 report to use physical affection.

The community role of teachers is considered of high importance among PCs (Table 20 in Appendix D). When thinking about quality of education, teacher’s experience or instruction is reported as the most important feature of a KG. Accordingly, PCs consider it as very important that KG teachers know children’s needs, encourage child’s learning of math and literacy, provide materials for play and measure children development. PCs also find teacher’s role also very important in terms of improving child’s interaction with others, child’s behaviour and emotions.

In contrast, PCs assign relatively less weight to teachers' role to connect families to community resources or to care about the whole family.

4.6 Volunteer mothers and their children

Not every mother decides to become a VM in the *Lively Minds* intervention. On average in each community, around 40 women typically put themselves forward as VMs. It is interesting to compare the characteristics of those who do and those who do not, and in this section, we compare baseline characteristics of VMs and non-VMs themselves as well as those of their children, restricting the sample to VMs that were found to be PCs in our sample.

Table 10 in Appendix D shows that children of VMs and those of non-VMs are similar in age and in gender. However, despite this demographic similarity, children of VMs were significantly more likely to go to school already prior to the launch of the VM scheme (Table 11 in Appendix D). Whereas 77% of children of non-VM caregivers were attending school at baseline, 84% of VM children were already going to school. Interestingly, as can be seen in Table 21 in Appendix D, VM children were also significantly more likely to have an older sibling than non-VM children. This could at least partially explain why their main caregivers were more likely to become VMs, if having older children already in school make caregivers relatively more familiar with the schooling system.

We do not observe any consistent differences in terms of children's baseline health and socio-emotional development characteristics (Table 7 and Table 10 in Appendix D). However, VM children score significantly better on some IDELA tasks related to fine motor development and executive function and they are also significantly more likely to list good times for hand washing. These patterns could potentially be explained by the fact that VM children on average have more schooling than non-VM children (see result discussed above), and/or by the fact that they grow up in different home environments.

That being said, we do not observe striking differences in the socio-demographic profile of VMs versus that of non-VMs (Table 16 in Appendix D). VMs and non-VMs are about 37 years old, about half of them are Muslim and most of them have not benefited from any form of education. VMs are however more likely to be the biological mother of the TC as well as their PC since birth, and they allocate more time to farm work and less time to paid work and to play with children than those who do not volunteer.

Interestingly, however, prior to the programme VMs were already more involved in school activities than non-VM mothers (Table 14 in Appendix D). In fact, more than half of them already knew their child's teacher's name, compared to less than 4 in ten of the non-volunteering PCs; and they were 6-7 percentage points more likely to have visited the school in the last month or to have attended a parent teacher association meeting. This could be explained by the fact that they were more likely to have a child older than the TC (see results discussed above). Nevertheless, they have similar beliefs about child development (Table 18 in Appendix D) and a similarly low number of play materials available at home (Table 17 in Appendix x).

5 Programme Implementation

5.1 Compliance

The PS were designed to run 4 times per week in each school during school term with different children and mothers attending each session. Each child would participate in indoor games at least once per week and in outdoor play 2 or 3 times per week. Over the course of the academic year, the PS ran over three terms, with the first term being 1st of November 2017 - 10th of December 2017 (6 weeks). The second term started the 15th of January 2018 and ran until the 15th of January 2018 (13 weeks) and, the last term went from the 14th of May 2018 and finished on the 24th of July 2018 (11 weeks). If the PS had run at its maximum capacity, then it would have had an intensity of 30 weeks in total, 30 hours per child during the academic year.

We have two sources of information to assess the level of compliance: process data collected by either *Lively Minds* supervisors or GES officials during their random visits to schools on the one hand; and data collected by the research team through the primary caregiver and teacher's survey. The evidence we find from both types of data is broadly consistent and suggests that compliance was high. Below we briefly describe the findings from each type of data.

Process data on compliance

Over the course of the academic year, *Lively Minds* and GES shared supervision of the schools where the PS took place. The 40 treated schools received random visits from either *Lively Minds* supervisors or GES officials, with the latter taking care of 55% of all visits. Each school received on average 9-10 visits in total, with some having had only 3 visits and others 16 visits in the same period. Disproportionally more visits took place in October-November 2017 and February 2018, with approximately 40% of all observed visits happening during the first month of the intervention.

The PS were found running in 80% of all monitoring visits. The main reasons for the PS not running at a particular visit were bad weather, community funerals, other school/community event or low attendance of VMs, the latter being the most frequent reason recorded by the monitoring team. Excluding incidences of bad weather and funeral, the percentage of PS successfully observed was about 85%. However, when calculating the average compliance rate per school, the rate is slightly lower: 82% when excluding bad weather and funerals and 78% when including those. Using the latter rates as a proxy of the actual proportion of PS that were running suggests that a child was effectively exposed to an average of 23-25 hours of the scheme during the school year.

On most monitoring visits, there were 2 teachers present, the average number of volunteer VMs observed was 8, the minimum number was 6 and there were around 24 children. In 93% of the cases hand washing practices were observed and in 77% of cases also the use of soap during hand washing. In most cases mats were organised as expected and in more than half of the cases the practice of discovery-based teaching was observed.

In addition, each school received on average between 4 and 5 monitoring visits at the monthly parental workshops organised over the academic year. In 88% of the visits, the workshop was taking place. Again, among the most important reasons for a workshop not to take place was low

attendance of VMs. In most cases both teachers were present with an average attendance of 30 mothers per session. Only in 14% of the workshops observed, the head teacher was also present.

Survey data on compliance

In addition to the processing data discussed in the previous section, we also assess compliance to the *Lively Minds* intervention by considering a set of indicators available in the endline survey data collected by the survey team independently from the *Lively Minds* programme team. For example, PCs were asked whether in the last 12 months they or anyone else in their household had offered their services voluntarily to a school, church, health centre or any other service provider. If anyone in the household had been involved in any volunteering work in the school, specifically, the respondent was asked to describe the type of volunteering activities, which the enumerator then categorised (without prompting) as being consistent with PS activities or not. Those women who indicated to have volunteered in any PS were then asked whether they could show a PS training certificate and/or a Lively Minds training certificate.

Table 3. Volunteering

	(1) Any volunteer work	(2) Volunteering in school	(3) Volunteering in PS	(4) Training certificate from PS	(5) Training certificate from LM
<i>Average Mother</i>					
Treatment Effect	0.196*** (0.0277)	0.225*** (0.0278)	0.239*** (0.0254)	0.209*** (0.0263)	0.196*** (0.0253)
Control mean	0.289	0.070	0.010	0.006	0.006
<i>By VM status</i>					
Treatment Effect on VM	0.708*** (0.0386)	0.906*** (0.0288)	0.980*** (0.0105)	0.857*** (0.0383)	0.806*** (0.0386)
Treatment Effect on Non-VM	0.024 (0.0324)	-0.003 (0.0158)	-0.006 (0.0040)	-0.003 (0.0085)	-0.004 (0.0093)
Volunteer Mother (VM)	0.038 (0.0337)	0.050** (0.0238)	0.014 (0.0105)	-0.003 (0.0055)	-0.003 (0.0055)
P-value (F-Stat)	0.000	0.000	0.000	0.000	0.000
Control mean VM	0.326	0.114	0.022	0.005	0.005
Control mean non-VM	0.282	0.061	0.008	0.007	0.007
Observations	2206	2206	2206	2206	2206

Notes: All specifications control for district, strata, and imbalance controls. Standard errors clustered at the school level are reported in parenthesis. *, **, *** correspond to the 10%, 5%, and 1% levels of significance, respectively. The outcome in column (1) corresponds to whether or not the Primary Caregiver has conducted any volunteering work in the last 12 months. This includes volunteering in the church, in health centres, in the school and in other places. Column (2) examines the probability of volunteering in the school. Column (3) estimates the probability of volunteering for Play Schemes (PS). Columns (4) and (5) estimate the probability of having a training certificate and a LM certificate, respectively. All specifications have been estimated using linear probability model (OLS).

Table 3 presents the results of regressions estimating the difference in responses to these questions between PCs in treatment schools and those in control schools. The upper panel shows the average treatment effect across all PCs whereas the bottom panel shows treatment effect by whether or not the PC was identified (ex-post) as being a VM.

The results in Table 3 are clearly consistent with high compliance of treatment schools to the *Lively Minds* activities related to VM volunteering. On average, PCs (irrespective of their VM status) in treatment communities were 20 pp more likely to report to have done any volunteering work for any service provider in their community, 23 pp more likely to have done volunteering work in schools specifically, and 24 pp more likely to have volunteered in school PS. Considering treatment effects by VM status shows that, as expected, these effects are entirely driven by PCs who were identified as VMs. Virtually all VMs in treatment schools reported to have done volunteering in school PS (98%), compared to virtually zero in the control schools. Out of those who reported to have volunteered in PS, more than 9 out of 10 reported to have volunteered at least once a week and one in ten reported to have volunteered twice a week (not reported in results Tables). Almost nine in ten VMs in treatment schools could present the survey enumerator with a PS or *Lively Minds* certificate to confirm that they did indeed complete VM training.

Similarly, if the Lively Minds programme was implemented in the way that it was intended, then we would expect teachers in treatment schools to be significantly more likely to report to have received parental support to their KG activities. Moreover, we would expect to see a change in the teacher's description of the KG class activities on a typical KG day. For example, we would expect to see more toys and games available for play in their classrooms and we would expect them to be more likely to report play in small groups.

Table 4. Teacher reports related to compliance

	(1) Parents contribute	(2) Parents support	(3) Games for free play	(4) Play in small groups
Programme impact	4.814*** (2.017)	0.435*** (0.125)	0.338*** (0.100)	0.0300 (0.0320)
Control mean	2.188	0.388	0.525	0.887
Observations	160	155	156	139

Notes: All specifications control for district, strata, imbalance controls and the average teacher outcome at baseline. Standard errors clustered at the school level are reported in parenthesis. *, **, *** correspond to the 10%, 5%, and 1% levels of significance, respectively. The dependent variable in the first column is an ordered five categorical variable measuring the extent to which parents contribute to KG activities as per teacher report, ranging from 'not at all' to 'quite a lot'. We estimate the impact on this variable using an ordered logit model and the estimate of the treatment effect is presented in terms of a proportional odd ratio. Columns 2-4 show marginal effects of probit regression estimation results. The second column estimates the probability of teachers reporting that they received active support from parents of pupils in their class at least once in the past academic year. The last two columns estimate the probability of teachers reporting to have games and toys available for children during free play and whether or not any playing activities in their class are done in small groups, respectively. Observations in Columns 2-4 are reduced because of treatment predicting the outcome perfectly in some of the strata.

Table 4 presents results in terms of such teacher reports related to compliance. In the first column, we show ordered logit regression results of an ordered categorical variable that measures the extent to which teachers report that parents are actively contributing to the KG activities organised in their schools. Responses ranged from ‘Not at all’ (code 1) to ‘quite a lot’ (code 5). The ordered logit results presented in Column 1 are displayed as proportional odd ratios. We find that teachers in treatment schools were almost 5 times more likely to agree with the statement that parents actively contribute to the KG activities than teachers in the control schools.

Columns 2-4 in Table 4 present marginal effects of three different Probit regressions. Consistent with the result in Column 1, Column 2 shows that parents were 44 pp more likely to report that they received active support from parents in the running of KG activities at least once in the last academic year. They were also significantly more likely (34 pp) to report that there are any toys, games, or other equipment accessible for children during free play on a typical KG day (Column 3). Interestingly, we do not find any difference in terms of teacher’s probability to report that some play activities are done in small groups on a typical KG day. However, the fact that even in control schools nearly 9 in 10 KG teachers reported play in small groups suggests that perhaps the question asking about play in small groups was not able to sufficiently capture the difference between *Lively Minds* PS and other forms of play that already pre-existed in KGs in our study area.

5.2 Implementation challenges

Before moving to the impact results, it is important to note that there have been some important implementation challenges related to the evaluation design which may have affected the performance of the *Lively Minds* activities and therefore also the impacts that we observe.

For example, *Lively Minds* usually targets the most deprived schools in remote areas, rather than targeting all schools within a particular district. To maximise sample size within districts, however, for this experiment *Lively Minds* was asked to intervene in some schools where they would not necessarily have operated otherwise. *Lively Minds* would normally also drop poorly performing schools much more rapidly (and instead move their efforts elsewhere) than they were able to in this study in order to maximise compliance. Also, to make use of economies of scale (in terms of training and supervision) *Lively Minds* usually saturates an entire circuit before moving to the next one. For this study, however, we randomised schools *within* circuits to improve comparability of control and treatment schools.

Another example relates to the speed at which learning feeds back into programme design. *Lively Minds* would normally continuously test their programme using rapid assessments and immediately adapt their activities accordingly. To maintain the integrity of the RCT study design, however, in this study they were for example asked not to pass on the midline assessment results to the study communities. Finally, whereas *Lively Minds* would normally embrace and encourage spillover effects to other communities, we put measures in place to minimise possible contamination from treatment to control communities.

Overall, these challenges imply that for the evaluation to work *Lively Minds* had to some extent deviate from their typical model in a way that might have made their activities perhaps somewhat

less effective than otherwise. One could therefore argue that the impacts discussed in this report should be considered lower bounds.

6 Empirical strategy

In this section, we present the details of our empirical strategy to evaluate the *Lively Minds* programme. We start with a description of our estimating equation. We then proceed by describing the results of tests of baseline balance between the control and the treatment group and we conclude by describing the procedure through which we construct our outcome measures.

6.1 Estimating equation

Our experimental design allows us to study the causal impact of *being eligible for the Lively Minds* programme (intent-to-treat effect – ITT). In other words, we compare outcomes of children who were reported during pre-baseline census to be attending one of the “treatment” schools to those reported to be attending one of the “control” schools.

While the experimental design would allow us to compare simply the means (or the distributions) of the outcomes of interest between treatment and control sample, conditioning the outcomes at endline for a set of observable variables (including the outcome at baseline) might improve the precision of our estimates and, given the sample size, this can be very useful. Formally, we identify ITT of *Lively Minds* on each outcome of interest with the help of the following regression:

$$Y_{is,1} = \alpha + \beta Treat_{s,0} + \sigma Y_{is,0} + \gamma X_{is,0} + \theta Strata_s + \delta District_s + \varepsilon_{is,1} \quad (1)$$

where $Y_{is,1}$ is the outcome of interest for child i in school s measured at endline; $Treat_{s,0}$ is a dummy equal to 1 if the school s received the *Lively Minds* programme; $Y_{is,0}$ is the same outcome measured at baseline⁹; $X_{is,0}$ is a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline (see Section 6.2); $Strata_s$ is a fixed effect for the randomisation strata of school s ; $District_s$ is an indicator for whether the school is located in Bongo or Tolon district; finally $\varepsilon_{is,1}$ is the random error term, clustered at the school level (unit of randomisation).

For continuous outcomes, we estimate equation (1) by OLS so that β is the estimated average ITT impact of *Lively Minds* on outcome $Y_{is,1}$. For binary outcomes we estimate equation (1) using a logit model and report the estimated adjusted difference in proportion of the outcome between the treatment group and the standard error of that difference using the STATA margins command.

6.2 Balance tests

One of the key assumptions underlying our evaluation design is the absence of systematic pre-intervention differences in trends and levels of characteristics of the treatment and the control group. We test this by examining if there are any significant differences in a range of observable characteristics of children, caregivers, households, teachers and communities (including those described in the previous section) between the two groups. Specifically, we test for differences

⁹ 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.

between the two groups by estimating the following regression using Ordinary Least Squares (OLS):

$$Y_{i,j} = \alpha + \beta T_j + s_j + \varepsilon_{i,j}$$

Where $Y_{i,j}$ is an indicator of interest for unit i (child, caregiver, household, teacher or community) in school j , and T_j is an indicator variable equal to 1 if j is a treatment school and s_j is a fixed effect for the randomisation stratum that school j belongs to. We report p-values for the hypothesis test that the mean of $Y_{i,j}$ in the control group is equal to the mean in the treatment group. The p-values for these statistical tests inform on the probability that a difference as big as the one we see could be due to chance if, in fact, no difference was present. By chance, we expect there to be 10 imbalances in every 100 characteristics at the 10% significance level or lower (5% or 1%). The higher the p-value the more similar our study groups are, statistically speaking, for that particular outcome.

To test for potential imbalances, we consider a set of characteristics that best describe the TC's environment at different levels: community, household, teachers, PCs, and TC. All child, household, caregiver and community characteristics considered in the tests were collected at baseline. We also present results of balance tests in terms of a set of time invariant endline characteristics of KG teachers in our endline sample (since we do not have baseline data available for half of the teachers in the endline sample, see Section 3.2 for a discussion).

The tables in Appendix E report the mean of each group and the p-value reflecting the statistical significance of this difference. We find for the TC/PC sample that out of the 189 variables examined, 18 (10%) of these are significantly different between control and treatment group, 8 of these at the 10% level of significance. This is exactly what would be expected by chance (at a minimum of 10% significance level) given the number of variables considered in this balance test. Similarly, for the VM sample we observe 19 out of 189 imbalances (10%) and for the teacher sample we observe 6 imbalances in 63 characteristics that we considered (10%). To account for such differences that occurred by chance, the imbalanced variables are used as controls in the main specification outlined in Section 6.1.

6.3 Construction of primary outcome measures

Measures of Child Development

As set out in Section 3.3, we administered three tools to measure TC's child development: the IDELA, Spelke Tasks instruments and the Strengths and Difficulties Questionnaire (SDQ). Together, these tools measure five developmental domains: emergent numeracy, emergent literacy, executive functioning, fine motor skills and socio-emotional development. Table 5 shows the domains measured by each of the tasks/questions in these three assessment tools.

Our aim is to create measures for each of the constructs listed in the table. In addition, we also create summary measures of cognitive and socio-emotional development. As Table 5 also shows, for emergent numeracy, we further construct measures of different aspects of emergent numeracy, namely spatial ability, numerical estimation, numerical vocabulary, and arithmetic.

Table 5. Mapping of tasks to child development domains

Constructs			Tasks
Cognitive	Emergent math	Spatial ability	Geometric intruder Comparison by size and length (Idela 2) Sorting and classification (Idela 3) Shape identification (Idela 4) Puzzle completion (Idela 8)
		Numerical estimation	Panamath
		Numerical vocabulary	Point to number Number identification (Idela 5) One to one correspondence (Idela 6)
		Arithmetic	Extra number Addition and substraction (Idela 7)
	Emergent literacy		Vocabulary assessment 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 Mental simulation / rotation Short term memory (Idela 13) Inhibitory control (Idela 14)
	Fine motor		Copying a shape (Idela 21) Drawing a person (Idela 22) Folding a paper (Idela 23)
Socio-emotional	Emotional awareness		Point to emotion Self-awareness (Idela 1) Friends (Idela 9) Emotional awareness/regulation (Idela 10) Empathy/perspective taking (Idela 11) Solving conflict (Idela 12)
	Externalising behaviour		Conduct problems (SDQ 5, 7*,12, 18, 22) Hyperactivity (SDQ 2, 10, 15, 21*, 25*)
	Internalising behaviour		Emotional problems (SDQ 3, 8,13, 16, 24) Peer problems (SDQ 6, 11*, 14*, 19, 23)
	Pro-social behaviour		Prosocial behaviours (SDQ 1, 4, 9, 17, 20)

* refers to sub-items for which we have reserved the scoring

Similarly, we construct measures of four distinct sub-domains of socio-emotional development: emotional awareness, which is the ability to recognise and make sense of not just one own emotions, but also those of others; externalising behaviour, which are problem behaviours that are directed toward the external environment and include physical aggression, disobeying rules, cheating, stealing and destroying things; internalising behaviour, which are negative behaviours

that are focused inwards and include fearfulness, social withdrawal and anxiety; and pro-social behaviours, which is a social behaviour that benefits other people, such as helping, sharing and co-operating.

We construct these measures is twofold. In the first step, we use latent variable modelling techniques to create a specific continuous score to measure the child's performance on each task. Depending on the nature and number of the questions in each task, we estimate different models to estimate this score. For example, for tasks including questions with binary answers (yes/no), we estimate a 1 or 2 parameter IRT analysis. For tasks including questions with ordinal answers (e.g. not true/ somewhat true/absolutely true) we estimate a graded response model in order to adequately account for the nature of the items.

There are two main advantages of using latent variable models for this type of analysis. First, since there is no prior evidence on the performance of our instruments in the study context, an important step is to assess how well the measures performed in this population. IRT allows us to estimate their reliability. Second, unlike a raw score, IRT allows assessment items to vary by how difficult they are and how well they discriminate between children of different underlying abilities. As such, conditional on key assumptions holding, IRT is an efficient approach to extract information about children's development from the measures, while purging them from the noise that they may contain.

Based on this analysis, we are able to identify and exclude from the analysis all items/assessments that performed poorly. We then repeat the procedure above to construct item-specific latent scores only using the sub-items and items that performed well. A detailed discussion of the IRT approach, our application of it in this study, as well as analysis of the validity of the measures is available on request.

The second step of the procedure consists in constructing composite measures based on the item-specific scores produced in the first step. Following the mapping of tasks to developmental domains in Table 5, we perform a confirmatory factor analysis of the latent scores of items measuring the same domain or sub-domain and predict a latent score for the corresponding domain or sub-domain. The constructed factors are then re-scaled to have a mean zero and standard deviation of one for the control group.

Measures of Child Health

We construct and analyse two different sets of measures of child health. The first set includes a measure of upper-arm circumference (in cm) and an indicator of malnutrition, which takes the value 1 if arm circumference is below the relevant threshold for the age and sex of the child (see Fiorentino et al., 2016 for threshold tables). The second set includes three measures based on PC reports of the child's morbidity over the last 30 days, as well as due to longer-term conditions. Specifically, caregivers were asked about the target child's long-term health (e.g. whether he/she has physical or mental disability, skin problems etc.). They were then asked about the target child's health over the last 30 days (e.g. stomach pains, coughs, fevers, etc.). We apply 2-parameter IRT to these data and construct three measures: morbidity, chronic illness and overall

health (which combines all the items). The constructed factors are then re-scaled to have a mean zero and standard deviation of one for the control group.

7 Impacts on primary outcomes

Having described how we construct our primary outcome measures, we now turn to presenting and discussing estimates of programme impacts on cognitive skills, socio-emotional skills and health. In the next section, we then explore impacts on secondary outcomes which we hypothesise constitute the mechanisms through which *Lively Minds* could impact targeted outcomes. Finally, we study heterogeneity in main impacts by child, parent, household and school characteristics. We adhere to the project protocol (Appendix A) in the empirical strategy we apply, the primary and secondary outcomes that we study and the hypotheses that we test.

7.1 Cognitive domains

Panel A, Col(1) of Table 6 shows impacts on our aggregated measure of cognition, obtained, as mentioned above, from the IDELA items and the additional items from the Spelke lab; this is followed by four columns showing impacts on individual sub-domains which make up the cognitive domain. These include emergent numeracy, emergent literacy, executive functioning, and fine-motor skills.

Lively Minds has a significant impact on cognition of pre-school children, raising it by, on average, 14.2% of a standard deviation. Breakdowns in columns (2)-(5) suggest that this is being driven by significant improvements across several sub-domains, including emergent numeracy, executive functioning and fine-motor skills. The effect sizes are very similar across these – between 15 and 15.8% of a standard deviation. We do not see any significant overall improvement in emergent literacy.

Given the sizeable effect observed on numeracy assessment, it may be interesting to decompose it into four sub-domains: spatial ability, numerical estimation, numerical vocabulary and arithmetic. Looking at programme impacts on these sub-domains can be informative for unpacking what is driving the observed overall improvement in numeric ability. The results, presented in Panel B of Table 6, suggest that *Lively Minds* is especially effective at improving spatial ability and numeric vocabulary. Spatial ability relates to one's capacity to understand and remember the spatial relations among objects and is measured by tasks involving shapes and puzzles, while numeric vocabulary reflects children's ability to recognise and name numbers.

7.2 Socio-emotional domains

Panel C of Table 6 presents results for programme impacts on socio-emotional skills. The first column shows impacts on the aggregated measure and is followed by four columns showing impacts on four sub-domains – emotional awareness, externalising, internalising and pro-social behaviours. While there is no significant impact on the aggregated measure, we see a significant reduction in externalising behaviours. On average, *Lively Minds* results in a 14% of a standard deviation reduction in externalising behaviours, including both conduct problems (e.g. lying, cheating, bullying) and hyperactive behaviours (e.g. being fidgety, being unable to concentrate on a task). There is no evidence of overall impacts on emotional awareness, internalising or pro-social behaviours.

Table 6. Programme impacts on primary outcomes

Panel A: Cognitive (COG) domains					
	(1) COG (All)	(2) Numeracy	(3) Literacy	(4) Executive Function	(5) Fine Motor
Treatment effect	0.142** (0.0599)	0.149** (0.0601)	0.066 (0.0685)	0.150** (0.0675)	0.158*** (0.0527)
Panel B: Numeracy sub-domains					
	(1) Spatial Ability	(2) Numeric Estimation	(3) Numeric Vocabulary	(4) Arithmetic	
Treatment effect	0.143** (0.0660)	0.042 (0.0597)	0.159*** (0.0542)	0.087 (0.0661)	
Panel C: Socio-emotional (SE) domains					
	(1) SE (All)	(2) Emotional Awareness	(3) Externalising Behaviour	(4) Internalising Behaviour	(5) Pro-social Behaviour
Treatment effect	0.103 (0.0916)	0.044 (0.0851)	-0.141** (0.0681)	-0.100 (0.0888)	-0.040 (0.1039)
Panel D: Health					
	Nutrition		General health (Caregiver reports)		
	(1) MUAC (cm)	(2) Acute Malnutrition	(3) Overall health	(4) Morbidity	(5) Chronic Problems
Treatment effect	0.213*** (0.078)	-0.075*** (0.028)	0.074 (0.085)	0.063 (0.084)	0.054* (0.028)
Control Mean	15.712	0.333			
Observations	2206	2206	2206	2206	2206

Notes: This table reports the coefficient on an indicator for whether the school attended a school that received the Lively Minds programme. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. For outcomes in Panels A and C, we also control for the child's baseline score in the same domain as the outcome. For Panel B, we also control for the child's baseline score in emergent math. For Panel D, we control for the child's baseline cognitive score and baseline overall health score. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

7.3 Health

Finally, in Panel D of Table 6 we show impacts on child health. There is a significant increase in children's mid-upper arm circumference of 21mm. This is equivalent to 18% of a standard deviation of the average outcome. It also translates into a reduction in acute malnutrition rate, defined using the cut-offs described above. The results suggest that children who attended *Lively Minds* schools were 7.5 percentage points less likely to be acutely malnourished at the time of the endline; this is equivalent to a 22.5% reduction relative to the mean of the control group.

These improvements are not reflected in health measures constructed using caregiver reports. Specifically, based on these, there is no change in the overall health or the morbidity sub-domain. There appears to even be a slight increase in reporting of chronic health problems among mothers of children in treatment schools. This increase is hard to explain, especially for chronic illnesses that are likely to evolve over a longer time-horizon than that covered by the study. A possible explanation is that caregivers in the treatment group become more aware of child health and illness symptoms.

7.4 Discussion

Before moving onto a more in-depth investigation of these findings we discuss the size of the effects, turning first to effects on measures of cognitive and socio-emotional development. The effect sizes we have presented so far are standardised statistics measuring the difference in test scores between the treatment and the control group expressed in terms of percentage change of a standard deviation (SD) of the distribution of the test scores of the control group. One way to assess whether the effect is large or small, could be to compare it to effect sizes achieved in other related interventions and studies. It is common practice in education, developmental psychology and related disciplines to review findings across categories of experiments and provide benchmarks for what constitutes a small, moderate and large effect size for a given type of experiments. The most relevant review for the purposes of this study suggests that in education experiments an effect size which is less than 0.05 SD should be considered small, an effect bigger than 0.20 SD of a standard deviation - large and anything in between as medium /moderate (Kraft et al, 2018). The review shows that raising academic achievement through school-based programmes is difficult; more than 1 in 5 of the 242 studies reviewed had an effect size of 0 or smaller, with many more failing to obtain traditional levels of statistical significance. Applying these benchmarks to our main finding, *Lively Minds* consistently has medium size impacts, edging towards large. What is especially striking about the *Lively Minds* results is that it achieves impacts across several developmental domains – cognitive, socio-emotional and health.

There are many critical caveats to consider when comparing effect sizes across studies in this way, both statistically and conceptually. However, without either having comparable assessments across studies or ways to link changes in test scores to changes in “objective” measures such as wages (neither of which we have in this context), we are forced to rely on these “back-of-the envelop” comparisons for contextualising our findings.

The finding of a strong impact on arm circumference with an effect size of 0.21 is especially striking given how few programmes have successfully achieved impacts on objective indicators of child health. For example, only very few programmes have successfully reduced stunting in

early childhood (perhaps the most widely used objective measure of child health)¹⁰. A recent review of evidence-based approaches to childhood stunting in LMICs by Hossain et al (2017) concludes that complex multi-sector approaches (e.g. a combination of nutrition education, WASH, growth monitoring and promotion, immunisation and social safety net programs) are usually required to effectively reduce stunting.

¹⁰ Notable exceptions include, for example, the well-known INCAP nutritional supplement intervention in Guatemala (Hoddinott et al, 2013) and Progresas's Conditional Cash Transfer program (Behrman et al, 2005)

8 Mechanisms

We now turn to exploring potential pathways for the impacts that we find. The pathways we consider are aligned with the programme theory of change, set out in Section 2.2 and the study hypotheses specified in the study protocol. We do this by examining impacts on a set of intermediary outcomes, which we hypothesise, may mediate impacts on the primary outcomes.

8.1 Impacts on intermediary outcomes

A key feature of *Lively Minds* is that it has the potential to affect children's outcomes not only through the direct impact of participation in the play-schemes, but also through educating, improving the well-being and changing the behaviour of the key actors in the kids' lives – including their caregivers and teachers. We start by studying evidence of programme impacts on primary caregivers and teachers, as well as the extent to which the specific significantly affected outcomes mediate the impact of the programme on child development. We then come back to consider how much of the programme impact can be attributed to direct effects of the play-schemes.

Impacts on Primary Caregivers

The PCs who are most likely to be affected by the *Lively Minds* programme are the VMs, who are trained on the curriculum and attend monthly workshops where they receive further training and information. However, we hypothesise that the programme may also affect caregivers who are not directly involved through interaction with the VMs. For all PC outcomes we therefore estimate impacts on the VMs as well as the non-VM PCs in the sample. To be clear, impacts on VMs compare the outcomes of mothers who actually volunteered in a treatment school during this trial with those of mothers who signed up to volunteer at a control (waitlisted) school for the following academic year. Impacts on non-VMs compare the outcomes of mothers who did not sign up to volunteer in the treatment schools with those who did not sign up to volunteer in the control schools. Despite the difference in how VMs and non-VMs were defined in control versus treatment communities, balance tests in Appendix E confirm that at baseline the sample of VMs in the treatment group was not significantly different from the sample of potential VMs in the control group.

Knowledge and Behaviour

We start by looking at Primary Caregiver knowledge relating to two areas. The first of these relates to knowledge of child development assessed (as described in Section 3.3) using six items from the Knowledge of Infant Development Scale (KIDI). The items ask about ways in which parents can support children's learning and development, how children learn and why it is. We construct the outcome measure using IRT to combine the six items into one factor and standardise it using control group mean and standard deviation. The second measure hones in specifically on parental knowledge relating to pre-school quality including most productive care and teaching strategies, key features of pre-school infrastructure, and parental involvement. Again, we use IRT to combine the four items into one factor and standardise relative to the control group.

Table 7. Programme impacts on VMs and non-VMs’ knowledge of child development and pre-school quality

	(1) KIDI	(2) Pre-school quality
Impact on VMs	0.290*** (0.0879)	0.316*** (0.1191)
Impact on non-VMs	0.034 (0.0788)	0.178** (0.0764)
P-value	0.021	0.256
Observations	2206	2206

Notes: This table reports estimates of the impact of the *Lively Minds* programme on outcomes of VMs and non-VMs. These are calculated from estimates of a regression of the outcome on an indicator for whether the mother’s child is in the treatment school (treatment dummy), a VM indicator taking the value 1 if the mother is a VM (in the treatment group) or a potential VM (in the control group), an interaction between the treatment dummy and the VM indicator. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The third row reports the p-value obtained from an F-test testing that the impact of the programme on VMs is equal to that on non-VMs. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

We see clear evidence that *Lively Minds* improves both general knowledge of the VMs about child development and pre-school quality. VMs achieve over a quarter of a standard deviation higher score on the KIDI items (Table 7). Analysis of the programme impact on individual items of the KIDI scale shows that the significant effect seems to be mainly driven by programme impacts on two items: VMs in the treatment group are more likely to agree with the statements that (1) it is possible for parents to engage with or engage children in games while doing their daily work; and (2) praising children when he/she tries to do something new is important.¹¹ VMs in treatment schools also score nearly a third of a standard deviation more on the pre-school quality knowledge scale. Further analysis of the individual components of this scale suggests that this increase is driven by treatment increasing the likelihood that a VM thinks that (a) a shabby building with books, posters on the wall and a separate play area is a higher quality pre-school than a polished new building without any of these; and (b) play based learning is better for kids than rote based learning.¹² While we see no evidence that there are spillovers to non-VM caregivers in general child development knowledge, we do see a statistically significant improvement in knowledge about pre-school quality among non-VM caregivers of children in treatment schools. This is not necessarily evidence of a spillover effect as all caregivers of children attending treatment KGs witnessed changes to the KG as the result of the programme.

To complement the knowledge items, we assessed parental behaviour through both parental report and direct observation of parents engaging in structured play-activities with kids. We administered the well-established parental report “Family Cares Indicators” questionnaire, which has been validated across many LMIC settings, as well as a more detailed set of questions on parenting strategies adapted from existing instruments validated for comparable contexts (see

¹¹ Results are not reported, but available from the authors upon request.

¹² Results are not reported, but available from the authors upon request.

Section 3.3). Impacts on all our measures of parenting behaviours are reported in Table 8. Overall, we find strong evidence of significant impacts across these measures on VMs; but nothing to suggest that there are any spillover impacts on behaviour of non-VM PCs of children in treatment schools.

Table 8. Programme impacts on VMs and non-VMs’ parenting behaviours

	FCI		Teaching style			Punishing behaviours		
	(1) Toys	(2) Time	(3) Observed	(4) Reported	(5) Combined	(6) Observed	(7) Reported	(8) Combined
Impact on VMs	0.098 (0.1549)	0.467*** (0.1030)	0.271** (0.1215)	0.274* (0.1386)	0.294** (0.1349)	0.004 (0.1030)	0.367*** (0.1302)	0.359*** (0.1287)
Impact on non-VMs	-0.069 (0.1400)	0.025 (0.0900)	0.094 (0.0753)	0.187* (0.1096)	0.150 (0.0912)	-0.001 (0.0981)	0.054 (0.1083)	0.052 (0.1081)
P-value	0.110	0.000	0.054	0.367	0.128	0.956	0.009	0.010
Observations	2206	2206	2206	2206	2206	2206	2206	2206

Note: This table reports estimates of the impact of the *Lively Minds* programme on outcomes of VMs and non-VMs. These are calculated from estimates of a regression of the outcome on an indicator for whether the mother’s child is in the treatment school (treatment dummy), a VM indicator taking the value 1 if the mother is a VM (in the treatment group) or a potential VM (in the control group), an interaction between the treatment dummy and the VM indicator. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The third row reports the p-value obtained from an F-test testing that the impact of the programme on VMs is equal to that on non-VMs. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

The FCI measures behaviour with respect to material goods parents make/buy that are good for child development, as well as activities they undertake with the kids. We construct two factors distinguishing between these two types of behaviours. As shown in the first two columns of Table 8, we find no impact whatsoever on the material factor (FCI toys), but we see a strong significant positive effect of nearly half of a standard deviation on “productive” time VMs report spending with their kids. Item specific analysis suggests that this effect is driven especially by higher likelihood of VMs in the treatment group reporting that someone in the household did the following activities with the target child in the last 3 days: reading, stories, singing and playing.

While the FCI time measure captures *whether* parents are engaging in developmentally appropriate activities with their kids, it is not informative about the quality of parent-child interaction, which is critical for child development. In addition to the FCI, we therefore administered measures capturing teaching and punishment strategies. For this, parents and observers were asked a common set of items – parents in relation to a “typical month” and observers in relation to what they observed during play tasks undertaken by PCs and target children as part of the endline survey. The advantage of the parental report measure is that it informs on what is “typical” for the child. The observer measures, on the other hand, may be less subject to reporting biases such as the social desirability bias (Edwards, 1957). Since both of these reflect possibly different dimensions of the same underlying constructs/types of behaviours, we construct separate measures for observed and reported behaviours, as well as combine reported and observed items

relating to the same broad parenting strategy to construct a single measure of that strategy using IRT.

The items we administer relate to two different types of behaviour – encouraging /productive teaching style and punishing behaviours. The higher the score on the punishing scale, the more harsh the punishment strategies that parents report/are observed to adopt in dealing with bad behaviour. The higher the score on the productive teaching style scale the more encouraging and constructive the behaviour parents report/are observed to adopt in encouraging good behaviour / teaching a task.

The results suggest that *Lively Minds* has an impact on both of these types of behaviours among Volunteer Mothers. First, there is a positive and statistically significant effect on practicing productive teaching strategies (columns 3 to 5 of Table 8). The effect on the combined measure is very similar to the effect on the observed and reported measures at around 29% of a standard deviation. Analysis of the programme impact on individual behaviours included in these overall scales shows that the effect is driven by VMs in treatment group being more likely to act productively across most items including, for example, praising kids, providing feedback, appearing confident in performing tasks with the child, breaking task down into small steps, turning the task into a game and using descriptive words for objects and actions.¹³ Again, effects are restricted to Volunteer Mothers, with no evidence of spillover effects on other PCs of children in treatment schools.

The last three columns of Table 8 show a more puzzling result. We see that while on average potential VMs in the control group report and are observed to use less harsh punishment strategies (e.g. hitting, calling names, shouting) than non-VMs in the control group, VMs in the treatment group are significantly more likely to use harsh punishment strategies than potential VMs in the control group. While we see a significant impact on the overall score, this is driven entirely by significantly higher rates of *reporting* such behaviours: there is no evidence of an impact in what was observed. This is unlike what we find for the other set of behaviours where the treatment coefficient is the same sign and similar magnitude for both reported and observed behaviours. Given the stark difference between what was observed and what was reported in this domain, it is unclear how to interpret this increased reporting of punishing behaviours. A possible explanation for this pattern is that the VMs in the treatment group spend more time with their children as a result of the intervention and may report higher incidence of all types of interactions with their children.

Maternal well-being

While there is clear evidence that participation in *Lively Minds* improves the knowledge and changes behaviour of the participating mothers, we find little to suggest that it has any impact on their well-being including mental health and self-esteem. In line with the other measures, we construct mental health and self-esteem factors using responses to the items in each of the scales. We show impacts on both the factor and raw score. None of the impacts on either the Volunteer Mothers or the other caregivers are statistically significant (Table 9).

¹³ Results are not reported, but available from the authors upon request.

Table 9. Programme impacts on mothers' well-being

	Mental health		Self-esteem	
	(1)	(2)	(3)	(4)
	Raw score	IRT	Raw score	IRT
Impact on VMs	0.757 (0.6536)	0.167 (0.1251)	0.043 (0.4390)	-0.039 (0.1289)
Impact on non-VMs	0.657 (0.4779)	0.139 (0.0952)	-0.333 (0.3274)	0.026 (0.0824)
P-value	0.870	0.811	0.376	0.575
Observations	2206	2206	2206	2206

Notes: This table reports estimates of the impact of the *Lively Minds* programme on outcomes of VMs and non-VMs. These are calculated from estimates of a regression of the outcome on an indicator for whether the mother's child is in the treatment school (treatment dummy), a VM indicator taking the value 1 if the mother is a VM (in the treatment group) or a potential VM (in the control group), an interaction between the treatment dummy and the VM indicator. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The third row reports the p-value obtained from an F-test testing that the impact of the programme on VMs is equal to that on non-VMs. Outcome in column (1) is measured on a 20-point scale. Outcome in column (3) is measured on a 30-point scale. Outcomes in columns (2) and (4) are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

Impacts on teachers

In addition to changing the knowledge and behaviour of the Primary Caregivers, we hypothesise that *Lively Minds* may change the knowledge and practices of the teachers. There are several mechanisms through which this could happen. First, *Lively Minds* provides teachers with training, both on the *Lively Minds* curriculum of educational play strategies, as well as on broader information relating to child development and family well-being. Second, teachers receive more support in the classroom through the Volunteer Mothers and may be able to allocate their time more productively. Third, teachers may feel more motivated and more accountable to parents as the result of greater parental involvement.

The impact of *Lively Minds* on the schools where they operate and the teachers they directly work with is a first order question, especially as the programme continues to be scaled through the education system. However, the focus of this evaluation is on studying impacts on the participating children. The samples of schools and teachers (80 and 160 respectively) are not nearly big enough to provide the statistical power necessary to evaluate school and teacher impacts rigorously. We collected data on teachers, nevertheless, and present estimates of treatment effects on some key outcomes. A critical caveat to bear in mind when considering the findings we now present is that being underpowered means that there is a high chance of not being able to reject hypotheses of no effects, when in reality there are some.

We study impacts on three broad categories of outcomes: teacher knowledge, behaviour and well-being. Teacher knowledge is captured through a literacy assessment and questions about perceptions relating to optimal strategies and practices for fostering development in early

childhood. We measure teacher behaviour using several indicators of their time-allocation, as well as reported practices in the classroom, including those related to teaching style and discipline. Finally, we administer scales measuring mental health, motivation, agency, burn-out and job satisfaction to assess teacher well-being. While our empirical strategy for estimating impacts is the same as above (see Section 6), because we are now looking at a sample of 160 teachers, we control for the variables that are imbalanced at baseline in this sample (rather than the full child sample): Table 5 in Appendix E shows results for teacher sample balance analysis. Further, as discussed in Section 3.2, due to significant teacher attrition we have baseline data available only for half of the endline teacher sample. Therefore, instead of controlling for individual baseline levels of outcomes (as we do in our child development outcome regressions), here we control for average baseline teacher outcomes to improve the precision of our estimates of impacts on teacher outcomes.

We find no evidence that *Lively Minds* has an impact on teachers' knowledge (Table 10): there are no significant impacts on either the teacher literacy or ECD perceptions factors, such as the value they attach to KG teacher's knowledge of developmentally appropriate practice (e.g. know about children's age specific needs, encourage children to recognise letters or words), the extent to which KG teachers support children's social and emotional development (e.g. help children to build relationship with peers and adults) and whether they engage in family-sensitive caregiving (e.g. consult with parents, connect families to external resources).

Table 10. Programme impact on teacher's literacy and knowledge about ECD practices

	Knowledge of early childhood development practices				
	(1) Literacy	(2) ECD Practice	(3) Socio- emotional Development	(4) Family	(5) All
Treatment effect	-0.195 (0.164)	-0.324 (0.281)	-0.175 (0.129)	-0.0666 (0.187)	-0.217 (0.181)
Observations	160	160	160	160	160

Notes: This table reports the estimate of the coefficient on an indicator for whether the teacher works in a treated school in a regression of the teacher's outcome on such indicator, a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** < 0.05, * < 0.1.

The evidence in Table 11 also shows that teachers in *Lively Minds* schools do not significantly change their time allocation either across the activities that they do with the kids during the school day, or total hours working and preparing. However, the results in Table 12 indicate that there may be an impact on teacher practices in the classroom. While there is no impact on scales reflecting overall level of engagement with parents (e.g. frequency of meetings with parents) and in productive activities in the class (e.g. teach colours, improve problem solving, singing, etc.), *Lively Minds* appears to change the way in which teachers deal with (or at least report dealing

with) good behaviour: teachers in *Lively Minds* schools are significantly more likely to use words through praising/cheering and significantly less likely to give the child something, like a sweet.

Table 11. Programme impact on teachers' time use

	Weekly hours		Hours spend daily on different classroom activities					
	(1) Prep	(2) Total	(3) Admin	(4) Discipline	(5) Teaching	(6) Superv. Free play	(7) Superv. teachers	(8) Break
Treatment Effect	1.454 (0.997)	1.798 (1.386)	-0.0923 (0.255)	0.117 (0.181)	0.496 (0.329)	-0.0291 (0.309)	-0.315 (0.218)	-0.178 (0.182)
Observations	160	160	160	160	160	160	160	160

Notes: See note to Table 10.

Table 12. Programme impact on teachers' classroom practices

	(1)	(2)	(3)	(4)	(5)
	Engagement with parents	Productive activities	Responding to good behaviour		
			Words	Sweets	Always praises
Treatment Effect	-0.248 (0.175)	0.115 (0.219)	0.178** (0.0701)	-0.169*** (0.0586)	0.0141 (0.0756)
Observations	160	160	160	160	160

Notes: See note to Table 10.

Finally, in Table 13 we do not find any impacts of *Lively Minds* on direct measures of well-being of teachers, including mental health, burnout and agency. There are, however, indications in Table 14 that teacher job satisfaction is improved in two important domains – feeling more supported by the parents and the supervisors. First, teachers were asked to respond to “problems” scale, focusing on problems related to school and parents. School-related problems included classes being too large, lack of human and financial resources, and inadequate curriculum. Issues related to parents include unreasonably demanding parents, as well as insufficient interest and involvement among parents in the education of their children and KG activities. The results suggest that teachers in *Lively Minds* schools report less problems with parents. Although there is no overall effect on problems reported with school, analysis of a more detailed job satisfaction scale covering four core domains: (1) relationship with colleagues; (2) relationship with supervisors; (3) general work motivation; (4) satisfaction with school and classroom capacity suggests that there is a significant improvement in teachers' relationship with supervisors in *Lively Minds* schools. This sub-scale contains items relating to teachers feeling respected by the supervisor, as well as feeling that they can count on him/her. There is a 43% of a standard deviation increase in this sub-domain as the result of the programme.

Table 13. Impact of the programme on teacher's well-being

	(1)	(2)	(3)	(4)	(5)
	Mental Health		Motivation Factor	Agency Factor	Burnout Factor
	Raw score	IRT score			
Treatment Effect	0.106	0.185	-0.321	0.0660	-0.0880
	(0.561)	(0.159)	(0.229)	(0.174)	(0.161)
Observations	160	160	160	160	160

Notes: See note to Table 10.

Table 14. Impact of the programme on teachers' job satisfaction

	Problems with:		Job satisfaction items			
	(1)	(2)	(3)	(4)	(5)	(6)
	Parents	School	Satisfaction with colleagues	Satisfaction with supervisor	Work motivation	Capacity
Treatment Effect	-0.315*	-0.0221	0.101	0.432**	0.0232	-0.0807
	(0.175)	(0.156)	(0.225)	(0.180)	(0.177)	(0.195)
Observations	160	160	160	160	160	160

Notes: See note to Table 10.

8.2 Mediating analysis

Overall, we see strong impacts of *Lively Minds* on knowledge and practices of Volunteer Mothers, much less evidence of impacts on teachers (though our design is not suited to robust evaluation of these) and almost no evidence of impacts on caregivers of children in treated schools who are not Volunteer Mothers. We, therefore, focus on two key pathways through which the programme may have achieved the overall effects on child outcomes that we find (Section 7) – direct effects of the play-schemes/being in a treated school and indirect effects through changes in parenting practices among Volunteer Mothers.

To investigate this further, we now look at whether there are differences in programme effects by whether the child's mother is a Volunteer Mother. The Volunteer Mother coefficient ('Child's mother is VM') informs on the difference between caregivers who volunteer to participate in the *Lively Minds* programme and those who do not, in a control community. The results tables show separately the treatment effect on Volunteer Mothers and non-Volunteer mothers and the P-value of the F-stat informs on the significance of the difference between these two treatment effects.

Table 15. Programme impact on primary outcomes of children of VM and children of non-VMs

Panel A: Cognitive (COG) domains					
	(1)	(2)	(3)	(4)	(5)
	COG (All)	Numeracy	Literacy	Exec. Function	Fine Motor
Impact on VM child	0.083 (0.0789)	0.159 (0.0964)	-0.039 (0.0941)	0.119 (0.1011)	0.015 (0.0719)
Impact on non-VM child	0.145** (0.0623)	0.131** (0.0606)	0.078 (0.0672)	0.152** (0.0739)	0.184*** (0.0580)
Child's mother is VM	0.172*** (0.0539)	0.146** (0.0646)	0.225*** (0.0575)	0.090 (0.0833)	0.221*** (0.0583)
P-value	0.399	0.751	0.133	0.763	0.032
Panel B: Socio-emotional domains					
	(1)	(2)	(3)	(4)	(5)
	SE (All)	Emotional awareness	Externalising behaviour	Internalising behaviour	Pro-social behaviour
Impact on VM child	0.219 (0.1342)	0.093 (0.1151)	-0.294** (0.1190)	-0.233* (0.1328)	-0.107 (0.1588)
Impact on non-VM child	0.071 (0.0905)	0.022 (0.0857)	-0.101 (0.0668)	-0.061 (0.0894)	-0.038 (0.1011)
Child's mother is VM	-0.057 (0.0880)	0.057 (0.0613)	0.107 (0.0854)	0.057 (0.0914)	0.197* (0.1053)
P-value	0.179	0.413	0.083	0.139	0.592
Panel C: Health					
	Nutrition		Health (Caregiver reports)		
	(1)	(2)	(3)	(4)	(5)
	MUAC (cm)	Acute malnutrition	Overall health	Morbidity	Chronic Problems
Impact on VM child	0.244** (0.1129)	-0.065 (0.0498)	0.130 (0.1245)	0.117 (0.1242)	0.082 (0.0497)
Impact on non-VM child	0.199** (0.0799)	-0.075*** (0.0289)	0.066 (0.0866)	0.058 (0.0848)	0.043 (0.0298)
Child's mother is VM	0.035 (0.0744)	-0.035 (0.0353)	-0.117 (0.0784)	-0.127 (0.0794)	0.013 (0.0353)
P-value	0.649	0.837	0.547	0.572	0.459
Observations	2206	2206	2206	2206	2206

Notes: This table reports estimates of the impact of the *Lively Minds* programme on outcomes of children of VMs and non-VMs. These are calculated from estimates of a regression of the outcome on an indicator for whether the child is in the treatment school (treatment dummy), an indicator “Child’s mother is VM” taking the value 1 if the child’s mother is a VM (in the treatment group) or a potential VM (in the control group), an interaction between the treatment dummy and the latter indicator. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The fourth row reports the p-value obtained from an F-test testing that the impact of the programme on VM children is equal to that on non-VM children. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

First, consistent with the descriptive findings in Section 4, we find that in both treatment and control schools the women who opt into volunteering for *Lively Minds* have children with significantly higher cognitive development than children of women who do not. This is true for their overall cognitive development, as well as across each of the constituent domains, except for executive function (Table 15, Panel A).

We further see that the sizes and significance of the treatment coefficients are lower for children of VMs across most of the cognitive outcomes. In fact, while *Lively Minds* has a significant positive effect on three out of four cognitive domains and our aggregate measure of cognitive development on non-VM children, none of the effects on children of VMs are statistically significant (see Table 15, Panel A). It should be noted that only one of the *differences* in treatment effects between VM and non-VM children is statistically significant (fine motor skills) and that because of the small sample of VMs with children in our study sample, this analysis suffers from lack of statistical power. At the very least, however, the fact that we clearly do not observe greater impacts on VM kids combined with our earlier finding that the programme only affects knowledge and behaviour of VMs suggest that any cognitive gains from *Lively Minds* are likely to come from attending a participating school, with no additional gains being generated by changes in parental behaviour at home.

This is different from what we see for effects on socio-emotional skills in Panel B of Table 15. There is no consistent evidence that the women who opt into volunteering for the *Lively Minds* programme have children with higher socio-emotional skills than those who do not. Further the effect of *Lively Minds* on reducing negative externalising behaviours is only statistically significant for children of VMs, with a coefficient on the treatment indicator which is around three times the size of the coefficient on the treatment indicator for kids of non-VMs. We also see some indication of a significant effect of *Lively Minds* on reducing negative internalising behaviours among VM kids and a marginally significant effect on the overall measure of socio-emotional skills (p-values of 0.107). While the *differences* between the VM and non-VM kids are not quite statistically significant, the results do suggest that unlike in the case of cognitive skills, participation in the play-schemes is not sufficient for significant gains in socio-emotional skills and changes in parental behaviour are key.

Perhaps surprisingly, there is also no evidence that the health benefits of the programme are greater for children of VMs than non-VMs (see Panel C of Table 15). This suggests that the effects are more likely to reflect changes in WASH practices in treatment schools rather than changes in nutritional practices at home.

In order to investigate these important potential differences in programme impact mechanisms across the different child outcomes further we conduct mediation analysis (Imai et al., 2011). The analysis allows us to investigate directly the degree to which the effects of *Lively Minds* on child outcomes are explained by its impact on caregiver behaviour (especially that of the VMs). We start by assessing the direct relationship between our hypothesised mediators (caregiver knowledge and child related practices) and the child outcomes which we find to be significantly impacted by the programme. We then test the degree to which those hypothesised mediators that are significantly associated with the outcomes explain the impact of *Lively Minds* on the outcomes by adding them into our main estimating equation (Equation 1). To the extent that

some of the *Lively Minds* effect works through the channels we are testing, we would expect the size of the coefficient on the treatment effect to decline significantly.

Table 16 shows the results of this analysis for child cognitive outcomes. Columns (1)-(4) show associations between the caregiver knowledge and behaviour measures (which we find to be significantly impacted by the programme – see Table 8) and the cognitive outcomes which we find to be significantly positively impacted by *Lively Minds*, including numeracy, executive function, fine motor skills and the overall measure of cognitive development (Table 6, Panel A). While there is little evidence of significant associations between the caregiver variables and the numeracy, executive function and fine motor sub-domains (Cols 2-4), there is a significant positive association between caregiver teaching style and child cognitive development overall, as well as a negative association between practice of harsher punishment and child cognitive development (Column 1). However, results presented in Columns (5) and (6) suggest that these are unlikely to be important mediators of the effect of *Lively Minds* on child development: there is almost no change in the size of the treatment effect coefficient with the addition of the teaching style and harsh punishment variables. Consistently with this we cannot reject the hypothesis of equality of treatment coefficients in the specification with and without controls for these caregiver behaviours. This pattern of results is consistent with our earlier conjecture that cognitive gains from *Lively Minds* are likely to come from attending a participating school, with no evidence of substantial additional gains being generated by changes in parental behaviour at home.

In contrast, mediation analysis does offer further suggestive evidence that parental behaviours mediate impacts of *Lively Minds* on socio-emotional development. Table 17 shows analysis for the whole study sample. Column (1) shows that children with mothers who have better knowledge about pre-school quality, as well as a more developmentally appropriate teaching style have children with fewer problems of externalising behaviours. Columns (2) and (3) then show that adding these to the main treatment effect specification significantly reduces the size of the treatment effect of *Lively Minds* on externalising behaviours by nearly 35%. We extend this analysis in Table 18 by restricting the analysis sample to kids of VMs and potential VMs, as this is the sub-sample for which we see biggest impacts on socio-emotional development and caregiver behaviour. The findings show clearly that across the socio-emotional domains, 25-35% of the positive impact of *Lively Minds* is mediated by improvements in caregiver knowledge and behaviour achieved by the programme.

Finally, as in the case of cognitive development, we see no evidence that improvements in parental knowledge and practice mediate programme impacts on child nutrition (Table 19). Again, this is consistent with earlier results which suggest that the impact on nutrition is driven primarily by changes that take place at the school rather than home level as the result of the programme. An important caveat to the mediation analysis is that unlike the analysis that we have presented up to this point, the randomisation is not sufficient here for ensuring that we are comparing like with like and it is plausible that any findings are, at least in part, driven by unobserved factors. It is, nevertheless informative, as it supports the narrative that started to emerge from our earlier results, adding credibility to the key finding that the mechanisms for programme impacts are likely to differ for different child outcomes.

Table 16. Mediation analysis of the programme impacts on cognitive domains

	(1) COG (all)	(2) Numeracy	(3) Executive function	(4) Fine Motor	(5) COG (all)	(6) COG (all)
KIDI	-0.012 (0.0165)	-0.016 (0.0185)	0.011 (0.0259)	0.010 (0.0268)		
School quality knowledge	0.025 (0.0157)	0.024 (0.0176)	0.025 (0.0214)	0.025 (0.0205)		
FCI (time)	0.009 (0.0173)	0.014 (0.0183)	0.014 (0.0218)	0.015 (0.0216)		
Teaching style	0.035** (0.0176)	0.038* (0.0199)	0.042 (0.0265)	0.042 (0.0266)		0.034* (0.0173)
harsh Punishment	-0.029* (0.0163)	-0.018 (0.0184)	-0.013 (0.0212)	-0.008 (0.0226)		-0.033** (0.0162)
Treatment effect					0.142** (0.0599)	0.140** (0.0600)
P-value (diff in treatment coeff)					0.518	
Observations	2206	2206	2206	2206	2206	2206

Notes: The tables reports the results from regressions of different measures of cognitive development on the variables listed in the table, as well as a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The second to last row entitled "P-value (diff in treatment coeff) reports the p-value obtained from an F-test testing the hypothesis that the treatment effect in the regression reported in column (5) is equal to the treatment effect in the regression reported in column (6). All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

Table 17. Mediation analysis of the programme impacts on externalising behaviour

	(1)	(2)	(3)
	Externalising behaviour		
KIDI	-0.041 (0.0293)		
School quality knowledge	-0.147*** (0.0273)		-0.152*** (0.0258)
FCI (time)	0.023 (0.0210)		
Teaching style	-0.085** (0.0416)		-0.068* (0.0389)
Harsh Punishment	0.049 (0.0316)		
Treatment effect		-0.141** (0.0681)	-0.093 (0.0631)
P-value (diff in treatment coeff)			0.002
Observations	2206	2206	2206

Notes: The tables reports the results from regressions of the measure of externalising behaviour on the variables listed in the table, as well as a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The second to last row entitled "P-value (diff in treatment coeff) reports the p-value obtained from an F-test testing the hypothesis that the treatment effect in the regression reported in column (2) is equal to the treatment effect in the regression reported in column (3). The outcome variable is a continuous standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

Table 18. Mediation analysis of the programme impacts on socio-emotional domains amongst children of VM mothers

	(1)	(2)	(3)	(4)	(5)	(6)
	SE (All)		Externalising Problems		Internalising Problems	
KIDI	0.051 (0.0602)		-0.052 (0.0657)		-0.050 (0.0619)	
School quality knowledge	0.175** (0.0677)		-0.179*** (0.0672)		-0.124* (0.0675)	
FCI (time)	-0.073 (0.0442)		0.061 (0.0566)		0.056 (0.0472)	
Teaching style	0.169** (0.0726)		-0.183** (0.0730)		-0.167** (0.0717)	
Harsh Punishment	0.060 (0.0649)		0.047 (0.0570)		-0.088 (0.0719)	
School quality knowledge		0.158** (0.0659)		-0.169*** (0.0619)		-0.105 (0.0651)
Teaching style		0.166** (0.0669)		-0.145** (0.0677)		-0.171** (0.0665)
Treatment effect	0.331** (0.1543)	0.224* (0.1314)	-0.442*** (0.1416)	-0.340*** (0.1240)	-0.354** (0.1577)	-0.262* (0.1402)
P-value (diff in treatment coeff)		0.027		0.024		0.055
Observations		437		437		437

Notes: The tables reports the results from regressions of different measures of socio-emotional development on the variables listed in the table, as well as a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district on the sample of children whose mothers are VMs in the treated schools and potential VMs in the control schools. The second to last row entitled "P-value (diff in treatment coeff) reports the p-value obtained from an F-test testing the hypothesis that the treatment effect in the regression reported in column (5) is equal to the treatment effect in the regression reported in column (6). All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

Table 19. Mediation analysis of programme impacts on health

	(1) MUAC	(2) Mal- nourished	(3) MUAC	(4) MUAC	(5) Mal- nourished	(6) Mal- nourished
KIDI	-0.050 (0.0314)	0.008 (0.0112)				
School quality knowledge	0.019 (0.0239)	-0.004 (0.0102)				
FCI (time)	-0.004 (0.0269)	0.005 (0.0124)				
Teaching style	0.057* (0.0315)	-0.024** (0.0113)		0.058* (0.0303)		-0.023** (0.0106)
Harsh Punishment	0.035 (0.0257)	-0.009 (0.0101)				
Treat			0.209*** (0.0779)	0.198** (0.0781)	-0.078*** (0.0284)	-0.073*** (0.0279)
P-value (diff in treatment coeff)			0.19		0.17	
Observations	2206					

Notes: MUAC refers to Middle-Upper Arm Circumference and is measured in cms. Malnourished is an indicator for whether MUAC is below age and sex specific thresholds. The tables reports the results from OLS (probit) regressions of MUAC (an indicator for being malnourished) on the variables listed in the table, as well as a set of child, household, community and school characteristics measured at baseline that were imbalanced across the treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon district. The results reported in columns (2), (5) and (6) report marginal effects from probit regressions. The second to last row entitled "P-value (diff in treatment coeff)" reports the p-value obtained from an F-test testing the hypothesis that the treatment effect in the regression reported in column (5) is equal to the treatment effect in the regression reported in column (6). Standard errors in parentheses are clustered at the school level (unit of randomisation). *** p-value < 0.01, ** <0.05, * <0.1.

8.3 Spillover Effects

While the focus of the evaluation is on impacts on the target children, we hypothesise that the programme may have positive spillover effects on other children in the household. These are important to consider in order to avoid underestimating the full impacts of the programme. Spillovers may set in through interaction between siblings and target children with improved cognitive and socio-emotional skills, and/or through changes in parental behaviour.

A study design that would allow us to rigorously assess spillover effects was outside the scope of this evaluation. However, we did administer brief developmental assessments to the next oldest and the next youngest sibling of the target child (if they had one). Details of the assessments administered can be found in Section 3.3.

We do not find evidence of spillover effects either on older or younger siblings of target children (Table 20). However, there are important caveats. First, the sibling assessments were nowhere near as in-depth as those conducted for the target children and, therefore, do not discriminate between children as well as the target child assessments, making it difficult to detect especially small-moderate effects. Second, the sibling sample size is much smaller than the target child sample size, with 764 younger siblings and 432 older siblings.

Table 20. Impacts of the programme on younger and older siblings' development

	Younger sibling	Older sibling				
	CREDI	Point to emotion	Vocabulary assessment	Geometric intruder	Literacy test	Numeracy test
Treatment Effect	0.098 (0.1039)	-0.043 (0.1304)	0.014 (0.1268)	-0.163 (0.1227)	-0.164 (0.1200)	0.006 (0.1258)
Observations	764	432	432	432	432	432

8.4 Impact Heterogeneity

Finally, before drawing our findings together in a discussion, we examine whether there is evidence of differences in impacts by child and household characteristics, including sex, age, and baseline cognitive and socio-emotional skills of the target child, as well as the socio-economic status of his/her household. Observing impact heterogeneity in some dimensions could shed further light on the mechanisms for the impacts. Tables 21-23 show that while *Lively Minds* is equally effective at raising the cognitive skills of boys and girls, it is significantly more effective at raising socio-emotional skills of boys. There is a large and statistically significant overall effect on the socio-emotional skills of boys of 20% of a standard deviation, with a reduction in both externalising and internalising behaviours of a similar magnitude. Strikingly, we see no impacts of the programme on the socio-emotional skills of girls, who, on average score significantly higher across most of the socio-emotional domains than boys in both control and treatment schools.

We find little to suggest that programme effectiveness varies by child age. Children in the sample range in age between 46 and 85 months at endline. Although Tables 21-22 show that the effect-size of *Lively Minds* on both the cognitive and socio-emotional domains are higher for older children (kids above median age of 57 months), the difference between the two groups is never statistically significant. We also do not see stark differences in the effectiveness of *Lively Minds* by levels of baseline cognitive and socio-emotional development of the children, suggesting that effects are not confined to either developmentally more or less advanced children.

There is, however, a clear indication that *Lively Minds* is particularly effective at raising cognitive development of children from poorer households: the impact on overall cognitive development of children from households in the bottom quintile of the wealth distribution is more than twice as high as that on the rest of the kids and is nearly a quarter of a standard deviation (Table 21). Looking at cognitive sub-domains, we observe impacts on emergent numeracy for both the poorer and better off kids, but impacts on the former group are more than twice as high. Interestingly, whereas we did not find any impacts on emergent literacy for the average child in our overall sample (Panel A, Table 6), the results in Table 21 demonstrate significant impacts on literacy for the children living in the poorest quintile families. The exception to this pattern is executive functioning: here the trend appears to be reversed with the programme only having significant positive impacts on the relatively better off children (although differences in effects on executive functioning of poorer and better off children are not statistically significant).

Table 21. Heterogeneity in impacts on cognitive domains

	(1)	(2)	(3)	(4)	(5)
	COG (All)	Numeracy	Literacy	Exec. Function	Fine Motor
<i>By gender</i>					
Treatment Effect (girls)	0.129*	0.159**	0.040	0.111	0.134**
	(0.0673)	(0.0685)	(0.0769)	(0.0710)	(0.0660)
Treatment Effect (boys)	0.155**	0.138**	0.092	0.189**	0.182***
	(0.0672)	(0.0694)	(0.0787)	(0.0835)	(0.0559)
Girls	-0.002	-0.036	0.070	0.070	0.021
	(0.0438)	(0.0490)	(0.0555)	(0.0494)	(0.0475)
P-value (F-Stat)	0.676	0.763	0.479	0.306	0.433
<i>By age</i>					
Treatment Effect (younger child)	0.112*	0.124**	0.061	0.118*	0.138**
	(0.0578)	(0.0611)	(0.0669)	(0.0705)	(0.0575)
Treatment Effect (older child)	0.169**	0.183**	0.067	0.175**	0.161**
	(0.0752)	(0.0757)	(0.0851)	(0.0842)	(0.0709)
Younger child	0.077	-0.091	0.052	0.122	0.233***
	(0.0692)	(0.0806)	(0.0686)	(0.0869)	(0.0763)
P-value (F-Stat)	0.356	0.379	0.923	0.460	0.766
<i>By household wealth</i>					
Treatment Effect (bottom 20%)	0.248***	0.279***	0.192*	0.035	0.117
	(0.0928)	(0.1019)	(0.1055)	(0.1202)	(0.0884)
Treatment Effect (top 80%)	0.121**	0.125**	0.039	0.167**	0.158***
	(0.0581)	(0.0585)	(0.0666)	(0.0676)	(0.0548)
Bottom 20%	-0.049	-0.095	-0.084	-0.044	0.071
	(0.0508)	(0.0604)	(0.0600)	(0.0709)	(0.0648)
P-value (F-Stat)	0.058	0.065	0.054	0.217	0.627
<i>By cognitive measures at BL</i>					
Treatment Effect (bottom 20%)	0.147**	0.137**	0.075	0.101	0.169**
	(0.0673)	(0.0642)	(0.0791)	(0.0951)	(0.0832)
Treatment Effect (top 80%)	0.141**	0.152**	0.066	0.167**	0.155***
	(0.0668)	(0.0668)	(0.0766)	(0.0702)	(0.0562)
Bottom 20%	0.039	-0.114**	-0.193***	-0.264***	-0.225***
	(0.0485)	(0.0514)	(0.0530)	(0.0620)	(0.0654)
P-value (F-Stat)	0.931	0.828	0.907	0.505	0.874
Observations	2206	2206	2206	2206	2206

Notes: Table reports estimates of the impact of the *Lively Minds* programme on outcomes of different subgroups of children. These are calculated from estimates of a regression of the outcome on an indicator for whether the child is in the treatment group (treatment dummy), an indicator for whether the child is in a particular subgroup and an interaction between the treatment dummy and the subgroup indicator. The regressions also control for a set of child, household, community and school characteristics measured at baseline that were imbalanced across treatment and control group samples at endline, a set of strata indicators, and an indicator for whether the school is located in Bongo or Tolon. The third row reports the coefficient on the subgroup indicator. The fourth row p-value obtained from an F-test testing that the impact of the programme on VM children is equal to that on non-VM children. All outcomes are continuous scores standardised to have mean 0 and standard deviation 1. Standard errors in parentheses clustered at school level (unit of randomisation). *** p-value < 0.01, ** < 0.05, * < 0.1.

Table 22. Heterogeneity in impacts on socio-emotional domains

	(1)	(2)	(3)	(4)	(5)
	SE (All)	Emotional awareness	Externalising behaviour	Internalising behaviour	Pro-social behaviour
<i>By gender</i>					
Treatment Effect (girls)	0.003 (0.1004)	-0.009 (0.0891)	-0.090 (0.0823)	0.019 (0.0998)	-0.088 (0.1113)
Treatment Effect (boys)	0.201** (0.0980)	0.094 (0.0984)	-0.191** (0.0769)	-0.215** (0.0973)	0.006 (0.1073)
Girls	0.124** (0.0609)	-0.038 (0.0579)	-0.145** (0.0609)	-0.125* (0.0657)	0.111** (0.0533)
P-value (F-Stat)	0.012	0.208	0.207	0.008	0.171
<i>By age</i>					
Treatment Effect (younger child)	0.080 (0.1052)	0.068 (0.0980)	-0.106 (0.0810)	-0.083 (0.1065)	0.009 (0.1038)
Treatment Effect (older child)	0.130 (0.0914)	0.022 (0.0889)	-0.175** (0.0703)	-0.122 (0.0852)	-0.093 (0.1205)
Younger child	-0.012 (0.0848)	-0.050 (0.0823)	-0.049 (0.0808)	0.050 (0.0854)	-0.015 (0.0744)
P-value (F-Stat interaction)	0.498	0.554	0.313	0.619	0.247
<i>By household wealth</i>					
Treatment Effect (bottom 20%)	-0.001 (0.1311)	-0.137 (0.1269)	-0.131 (0.1253)	0.055 (0.1245)	-0.007 (0.1407)
Treatment Effect (top 80%)	0.120 (0.0908)	0.075 (0.0837)	-0.137** (0.0674)	-0.125 (0.0878)	-0.049 (0.1044)
Bottom 20%	-0.016 (0.0750)	0.084 (0.0793)	0.049 (0.0773)	-0.012 (0.0685)	0.034 (0.0719)
P-value (F-Stat interaction)	0.238	0.034	0.956	0.062	0.706
<i>By socio-emotional measure at BL</i>					
Treatment Effect (bottom 20%)	0.161 (0.1095)	0.072 (0.1050)	-0.126 (0.0993)	-0.112 (0.1131)	-0.260* (0.1532)
Treatment Effect (top 80%)	0.089 (0.0981)	0.032 (0.0909)	-0.144* (0.0763)	-0.098 (0.0959)	0.026 (0.1005)
Bottom 20%	-0.123 (0.0890)	-0.104 (0.0738)	0.089 (0.0924)	0.083 (0.0799)	0.235** (0.0901)
P-value (F-Stat)	0.485	0.683	0.868	0.894	0.031
Observations	2206	2206	2206	2206	2206

Note: See note to Table 21.

Table 23. Heterogeneity of impacts on health

	Nutrition		Health (Caregiver reports)		
	(1) MUAC (cm)	(2) Acute malnutrition	(3) Overall health	(4) Morbidity	(5) Chronic Problems
<i>By gender</i>					
Treatment Effect (girls)	0.297*** (0.0955)	-0.095*** (0.0341)	0.044 (0.0911)	0.032 (0.0903)	0.063* (0.0343)
Treatment Effect (boys)	0.131 (0.0857)	-0.055* (0.0334)	0.102 (0.0935)	0.094 (0.0914)	0.044 (0.0332)
Girls	-0.118* (0.0620)	0.017 (0.0266)	0.037 (0.0525)	0.047 (0.0532)	-0.044* (0.0249)
P-value (F-Stat)	0.079	0.281	0.425	0.387	0.604
<i>By age</i>					
Treatment Effect (younger child)	0.213** (0.0977)	-0.052 (0.0370)	0.024 (0.0911)	0.012 (0.0895)	0.055 (0.0338)
Treatment Effect (older child)	0.214** (0.0946)	-0.101*** (0.0361)	0.122 (0.0967)	0.113 (0.0954)	0.052 (0.0336)
Younger child	-0.031 (0.1208)	0.103** (0.0479)	0.074 (0.0865)	0.070 (0.0867)	-0.001 (0.0432)
P-value (F-Stat interaction)	0.990	0.289	0.224	0.211	0.939
<i>By household wealth</i>					
Treatment Effect (bottom 20%)	0.274** (0.1181)	-0.104* (0.0548)	0.088 (0.1202)	0.066 (0.1186)	0.068 (0.0505)
Treatment Effect (top 80%)	0.200** (0.0815)	-0.070** (0.0287)	0.071 (0.0864)	0.062 (0.0850)	0.052* (0.0281)
Bottom 20%	-0.024 (0.0790)	0.007 (0.0377)	0.018 (0.0765)	0.016 (0.0762)	0.033 (0.0335)
P-value (F-Stat interaction)	0.529	0.519	0.865	0.966	0.724
Observations	2205	2205	2206	2206	2206

Notes: See note to Table 21.

9 Discussion and conclusions

In this report we have presented findings from a randomised controlled trial of the *Lively Minds* programme. The aim of the trial was to test the effectiveness of the programme implemented through a scalable training of trainers model. A total of 80 schools participated in the trial which took place in the districts of Tolon and Bongo in the Northern and Upper East regions, respectively. Overall, our findings show that *Lively Minds* has significant positive impacts on multiple key domains of child development, including cognitive, socio-emotional and health, which constitute critical components of school readiness. More specifically, we find significant improvements in numeracy, executive functioning, fine motor skills and a reduction in externalising socio-emotional behaviours such as hyperactivity and conduct problems. We also find a significant reduction in acute malnutrition.

Overall these results suggest that the programme is effective at improving the school readiness of KG age children. The positive effects on cognitive development are particularly strong for children living in the most socio-economically deprived households, while the effects within the socio-emotional domain are driven primarily by improvement in the behaviour of boys. In contrast, there is suggestive evidence that the health benefits of the programme are likely to be bigger for girls than boys. While rigorous comparisons across evaluations are methodologically challenging, a “back-of-the-envelope” comparison to other educational experiments, suggest that the effect sizes (which are between 14% and 16% of a standard deviation) are in the moderate, close to large range. This is all the more notable when taking into account the fact that much of the evidence comes from high/middle income settings and highly controlled experiments. A key feature of *Lively Minds* is that it is a scalable ECCE programme which is being implemented through existing government infrastructure. This magnifies the challenge of the already difficult task of delivering high quality ECCE programmes in deprived, remote rural contexts.

A critical question (in this evaluation as well as in much of the existing evidence on impacts of ECCE programmes) is for how long impacts will be sustained. Although our results are from a short-run follow-up, we find promising evidence of behaviour change among participating caregivers which may reduce chances of medium/long-term fade-out of the observed impacts, as well as generate important spillovers on other kids in the household and mothers in the community who do not participate in the programme. We find significant improvements in caregiver knowledge about issues relating to the process of child development, best practice, and high quality schooling. These changes in knowledge are accompanied by changes in practice – we see significant improvement in the participating caregivers’ teaching style and approach to rewarding good behaviour. We also find higher prevalence of productive activities such as reading and playing with the child taking place in the households of mothers who volunteer for *Lively Minds*.

Analysis of pathways shows suggestive evidence of important differences in the mechanisms through which the programme affects child development. In addition to the direct effect of participation in the play schemes, *Lively Minds* has the potential to impact child outcomes through changing teacher practices, wider school environment and/or parental behaviour. Our sample of schools and teachers is too small for rigorous analysis of programme impacts on teacher behaviour and other school level factors. However, we see compelling evidence that

while participation in the play schemes (and possibly other teacher and school level changes that take place as the result of the programme) drive improvements in cognitive development and health, knowledge and behaviour change among the parents is an important mediator of the effect of *Lively Minds* on socio-emotional development. Consistently with this, the programme is significantly more beneficial for the socio-emotional development of children of participating caregivers, but is equally beneficial for the cognitive development and health of all children, irrespective of the participation status of their caregiver.

Combined, our results confirm the merit of simultaneously targeting the home and the pre-school environment, a key feature of the *Lively Minds* programme. One key take-away from our work is that the home environment remains a critical input for the development of pre-school age children. Policies and programmes that target only the pre-school environment may be overlooking an important pathway to improving children's school readiness. An immediate implication of this is that parental involvement should remain a key feature of the *Lively Minds* programme. Our results suggest that the change in parental behaviour which results from such involvement is important for achieving improvements in key developmental domains; if for example teaching assistants led the play-schemes instead of volunteer mothers, the programme may continue to have positive impacts on cognitive development and health (assuming same quality of implementation), but it would likely be less effective in improving socio-emotional skills.

However, our results also highlight that *only* targeting the home environment, as is typically done in for example early years home stimulation programs, might not be sufficient and that intervening in pre-schools is important. We find that impacts of the *Lively Minds* programme on cognitive development are strongest for children whose mothers decide *not* to volunteer in the Play Schemes. If those mothers would be similarly less likely to participate in other types of early years intervention, such as home visiting programs, then in the absence of any simultaneous pre-school intervention their children would be left out.

This evaluation makes an important contribution to the small body of evidence on effectiveness of ECCE programmes in deprived rural Low Income Country Settings and an even smaller evidence base on programmes implemented in a scalable way. The most recent Lancet Series on Early Childhood Development strongly emphasised the issue that "...Many efforts to promote early childhood development are dependent on non-governmental services, which are frequently limited in scope and inequitable in coverage. Interventions are highly dependent on skilled human resources and, unless built on existing service systems such as health, education, social and child protection, face severe supply-side constraints" (Lancet, 2016). One of the four action points that they put forward building on an in-depth review of the current evidence on ECCE programmes is the integration of early childhood development interventions into existing service delivery platforms. This is one of a handful of rigorous evaluations which shows that this can be done effectively even in a highly deprived context characterised by significant constraints in quality of infrastructure and key service delivery as well as skill levels of available human resources.

More broadly this work has the potential to add to the wider state of knowledge about child development in the critical first few years of life. We know that what happens during these years

has persistent long-lasting impacts on key life outcomes, which makes them a time of great vulnerability but also of opportunity for highly effective intervention. However, many questions remain about the process of child development: how do different capabilities and skills form and what are the key inputs. While it is well established that kids need a diverse range of skills and capabilities to thrive and meet their potential, we know very little about *differences* in how these form and the inputs they require. The suggestive early evidence we have already produced that quality of parenting may be more important for socio-emotional than cognitive development provide a solid motivation for future research investigating differences in key inputs for the formation of these two sets of skills, utilising the rich data we have collected as part of this evaluation.

A key set of questions which the current design does not allow us to address relates to the impact of the programme on teacher and wider school practices, as well as what features of the local context (school, community, circuit, district) are most important for enabling the programme to meet its potential. We know very little about what the key ingredients for successful integration of ECCE programmes into existing government service delivery platforms are, even programmes which have been shown to be effective in efficacy trials. The planned *Lively Minds* scale up provides a unique opportunity to make a substantive contribution to this agenda.

References

- Edwards, A. 1957. The social desirability variable in personality assessment and research. New York: The Dryden Press
- Andrew, A., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. 2018. "Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia." *PLoS Med* 15(4): e1002556.
- Attanasio, O., Rubio-Codina, M., Meghir, C., Varela, N. and Grantham-McGregor, S. 2015. "The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota," *Journal of Human Resources*, University of Wisconsin Press. 50(2): 464–483.
- Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., and Rubio-Codina, M. 2018a. Early Stimulation and Nutrition: the impacts of a scalable intervention. NBER Working Paper. 25059.
- Attanasio, O., Caeyers, B., Cattan, S., Cardona Sosa, L., Krutikova, S., Leighton, L., Masselus, L. and Yakubu, M. 2018. Improving early childhood development in rural Ghana through scalable low-cost community-run play schemes: Baseline Report. IFS Report. Institute for Fiscal Studies: London, UK.
- Attanasio, O., Bernal, R., Peña, X. and Vera-Hernández, M. 2019. "The effects of the transition from home-based childcare to childcare centers on children's health and development in Colombia." *Early Childhood Research Quarterly*, 47: 418–431.
- Berlinski, S., Galiani, S. and Gertler, P. 2009. "The effect of pre-primary education on primary school performance." *Journal of public Economics*, 93(1-2): 219–234.
- Bernal, R. and Fernandez, C. 2013. "Subsidized childcare and child development in Colombia: effects of Hogares Comunitarios de Bienestar as a function of timing and length of exposure." *Social Science & Medicine (1982)*, 97: 241–249.
- Briend, A., Garenne, M., Maire, B., Fontaine, O. and Dieng, K. 1989. "Nutritional Status, age and survival: the muscle mass hypothesis", *Eur J Clin Nutr.*, 43(10): 715–26.
- Briend, A., Maire, B., Fontaine, O. and Garenne, M. 2012. "Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children", *Maternal and Child Nutrition*, 8: p. 130–133.
- Brouwers, S. A., Van de Vijver, F.J.R. and Van Hemert, D.A. 2009, "Variation in Raven's Progressive Matrices scores across time and place", *Learning and Individual Differences*, 19: 330–338.
- Cawley, J., J. Heckman and Vytlačil, E.. 2001. "Three observations on wages and measured cognitive ability." *Labour Economics*, 8(4): 419–442.
- Fiorentino, M., P. Sophonneary, A. Laillou, S. Whitney, R. deGroot, M. Perignon, K. Kuong, J. Berger, and Wieringa F.T. 2016. "Current MUAC Cur-OFFS TO Screen for Acute Malnutrition Need to be Adapted to Gender and Age: The Example of Cambodia", *PloS One*, Feb 3:11(2)

- Frongillo, E., S. Sywulka and Kariger, P. 2003. UNICEF psychosocial care indicators project. Final report to UNICEF. Mimeo, Cornell University.
- Ghana Statistical Service - GSS, Ghana Health Service - GHS, and ICF International. 2015. "Ghana Demographic and Health Survey 2014." GSS, GHS, and ICF International, Rockville, Maryland, USA.
- Heckman, J., Stixrud, J. and Urzua, S. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411–482.
- Hoddinott, J., Behrman, J.R., Maluccio, J.A., Melgar, P., Quisumbing, A.R., Ramirez-Zea, M., Stein, A.D., Yount, K.M. and Martorell, R. 2013. "Adult Consequences of Growth Failure in Early Childhood." *The American Journal of Clinical Nutrition* 98, no. 5: 1170-78.
- Hoosen, N., Davids, E.L., De Vries, P.J. and Shung-King, M. 2018. "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.
- Hossain M, Choudhury N, Abdullah KA, Mondal P, Jackson AA, Walson J, and Ahmed, T. 2017. "Evidence-based approaches to childhood stunting in low and middle income countries: a systematic review", *Archives of disease in childhood*. 2017 Oct 1;102(10):903-9.
- Imai, K., Keele, L., Tingley, D. and Yamamoto, T. 2011. "Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies", *Am Polit Sci Rev*. 105(4):765–89.
- Kariger, P., Frongillo, E. A., Engle, P., Britto, P.M.R., Sywulka, S.M. and Menon, P. 2012. "Indicators of Family Care for Development for Use in Multicountry Surveys." *Journal of Health, Population and Nutrition* 30 (4): 472–86.
- Kraft, M. 2018. "Interpreting Effect Sizes of Education Interventions". Working Paper. Available on: https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf
- Osei, G. M. 2006. "Teachers in Ghana: Issues of training, remuneration and effectiveness". *International Journal of Educational Development*, 26(1), 38–51.
- Pisani, L., Borisova, I. and Dowd, A. J. (2015). International Development and Early Learning Assessment Technical Working Paper. Save the Children. Washington, DC.
- Rosenberg, M. 1965. "Society and the adolescent self-image". Princeton, NJ: Princeton University Press.
- Rosero, J. and Oosterbeek, H. 2011. "Trade-offs between different early childhood interventions: Evidence from Ecuador."
- SABRE Education. 2014. The Need – Kindergarten in Ghana <http://sabre.education/programmes/the-need/>
- Van der Gaag, J. 2010. "From child development to human development." In *Development and Prevention of Behaviour Problems*. 231–242. Psychology Press.

Vella, V., Tomkins, A., Ndiku, J., Marshal, T. and Cortinovis, I. 1994. “Anthropometry as a predictor for mortality among Ugandan children, allowing for socio-economic variables”, *Eur. J. Clin. Nutr.*, 48(3): 189-97.

Wolf, S. and McCoy, D.C. 2019. “Household Socioeconomic Status and Parental Investments: Direct and Indirect Relations With School Readiness in Ghana.” *Child Development*.

Wolf, S., J. Lawrence Aber, J. R. Behrman and Tsinigo, E. 2017. Developing and Testing Supply- and Demand-side Interventions to Improve Kindergarten Educational Quality in Ghana, SIEF Final Report. December 2017.

Wolf, S., J. Lawrence Aber, J. R. Behrman and Tsinigo, E. 2019. Experimental Impacts of the “Quality Preschool for Ghana” Interventions on Teacher Professional Well-being, Classroom Quality, and Children’s School Readiness, *Journal of Research on Educational Effectiveness*, 12:1, 10-37

World Health Organisation, 1994. A user’s guide to the Self-Reporting Questionnaire (SRQ), compiled by M. Beusenbergh and J.Orley, <https://apps.who.int/iris/handle/10665/61113>

Appendix A Study protocol

Title: Improving early childhood development in rural Ghana through scalable low-cost community run play schemes

Principal Investigators

Professor Orazio Attanasio (Department of Economics, UCL and co-director of the Centre for Evaluation of Development Policies) –

Contact: o.attanasio@ucl.ac.uk

Dr Sonya Krutikova – (Associate Director, IFS and co-director of the Centre for Evaluation of Development Policies)

Contact: sonya_k@ifs.org.uk

1. Background

Early childhood care and education (ECCE) is critical to a child's development and their success in adult life. Children who receive quality ECCE are proven to be healthier, do better and stay longer in school, and have better economic trajectories in adult life. But in the remote rural communities in northern Ghana, where most families live on less than US\$2 per day, children do not receive these vital opportunities. Although Ghana has relatively advanced ECCE policies and has introduced two compulsory years of Kindergarten (KG) into the primary education system (for ages 4-6), two barriers to ECCE persist. First, the quality of KG is low and marred by a lack of trained teachers, large class sizes, lack of play-based resources, teacher absenteeism and rote-based teaching. The second barrier to ECCE is the low level of maternal education, information and aspiration in deprived rural communities. The Ghana Demographic and Health Survey 2014 reports a median education attainment for women of 0.0 years (Northern) and 2.9 years (Upper East) compared to a national average of 7.2 years. Deprived from education and career opportunities, these women are insufficiently aware of the important role that they themselves can play as parents to fulfill their child's development and career potential, by offering stimulation and nurture in a safe home environment.

Lively Minds' innovation overcomes both of these challenges by training and empowering KG teachers and caregivers (usually mothers, so henceforth we refer to them as 'mothers') in highly deprived communities in Ghana to set up and run educational play schemes using local materials. This behaviour-change project incorporates health and hygiene activities to improve community health. To achieve sustainability, scalability and cost-effectiveness, play schemes are mainstreamed into and are designed to strengthen the Government KG system.

Initial pilot testing of a pathway to scale, delivered through the Ghana Education Service using a training of trainers approach, showed high rates of compliance and suggestive evidence of wide-ranging positive outcomes for teachers, mothers and children. The Institute for Fiscal Studies (IFS) will design and implement a randomised controlled trial to evaluate the impact of the

Lively Minds programme implemented through the training-of-trainer approach on the targeted children, their siblings and caregivers, volunteer mothers who run the play-schemes and teachers who train the volunteer mothers. The evaluation will also assess the key mechanisms for effective implementation of the programme contributing to the development of optimum strategies for scaling the programme. The evidence provided by this evaluation will be crucial for determining whether there is value in mainstreaming the programme across Ghana and replicating it in other countries.

2. Hypotheses

The intervention is intended to have the following final impacts:

1. Positive impact on child physical, cognitive and socio-emotional development and hygiene practices;

This is achieved through a combination of the following intermediate outcomes:

2. Direct impact from child's participation in the Play Schemes;
3. Indirect impact through improvements in Mothers' stimulation and care practices (through acquiring knowledge in the training sessions and interaction with Volunteer Mothers);
4. Indirect impact through improvements in Mothers' psychological wellbeing;
5. Indirect impact through a change in resource allocation (time and material) within the households by primary caregivers having more say;
6. Indirect impact through increased knowledge and awareness among Kindergarten Teachers on good stimulation and hygiene practices (through training) and reduced Teacher absenteeism (through higher motivation and improved accountability as a result of mothers' enhanced pre-school engagement);
7. Indirect impact on children who have not participated in the Play Scheme through interaction with siblings who have;

3. Study Design

3.1. The intervention

The intervention focuses on unlocking the potential of caregivers, both mothers and teachers, training and empowering them with the knowledge, skills and confidence to run educational Play Schemes in kindergarten classes and provide better care and stimulation at home, using local materials. The intervention will be carried out by Lively Minds, an award winning organisation that has been running the programme in rural Ghana (as well as Uganda) for 8 years. The structure of the program is as follows:

1. Kindergarten teachers trained at centralised workshops
 - a. There are ten practical and participatory sessions, which cover the importance of education and play, classroom management, how to use and make games, and how to train Mothers.
2. Teachers train 30 Mothers in their community.

- a. Training includes two community meetings and nine participatory workshops. It is designed for women who are illiterate and have never been to school. Content includes how to make and play games, child-friendly teaching, and health practices. The syllabus uses behaviour-change and play-based approaches to transform mindsets, gain buy-in and volunteerism. Teachers are supervised and supported by high performing Kindergarten Teachers from schools with existing Schemes.
3. Play Schemes run
 - a. A different group of 7 Mothers come into kindergarten each day of the week for an hour. 6 Mothers teach 30 kindergarten children indoors (1:5 parent child ratio). The remaining children and Mothers participate in outdoor play. The teachers supervise. Children learn by playing with a variety of games that strengthen six different skillsets (counting/numeracy; matching/sorting; imagination and creativity; reading/books; sensory awareness; and physical education). These crosscutting skills develop executive functions, providing the foundation for learning. Teaching uses discovery and playbased methods, rather than rote method which is the norm in school.
4. Health and hygiene activities are incorporated
 - a. Children have to handwash with soap before using the Scheme, sensitising them to this vital practice. Mothers are also taught how to erect simple handwashing devices (tippy-taps) at home. Once the Schemes are running, Mothers and teachers are given regular training on health and parenting topics to improve their childcare.
5. Teachers and Schemes are supported
 - a. Play Schemes are given regular supervisory visits by Lively Minds staff and GES officials to quality control. Regular ``top-up" training sessions are held for teachers where they discuss problems, share successes and also are trained to provide the Mothers with monthly skills workshops.
6. Mothers are supported
 - a. Mothers are given monthly workshops on parenting and health topics and life skills by Teachers (topics include nutrition, hygiene, child rights, play, communication, malaria prevention, financial awareness, self-esteem, inclusive education). This increases awareness on a variety of childcare and public health issues, reinforces new behaviours, and is a powerful incentive to keep the Mothers committed to volunteering.
7. Sustainability and scalability
 - a. District Education officials are involved in the mobilisation and training of schools. They monitor the Schemes and supervise the teachers and schools as part of their normal supervisory duties. High performing teachers and officials are trained to participate in the training and support of new cadres of teachers. Play Scheme Committees are established in each community.

3.2. Study design

The study is a cluster-randomised controlled trial, with the school as the unit of randomization. The trial will take place across two districts in rural Northern Ghana; Bongo district, Upper East region and Tolon district, Northern region. Across these two regions, 80 schools will be selected to be part of the study. Within these schools, 40 will be randomly allocated to receive the intervention, and 40 allocated to the control group, who will receive the intervention at the end of the study. Randomisation will stratify using circuit (a geographical cluster of on average 6 schools falling under one supervisor from the Ghana Education Service), and school size (defined as high or low, based on whether the school has above or below the median number of total Kindergarten children). Randomisation will be conducted using Stata 14.

We will conduct two main rounds of data-collection (baseline and endline) following (a) 2,400 pre-school aged children (30 per community) aged 4-5 in September 2017 (referred as ‘target children’ in our study); (b) their primary caregivers; (c) their oldest younger sibling (less than 3 years of age); (d) their youngest older sibling and in the 6-10 age range; (e) their main KG teacher.

Panel data collected at baseline and endline will include:

Community Survey, administered to well-informed community leaders, will capture information about the location and key characteristics of the community (such as population, infrastructure and availability of public services).

Household Survey, administered to the head or most knowledgeable person in the household, will capture basic information on all household members (age, gender, marital status, disability and education) with more detailed education data collected for individuals under the age of 19. We will further inquire about amenities, wealth and expenditure as well as shocks experienced.

Primary caregiver survey, administered to the person that spent most of the time caring for the child in the last 6 months when the child was not in school (usually mother of the target child), will have two groups of questions: 1) caregiver characteristics, well-being knowledge, attitudes, expectations and beliefs around care-giving and the development of her/his children. 2) Information about younger sibling and target child (e.g. development, parent-child interactions, care-giving practices, health, food consumption).

Where possible we will use well-established instruments previously implemented in Ghana or comparable contexts. These include the Family Care Indicators (Frongillo et al. 2013) to measure the level of stimulation in the home; SRQ-20 (Beusenberg and Orley, 1994) to measure caregiver mental health and the Rosenberg self-esteem scale (Rosenberg, 1965); Strengths and Difficulties Questionnaire (Goodman, 1997) and subscales from the Early Childhood Behaviour Questionnaire (Putnam et al., 2006) as measures of caregiver reported target child development; Caregiver Reported Early Childhood Development Index (McCoy et al., 2017) to assess the younger sibling’s development. (See Annex 1 for References)

Target child Assessment will be conducted using International Development and Early Learning Assessment (IDELA) tool developed by 'Save the Children' (Save the Children, 2015) for assessment of target children and Raven's Progressive Matrices for assessment of the youngest older sibling. In addition, we will collect child anthropometric measures (arm circumference only) during the endline survey.

Pre-school Teacher Survey will gather information on school infrastructure, teacher knowledge about child development, teaching practices, routines and quality of care in the class, as well as teacher well-being. To assess short-term impacts, we will also collect data on child development outcomes (using a reduced version of the IDELA test) at midline, 6 months into the project, on a sub-sample of an average of 10 study children in each community.

All surveys will be administered by experienced interviewers hired by our in-country collaborator, Innovations for Poverty Action (IPA) using tablets on the SurveyCTO platform who have specific experience in evaluation of programmes to improve quality of early childhood care provision in pre-schools in Ghana.

3.3. Study Population

The study sample will consist of 2400 children between 3 and 5 years of age, living in the two study districts.

This sample will be taken from an initial sample frame obtained through a census survey. The census survey covers the closest households (up to 150 per school) to each of the 80 study schools. From this sample frame, eligible children were defined as those meeting two criteria: (i) aged between 3 and 5 years as of the start of the school term on the 11th September 2017, (ii) reported by their primary caregivers to be either currently attending or intending to attend in the first coming academic year, one of the 80 study schools. From this sample of children, a random sample of 2400 will be drawn, with an equal number across schools (where possible) and only including a maximum of one child per household. For each child their primary caregiver, household head, and older and younger sibling where possible will also be surveyed.

4. Outcomes of interest

4.1. Primary outcomes

The primary outcome of interest is the developmental outcomes of the target child. This will be measured through use of the IDELA tool (Pisani, Borisova & Dowd (2015)). This provides measures of development along 5 core domains; emergent numeracy, emergent literacy, socio-emotional skills, motor skills, and executive function. This will be scored in two ways; (i) following the guidance from Save the Children, by computing a simple % right score on each domain, (ii) scoring each item using Item Response Theory (IRT) and aggregating the domains through the use of exploratory factor analysis.

4.2. Secondary outcomes

Health impacts: As well as primary impacts on target child cognitive and socio-emotional development, we will also assess impacts on child's health, by using data on incidence of

diarrhea, fever and respiratory infections using the definitions of the WHO as measures of morbidity. We will also measure arm circumference. We will also construct a hygiene knowledge score based on child's responses to questions such as what are good times to wash your hands, what material is needed to wash hands and what are reasons for why washing hands is important.

As well as impacts upon target children, we also intend to estimate the effects of the intervention on primary caregivers and siblings of children exposed to the program. The outcome measures used in each case are as follows:

Younger siblings: The primary outcome measure for younger siblings will be the CREDI short form (McCoy et al (2017)). This is a maternal report measure of the overall development of the child.

Older siblings: The primary outcomes measure for older siblings is the Ravens test (Raven (1936)), a widely used measure of fluid intelligence. In addition, we will measure executive function through forwards and backwards digit span, and literacy and numeracy through adapting a previous test used in the Ghanaian context.

Primary caregivers: We will collect information on the mother's knowledge of stimulation, and her beliefs regarding the importance of these for children's development. To test knowledge, we will rely on a selection of items from the Knowledge of Infant Development (KIDI). We will measure psychological outcomes of primary caregivers through the use of two well-established scales; the SRQ-20 measure of depression (Goodman (1997)), and the Rosenberg measure of Self-esteem (Rosenberg (1965)). These will be scored as a simple aggregate of the answers on each question.

Home environment: The presence of toys and learning materials in the house will be assessed together with parental involvement with the child, the child's routines and organisation of the child's time inside and outside the family house. This will be assessed using the Family Care Indicators, developed by UNICEF.

Teachers: Although limited in statistical power because of relatively small sample size, we will also assess changes in outcomes of *teachers*, using an instrument developed by (Aber, Berman and Wolf (2017)) for use in Ghana. This includes a variety of measures including teacher practices, burnout, job satisfaction amongst other things. The SRQ-20 will also be assessed on teachers.

4.3. Intervention mediators

In order to understand the mechanisms underlying the effectiveness of the intervention, we will investigate heterogeneity in the effects of the program along a number of dimensions. These will include; socio-economic status of the household, characteristics of primary caregiver (including social networks and whether they were a volunteer mother), quality of the school, characteristics of the teacher, characteristics of the child (gender, age, initial developmental outcomes, schooling).

5. Analysis Plan

5.1. *Primary outcomes*

Analysis primary outcomes will be conducted through estimating equation (1) by ordinary least squares (OLS) regressions in Stata 14.

$$y_{ist} = \alpha + \delta y_{is0} + \beta treat_s + \gamma X_{is0} + \sigma strata_s + \epsilon_{ist} \quad (1)$$

Where y_{ist} is one of the above-mentioned outcomes for child i in school s , at time t (either midline or endline). y_{is0} is the same outcome at baseline, $treat_s$ is an indicator for whether the child's school is a treatment school, X_{is0} is a vector of child/household/school level control variables at baseline; such as gender, age and household socio-economic status. Vector X will include variables for which we observe imbalances at baseline, if any, in addition to other baseline variables with high explanatory power to improve precision. $strata_s$ is an indicator for the randomization strata that the school is in. We will cluster standard errors at the level of the school. In this framework, coefficient β gives the causal effect of the intervention on a given outcome. We will also report results from a standard t-test, to determine whether the effect is statistically significant.

5.2. *Secondary outcomes*

Analysis of secondary outcomes will be conducted in an analogous way to the above; with secondary outcomes in place of y_{ist} and i denoting child, sibling, primary caregiver or teacher.

5.3. *Mediators*

The effect of mediators will be assessed through estimation of equation (2) through OLS:

$$y_{ist} = \alpha + \delta y_{is0} + \beta treat_s + \omega m_{ist} + \tau m_{ist} * treat_s + \gamma X_{is0} + \sigma strata_s + \epsilon_{ist} \quad (2)$$

Where m_{ist} is a given mediator. In this framework the coefficient on τ estimates the differences in the impact of the intervention on a given outcome. For example if m is an indicator for gender, τ estimates the difference in the effect of the intervention for boys and girls.

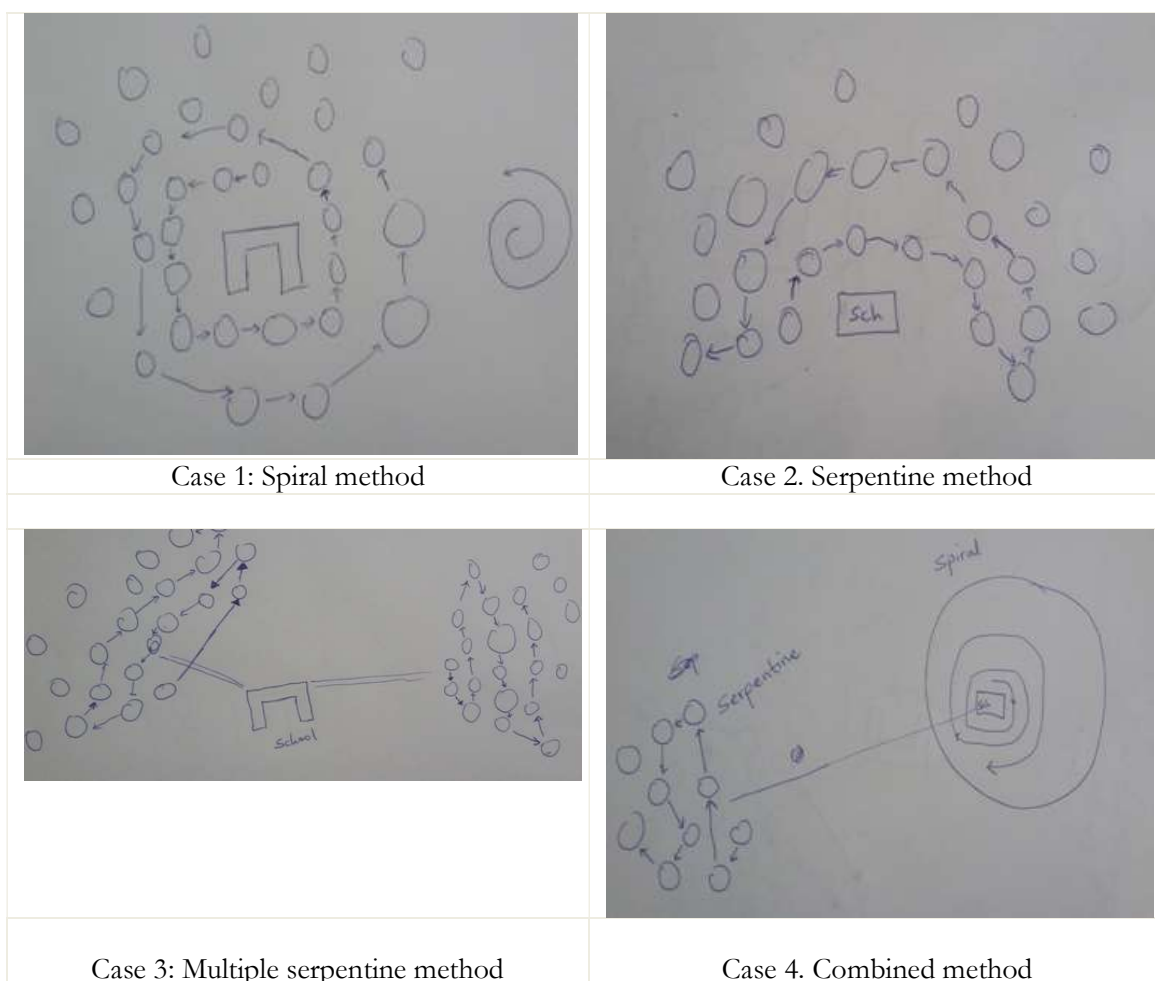
6. Ethical Considerations

Ethical clearance has been obtained from the UCL Institutional Review Board, Innovations for Poverty Action (IPA), and the Ghana Health Services Ethics Review Board.

Appendix B Census sampling strategy

To enumerate the households, one of four different approaches was used, depending on placement of a school within a community (see Figure 1). When the school was centrally located and surrounded by households, the spiral method (Case 1) was followed. In cases where households were located towards one side of the school, the serpentine method (Case 2) was implemented. Similarly, this method was followed when the school was in the middle of two different communities (Case 3), making it necessary to enumerate households in both communities using this approach in each of them. Finally, if after conducting either the spiral or the serpentine method, the number of households per school was not enough, then, the closest households in the next community were enumerated using the corresponding method, leading to a combination of methods (Case 4).

Figure 1. Census sampling techniques



Appendix C Data collection procedures

Survey project team

The survey project team was managed by the Research Manager (RM) at IPA-Ghana. The Survey & Data Coordinator (SDM) provided support on programming of the survey instruments. The SDM supported one intern with programming HFC do-files and data cleaning. The Research Associate (RA) oversaw Recruitment, Training and Project Management. Two Field Managers (FM) – one for each district - were in charge of managing the four Field Supervisors (FS) – two for each district - and handling field logistics. The field supervisors in turn each supervised three survey teams of five enumerators each. Each survey team was headed by a Team Leader whose task, in addition to surveying, was to report to the FS, to schedule appointments with households and to scrutinise survey work done by the other enumerators in his or her team.

Field Supervisors (FS), Team Leaders (TL) and Surveyors were all recruited through a job advert, which was posted on www.jobsinghana.com. Different application forms existed for the different positions. Candidates needed to have obtained at least the Higher National Diploma (NHD). Selection processes continued during the training stage. Suitable candidates were invited to training. In order to create a buffer, 30% more people than needed were invited to attend training. As such, immediate replacement would be possible in case one of the enumerators was unable to continue their work after the start of fieldwork. Furthermore, the buffer allowed for a selection of people based on their performance during training.

Field team training

Surveyors were divided into two groups; one group that focused on child assessments of the Target Child (TC), and the TC's Older- and Younger Siblings (OS and YS). The other group was tasked with the Household (HH), Primary Caregiver (PC), Teacher and Additional Volunteer Mother (aVM) Survey.

Different training sessions were organised for each group of surveyors. Each training session consisted of a classroom training session, focusing on familiarising with the survey instruments, followed by a field training session, which focused on practicing the instrument in a real field context.

The classroom training for the HH, PC, aVM and teacher survey interviewers was organised between 13 August - 18 August 2018. Eighty-eight people were invited to the training, of which 86 attended. Field practice sessions took place on August 20 and 21. The sessions took place in communities that were not part of the study area for this evaluation. Field practices were staggered, first in the Upper East region, and then in the Northern region, to ensure that there were enough time and bandwidth to monitor the practices in each of the districts.

TC Task Assessment Training was held on the 24th and 25th of August 2018. We invited 44 people to the training, of which 40 actually attended. An IPA staff member experienced in the TC task assessment came over from Accra to give the training. The IDELA classroom training happened separately from TC Task Assessment training. The same experienced IDELA trainer

was used as the one used at baseline and midline. IDELA Classroom training was organised on the 6th and 7th of September. The training was attended by 36 people (out of 44 invited). Field Practice for IDELA and TC Task Assessment took place on 12th – 14th of September 2018. Test-Retest was also done during the field practice.

Fieldwork

Table 1. Survey duration

Survey	Average Duration (in minutes)	Duration at Baseline (in minutes)	Duration (as % of Baseline duration)
Household	52.4	77.94	67.2 %
Primary Caregiver	74.1	63.70	116.3 %
IDELA	34.3	40.57	84.5 %
TC Task Assessment	23.1	n/a	n/a
Older Sibling Assessment	15.6	23.49	66.4 %
Younger Sibling Assessment	6.4	8.91	71.8%
Community	58.4	146.74	39.8 %
Teacher	74.8	92.31	81.0 %
School Observation	15.0	n/a	n/a
Additional VM	24.9	n/a	n/a

Endline data were collected in the months of September and November 2018. The HH/PC survey team started first, on the 2nd of September, followed by the child assessment team, which started the 17th of September. The child assessment team started later after the government schools had reopened. Government schools reopened on the 12th of September and we allowed for a week's lag before starting the TC assessments. At endline it was decided that TC assessments were done in schools, in contrast to the assessment at home during the baseline data collection.

Table 1 shows the average duration of each of the survey instruments during actual fieldwork. The table shows some differences between average survey times in the baseline and endline rounds of data collection. These differences can for the largest part be explained by differences in the survey instruments, but it could also at least partially be driven by a learning effect (enumerators being more familiar with the survey instruments).

Quality assurance

Monitoring happened by the Team Leader, Field Supervisor, Field Manager, and Research Associate. Monitoring happened either through accompaniments or spot checks. During an accompaniment, the person monitoring sat through the entire survey, whereas spot checks were shorter. For spot checks, the person monitoring walked in at a random point during the assessment, and witnessed a short part of the questionnaire. The goal for each team leader was to complete at least 5 monitoring activities (a combination of spot checks and accompaniments) per surveyor per week. In practice, this number was not always reached. Reasons for this was that team leaders were often called out to solve problems elsewhere, and that the monitoring forms were not always filled out (even when monitoring was done).

Each district also had two auditors who were engaged in audio auditing and backchecks alternately. These auditors were employed to ensure data quality. Backcheck auditors revisited a randomly selected 10% of the interviewed households, to re-ask a certain number of questions from the questionnaire. The questions that were re-asked are questions of which the answers are not expected to change much, such as the construction material of the house. The data collected by the backcheck auditor was then compared to the original survey data, to see if there were discrepancies. In the case of non-matching answers, the underlying reason for this was investigated.

In addition, at the beginning of each interview, the respondent was asked for permission to record portions of the interview. The large majority (97% in the HH Survey and 95% in the PC Survey) of respondents consented to the recording. The audio auditor listened to these recordings and checked if all procedures were followed and whether questions were asked correctly. A small report was written for every audio file. In case of a suspicious recording, the report was shared with the FS, who then further investigated the matter and made recommendations.

Finally, daily high-frequency checks were run on the HH, PC, IDELA and TC Task Assessment instruments to maintain data quality and rectify errors when data came in.

Fieldwork challenges

Tracking Primary Caregivers. Enumerators initially faced problems tracking PCs. This was a result of the fact that our survey began during the farming season. Most of the PCs were usually engaged in farm activities when enumerators visited. Other reasons include permanent relocation of some households (especially the Fulani households) and temporary rural-urban migration. This problem occurred in both districts but was larger in the Upper East region. Since there is only one farming season in the northern part of Ghana, and alternative livelihood opportunities are rare, migration to the south is very common. Especially for the PC survey, this significantly

affected the number of respondents in our sample. The women sometimes left the children behind or took them along.

Tracking Target Children (TCs). Some of the TCs in our sample were not enrolled in school yet/anymore and child assessors had to trace them to their homes to bring them to the school compound to assess them. This was not always feasible. Moreover, some of the TCs ended up attending different schools than our sample schools (the schools they had intended to attend during our census survey), both within and outside the community. Child assessors had to travel to these schools to bring these TCs to our sample schools to be assessed. To mitigate this problem, we recruited child mobilisers to mobilise TCs so that child assessors could continue with their normal duties. These child mobilisers were enumerators who had been part of the HH/PC team before. The introduction of child mobilisers was extremely useful, as it increased the productivity, as well as the percentage of children we were able to find and assess in each community.

Community accessibility. Most challenges we faced related to accessibility of communities. First of all, the large number of streams and rivers during the rainy season meant that some communities within the Bongo district were not accessible, especially on days of heavy downpour. These communities had to be revisited on different days. Second, some of the communities in the sample are very remote, leading to long travel times. Finally, transportation to the communities was a major problem in this study. About 90% of the Bongo district roads are feeder roads and rocky in nature. Due to the rough nature of the roads, motorbikes easily broke down and regular services were required.

Staff turnover. Halfway through fieldwork, a number of enumerators and child assessors dropped out and could not continue with fieldwork. Some had gotten different job offers, some went back to school whilst others found the job too stressful. Even though some of these dropouts were replaced, it did to some extent affect survey productivity.

Appendix D Study context descriptive

Table 1. Characteristics of communities where study schools are located

	Mean	Standard Deviation
Size		
Community size (people)	2101.26	2174.93
Community size (households)	300.11	323.21
Means of transportation to the capital		
By foot (%)	30	46.11
By bike (%)	35	48
By motorbike (%)	23.75	42.82
Roads		
Unpaved roads (engineered or motorised) (%)	63.75	48.38
Months the road is inaccessible	1.76	2.12
Average time to the capital (min)	68.63	45.07
Sanitation and electricity		
Main source drinking water: Tube well or borehole (%)	53.75	50.17
Main source drinking water: Surfaced water (river/dam/lake) (%)	36.25	48.38
Open defecation is common (%)	77.5	42.02
At least one electricity connection (%)	65	48
Commercial facilities		
Shop to buy children's clothes (%)	25	43.57
Shop to buy children's toys and games (%)	12.5	33.28
Health facilities		
Pharmacy (%)	67.5	47.13
Public hospital/clinic (%)	56.25	49.92
Education facilities		
Private school (%)	13.75	34.65
Public school (%)	88.75	31.8
Preschool available (%)	100	0
Junior High school (%)	66.25	47.58
Senior High school (%)	43.75	49.92
Observations		80

Notes: All statistics obtained from the baseline community survey.

Table 2. Household structure (country and regional average vs. study sample)

	Ghana			Study sample		
	National	Northern	Upper East	All	Tolon	Bongo
Household size	3.83 (2.49)	5.52 (3.45)	4.95 (2.87)	10.23 (6.34)	13.36 (7.03)	6.64 (2.35)
Adult men older than 16	1.00 (0.83)	1.34 (0.95)	1.32 (0.94)	2.53 (2.03)	3.44 (2.21)	1.49 (1.10)
Adult women older than 16	1.18 (0.88)	1.38 (1.02)	1.42 (0.96)	2.33 (1.59)	2.88 (1.83)	1.70 (0.92)
Children age 16 or younger	1.65 (1.77)	2.80 (2.43)	2.20 (1.91)	4.36 (3.55)	6.03 (3.98)	2.45 (1.42)
Children age 5 or younger	0.62 (0.88)	1.13 (1.17)	0.78 (0.94)	1.59 (1.81)	2.44 (2.04)	0.62 (0.72)
Female household head	33.54 (47.21)	11.86 (32.33)	25.23 (43.44)	17.32 (37.85)	2.04 (14.13)	34.82 (47.66)
Education people > 19 yrs						
No education (%)	21.76 (35.15)	60.80 (40.12)	51.90 (38.39)	73.42 (28.73)	83.50 (19.44)	61.88 (33.00)
Primary education (%)	12.93 (27.46)	8.16 (20.33)	15.80 (26.57)	12.10 (21.87)	5.03 (11.30)	20.19 (27.52)
Junior secondary school (%)	24.01 (35.65)	7.86 (20.09)	11.61 (22.85)	5.76 (13.51)	3.50 (8.90)	8.34 (16.98)
Senior high school (%)	12.73 (26.66)	9.22 (21.99)	8.42 (19.76)	6.02 (13.03)	5.27 (10.68)	6.88 (15.24)
Higher education (%)	11.16 (27.82)	10.29 (26.82)	7.92 (23.21)	0.50 (3.92)	0.33 (2.75)	0.68 (4.93)
Observations	26324	3816	2773	2206	1178	1028

Notes: National and regional statistics are obtained from the 2014 Ghana Demographic and Health Survey data (DHS) and from the 2017 Ghana Maternal Health Survey (GMHS). The study sample statistics are obtained from the baseline household survey. Standard deviations are shown in parentheses.

Table 3. Asset ownership (country and regional average vs. study sample)

	Ghana			Study sample		
	National	Northern	Upper East	All	Tolon	Bongo
Mobile phone (%)	90.38 (29.48)	87.78 (32.75)	84.34 (36.35)	89.26 (30.97)	95.59 (20.55)	82.00 (38.43)
Television (%)	63.55 (48.13)	45.45 (49.80)	35.81 (47.95)	27.48 (44.65)	35.94 (48.00)	17.80 (38.27)
Radio (%)	64.11 (47.97)	54.10 (49.84)	51.23 (49.99)	54.08 (49.84)	69.86 (45.90)	35.99 (48.02)
Motorbike/Scooter (%)	11.72 (32.16)	42.79 (49.48)	27.73 (44.78)	32.23 (46.75)	47.54 (49.96)	14.69 (35.42)
Bicycle (%)	21.10 (40.80)	62.43 (48.44)	59.00 (49.19)	78.38 (41.18)	93.72 (24.72)	60.80 (48.84)
Bed (%)	75.99 (42.71)	47.21 (49.93)	29.36 (45.55)	39.08 (48.80)	55.86 (49.68)	19.84 (39.90)
Table and chair (%)	80.18 (39.86)	63.19 (48.23)	48.78 (49.99)	39.26 (48.84)	37.18 (48.35)	41.63 (49.32)
Refrigerator/freezer (%)	40.42 (49.07)	19.42 (39.56)	16.91 (37.49)	3.45 (18.24)	3.90 (19.38)	2.92 (16.84)
Own livestock (%)	36.39 (48.11)	64.99 (47.71)	66.61 (47.17)	89.8 (30.27)	90.24 (29.69)	89.3 (30.93)
Observations	26324	3816	2773	2206	1178	1028

Notes: National and regional statistics are obtained from the 2014 Ghana Demographic and Health Survey data (DHS) and from the 2017 Ghana Maternal Health Survey (GMHS). The study sample statistics are obtained from the baseline household survey. Standard deviations are shown in parentheses.

Table 4. Household religion, main economic activity and shocks

	Mean	Standard Deviation
Polygamous household (%)	24.03	42.73
Religion		
Christian (%)	22.76	41.94
Islam (%)	52.67	49.94
Traditional African (%)	24.07	42.76
Economic Activity		
Main income source: Farming own land (%)	44.2	49.67
Main income source: Waged work (%)	27.79	44.81
Main income source: Profits from small enterprise (%)	18.4	38.76
Main income source: Livestock (%)	3.35	18.01
Has a loan (%)	34.72	47.62
Has savings (%)	51.9	49.98
Income from land used for crops/agriculture wage	65.59	47.52
Shocks experienced since pregnancy of TC (%)		
Fire	5.72	23.23
Severe flood	17.89	38.34
Severe drought	33.97	47.37
Decrease/change in food availability	33.02	47.04
Livestock died	39.78	48.96
Crops failed	38.31	48.63
Livestock stolen	12.17	32.7
Crops stolen	3.23	17.67
Death/reduction in household members	18.05	38.47
Job loss/loss source of income/family enterprises	4.86	21.51
Severe illness or injury	25.2	43.43
Victim of crime	1.87	13.54
Divorced or separated	2.14	14.48
Birth/new household member	19.04	39.27
Paying for child's education	15.47	36.17
Moved/migrated/fled	4.52	20.78
Observations		2206

Notes: All statistics are obtained from the baseline household survey. The statistics with regard to shocks measure the % of households that report to have experienced a particular shock since the Primary Caregiver found out she was pregnant of the Target Child.

Table 5. School readiness (IDELA)

	All	N	Non-VM	N	VM	N
Socio-emotional	27.7	2206	27.7	1769	27.7	437
Emergent Numeracy	27.9	2206	27.8	1769	28.5	437
Emergent Literacy	19.5	2206	19.3	1769	20.3	437
Motor skills	29.4	2206	28.9	1769	31.4*	437
Executive Function	36.2	2206	35.8	1769	38.2*	437
School readiness	26.1	2206	25.9	1769	27	437

Notes: The descriptive statistics in this table are obtained from the IDELA baseline dataset and show the proportion of items (%) that the TC scored correctly, on average for a particular developmental domain. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 6. School readiness (IDELA) (individual items)

	All	N	Non-VM	N	VM	N
Socio-emotional						
Self-awareness	58.3	2206	58.4	1769	58	437
Number of friends	31.2	2206	31.2	1769	31.2	437
Emotional awareness	6.45	2206	6.18	1769	7.55	437
Empathy/perspective taking	14.6	2206	14.6	1769	14.6	437
Solving conflict	28	2206	28.2	1769	27.3	437
Emergent Numeracy						
Comparison by size and length	78.9	2206	79	1769	78.6	437
Sorting and classification	24	2206	24.1	1769	23.6	437
Shape identification	29.1	2206	29	1769	29.7	437
Number identification	4.27	2206	4.12	1769	4.84	437
Counting	16.9	2206	16.3	1769	19.0**	437
Addition and Subtraction	29.8	2206	29.5	1769	30.8	437
Puzzle completion	12.5	2206	12.5	1769	12.9	437
Emergent Literacy						
Expressive vocabulary	25.7	2206	25.6	1769	26.1	437
Print awareness	25.8	2206	26	1769	25.2	437
Letter identification	2.31	2206	2.19	1769	2.83	437
First letter sounds	9.61	2206	9.5	1769	10.1	437
Emergent writing	13.5	2206	13.4	1769	14	437
Oral comprehension	40.1	2206	39.3	1769	43.3*	437
Fine Motor skills						
Copying a shape	35.9	2206	34.9	1769	39.7*	437
Drawing a person	26.5	2206	26.4	1769	27.2	437
Folding paper	25.8	2206	25.5	1769	27.2	437
Executive Function						
Short-term memory	54.1	2206	53.8	1769	55.5	437
Inhibitory control	32.4	2206	32.1	1769	33.8	437
Pencil tap	22.2	2206	21.5	1769	25.2**	437

Notes: The descriptive statistics in this table are obtained from the IDELA baseline dataset and show the proportion of TCs (%) that could successfully complete a particular task related to a particular developmental domain. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Figure 1. IDELA scores in perspective

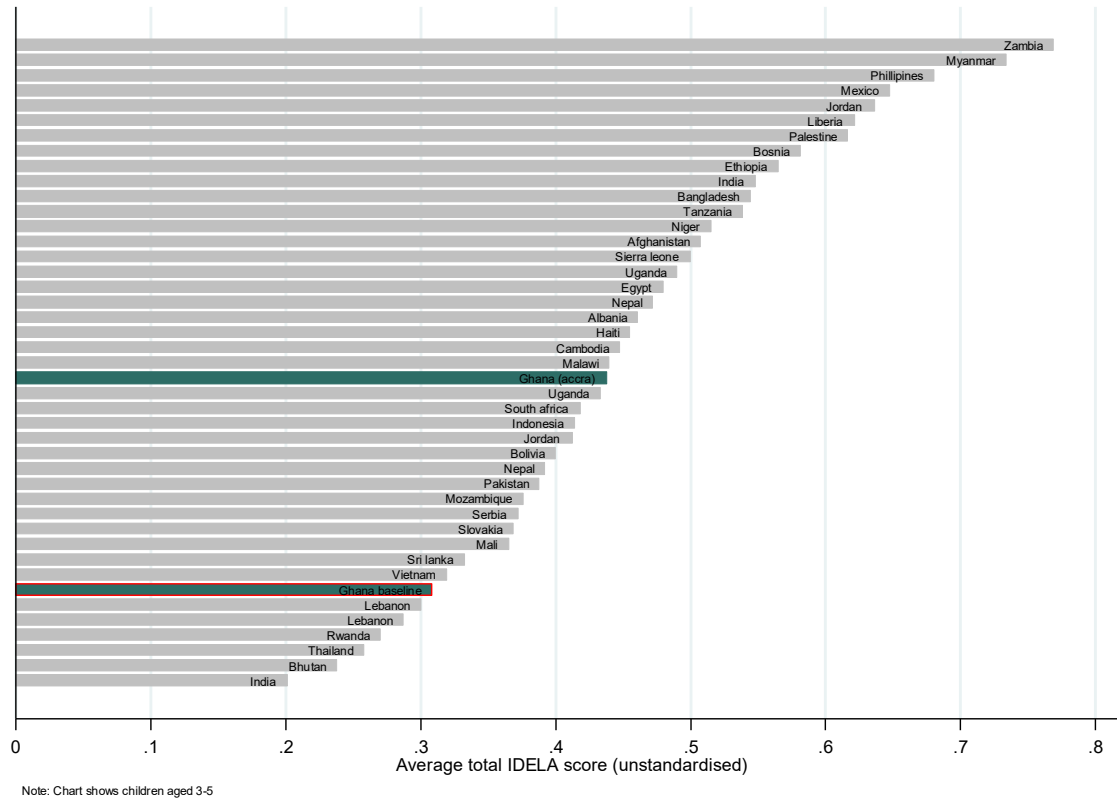


Table 7. Target Child behavioural problems (SDQ)

	All	N	Non-VM	N	VM	N
Individual scales						
Emotional problems scale	3.34	2206	3.34	1769	3.35	437
Conduct problems scale	2.41	2206	2.39	1769	2.48	437
Hyperactivity scale	4.06	2206	4.06	1769	4.04	437
Peer problems scale	3.28	2206	3.3	1769	3.17	437
Pro-social scale	6.6	2206	6.6	1769	6.62	437
Total difficulties score						
SDQ total score	13.1	2206	13.1	1769	13	437
Combined scales						
Externalising scores	6.47	2206	6.45	1769	6.52	437
Internalising scores	6.62	2206	6.64	1769	6.52	437

Notes: The descriptive statistics in this table are obtained from the Strengths and Difficulties Questionnaire (SDQ) baseline dataset and show the average SDQ core, by scale and in the aggregate. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 8. Primary Caregiver WASH practice

	All	N	Non-VM	N	VM	N
Number of times to wash her hands	1.96	2206	1.95	1769	2.02	437
Water is needed to wash her hands (\%)	86.6	2189	86.1	1761	88.8	428
Soap is needed to wash her hands (\%)	98.6	2198	98.4	1762	99.5	436
Water and soap are needed to wash her hands (\%)	84.7	2206	84.2	1769	86.5	437
One example for why hand washing is important (\%)	98.5	2201	98.4	1767	99.1	434

Notes: The descriptive statistics in this table are obtained from the baseline Primary Caregiver survey. The first column presents these statistics for the average PC in the overall sample. Columns 3 and 5 show the average scores for the sample of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between VMs relative to non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 9. Target Child WASH practice

	All	N	Non-VM	N	VM	N
Number of times to wash hands	0.7	2206	0.68	1769	0.77**	437
Water is needed to wash hands (%)	44.5	2206	44.1	1769	46.2	437
Soap is needed to wash hands (%)	55.2	2206	54.9	1769	56.3	437
One example for why hand washing is important (%)	11.3	2206	10.9	1769	12.8	437

Notes: The descriptive statistics in this table are obtained from the baseline IDELA survey. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *, ** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 10. Target child health status

	All	N	Non-VM	N	VM	N
Background						
Age in months	56.3	2206	56.3	1769	55.9	437
Male (%)	49.3	2206	49.3	1769	49	437
Care outside household or pre school (%)	6.12	2206	6.11	1769	6.18	437
Birth outcomes						
Birth weight (Kg)	3.26	2128	3.25	1709	3.31	419
Reported health (out of 5)	3.95	2206	3.95	1769	3.92	437
Health problems in the last 30 days (%)						
3 or more loose or watery stools	31.7	2199	31.4	1764	33.1	435
Blood in his/her stools	9.02	2194	8.86	1761	9.7	433
High fever	34.2	2206	33.4	1769	37.8	437
Cough	54.7	2205	54.7	1768	54.9	437
Very fast or difficult breathing	17.2	2206	16.6	1769	19.5	437
Vomiting everything	30.8	2205	30.4	1768	32.3	437
Stomach pain	50.8	2205	50.5	1768	51.9	437
Serious loss of appetite	43.1	2205	42.5	1768	45.3	437
Skin rashes	18.9	2205	19.1	1768	18.1	437
Sores on feet and legs	16.3	2206	16	1769	17.6	437
Convulsions	4.49	2203	4.13	1766	5.95	437
Unusual tiredness	9.28	2198	9.02	1763	10.3	435
Unconsciousness	1.5	2198	1.47	1763	1.61	435
Extreme lethargy e.g. extremely weak/listless	5.09	2181	5.21	1747	4.61	434
Long-term health problems (%)						
Physical disability	0.86	2204	0.9	1768	0.69	436
Mental disability	0.36	2203	0.34	1766	0.46	437
Skin problems	6.94	2205	6.45	1768	8.92	437
Asthma/respiratory problems	1.82	2202	1.53	1765	2.97	437
Anaemia	2.55	2199	2.61	1763	2.29	436
HIV/AIDS	0.14	2190	0.11	1756	0.23	434
Congenital illness	0.87	2195	0.57	1759	2.06**	436
Stomach ache/abdominal problems	15.2	2204	15	1767	16.2	437

Notes: All statistics are obtained from the baseline Primary Caregiver survey. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 11. Target Child schooling

	All	N	Non-VM	N	VM	N
Schooling						
Ever attended school (%)	77.2	2206	75.9	1769	82.4***	437
Currently in school (%)	78.3	2206	76.8	1769	84.2***	437
Reasons for not attending						
Child is too young (%)	59.3	504	59	427	61	77
Child doesn't want (%)	9.13	504	9.13	427	9.09	77
High school fees (%)	5.16	504	4.68	427	7.79	77
Distance to the school (%)	4.37	504	4.68	427	2.6	77
Road unsafe or inaccessible (%)	0.6	504	0.23	427	2.6	77
Family issue (%)	3.17	504	2.58	427	6.49	77
Uniforms are expensive (%)	2.18	504	2.58	427	0***	77
Books are expensive (%)	1.59	504	1.87	427	0**	77
Poor quality of the school (%)	0.6	504	0.7	427	0*	77
School not needed (%)	0.6	504	0.7	427	0*	77
Child is banned from school (%)	1.19	504	1.17	427	1.3	77
Disability or illness (%)	1.39	504	1.41	427	1.3	77
Child looks after other kids at home (%)	0.6	504	0.7	427	0*	77
Child needs to do domestic tasks (%)	0.99	504	0.94	427	1.3	77
Other (%)	9.13	504	9.6	427	6.49	77

Notes: All statistics obtained from the baseline Primary Caregiver survey. The first column presents these statistics for the average child in the overall sample. Columns 3 and 5 show the average scores for the sample of children of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between children from VM relative to children from non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *, ** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 12. Characteristics of teachers and teaching practices

	Mean	Standard Deviation
Background		
Age	33.05	7.15
Tertiary education (%)	86.09	34.72
Have ECD training (%)	73.51	44.27
Born in this community (%)	9.27	29.10
Live in this community (%)	17.22	37.88
Moved to this community for the job (%)	50.33	50.17
Experience and tenure		
Total years of experience	3.72	4.43
Total years in current KG	2.14	2.24
Allocation of weekly hours		
Working hours	22.09	9.03
Hours for class preparation	7.24	6.38
Teaching practices		
Always praise good behaviour (%)	70.86	45.59
Respond aggressively to bad behaviour (%)	49.67	50.17
Number of classroom activities		
Hygiene, pampering, dancing, etc.	10.48	2.53
Teaching numbers, letters, problem solving, etc.	11.08	5.00
Mental well-being		
Depression score (SRQ)	3.72	3.55
External control	16.66	3.88
Motivation	37.36	4.96
Job satisfaction	89.00	7.50
Burnout	77.56	13.24
Main problems of the school		
Classes are large (%)	58.28	49.47
Lack of financial resources (%)	94.70	22.47
Parents are not actively involved (%)	77.48	41.91
Parents are not sufficiently interested (%)	72.19	44.96
Observations		151

Notes: All statistics obtained from the baseline KG teacher survey.

Table 13. Characteristics of the school

	Mean	Standard Deviation
Average number of pupils per class	57.51	30.16
Desks/chair per pupil	0.12	0.18
Has electricity (%)	10	30.19
Has a blackboard (%)	95	21.93
School has a toilet (%)	38.75	49.03
Books available (%)	47.5	50.25
School has a major safety hazard (%)	45	50.06
Has at least one hand washing facility (%)	32.5	47.13
Observations		80

Notes: All statistics obtained from the baseline school observations made by survey enumerators.

Table 14. Primary Caregivers school involvement

	All	N	Non-VM	N	VM	N
Knows teachers' name (%)	41.2	1723	38.9	1368	50.1***	355
Visited school in the last month (%)	39.2	1724	38	1369	43.9**	355
Member of the PTA executive (%)	14.5	1588	14.2	1257	15.7	331
Ever attended a PTA meeting (%)	78.2	1298	76.6	1008	83.8***	290

Notes: All statistics are obtained from the baseline Primary Caregiver survey. The first column presents these statistics for the average PC in the overall sample. Columns 3 and 5 show the average scores for the sample of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between VMs relative to non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size. Sample restricted to households with children enrolled in KG at time of baseline and only those that responded to the survey two weeks after the start of the academic year (to allow for some time for the PCs to get to know the teacher and the PTA). There is one missing observation for knowing teacher's name. Questions related to PTA only asked if the PC was aware of the existence of a PTA in the school. Whether or not the PC attended a PTA meeting was only asked of PCs that reported to be a member of the PTA executive.

Table 15. PC and household KG school satisfaction

	Mean	Standard Deviation
PC School satisfaction		
Highly satisfied (%)	57.21	49.49
Satisfied (\%)	25.11	43.38
Moderately satisfied (%)	13.83	34.53
Dissatisfied (%)	3.85	19.25
Observations		2206
Household school quality rate		
Good (%)	66.63	47.17
Excellent (%)	9.22	28.95
Reasonably ok (%)	21.74	41.26
Bad or extremely bad (%)	2.41	15.34
Observations		1702

Notes: Statistics in the upper panel are obtained from the baseline PC survey and the statistics in bottom panel are obtained from the baseline household survey. Whereas school satisfaction was asked of all PCs, whether they had already a child in school or whether they were planning to send one to school, school quality rate was only asked of households that had a child that was already attending the pre-school at baseline.

Table 16. Primary Caregiver characteristics

	All	N	Non-VM	N	VM	N
Demographic background						
Age	36.5	2206	36.4	1769	36.6	437
Christian (%)	34.4	2205	34.2	1768	35.2**	437
Muslim (%)	54.9	2205	54.4	1768	57	437
No education (%)	80	2202	79.8	1768	80.9	434
Illiterate (%)	92.2	2206	92.1	1769	92.7	437
Caregiver since birth (%)	86.3	2206	85.5	1769	89.5*	437
Born outside community (%)	69.2	2206	69.2	1769	69.3	437
Born outside district (%)	16.2	2206	16.1	1769	16.7	437
Ravens score (%)	43.2	2206	43.4	1769	42.3	437
PC is biological mother (%)	78.5	2206	77.4	1769	82.8*	437
Mental health						
Raw Self-esteem (Rosenberg) score	18	2206	18	1769	17.8	437
Raw Depression (SRQ-20) Score	8.27	2206	8.2	1769	8.56	437
Depressed (SRQ greater than 7)	0.52	2206	0.51	1769	0.54	437
Raw Rumination Scale Score	10.8	2206	10.8	1769	10.7	437
Physical health (in the last 12 months)						
Weight loss (%)	61.6	2206	61.3	1769	62.9	437
Prolonged fever (%)	44.1	2206	43	1769	48.5	437
Diarrhoea (%)	20.4	2206	19.7	1769	22.9	437
Smoke cigarettes (%)	0.91	2206	1.02	1769	0.46	437
Smoke other form of tobacco (%)	3.13	2206	3.39	1769	2.06*	437
Time use (hours per day)						
Sleeping	8.05	2206	8.06	1769	8	437
Farm work	4.41	2206	4.35	1769	4.67***	437
Domestic tasks	4.38	2206	4.36	1769	4.47	437
Leisure	3.23	2206	3.24	1769	3.2	437
Caring for others	1.95	2206	1.93	1769	2.02	437
Paid work	1.15	2206	1.2	1769	0.94*	437
Playing with children	0.62	2206	0.65	1769	0.51***	437
Collecting children	0.21	2206	0.22	1769	0.18	437

Notes: All statistics are obtained from the baseline Primary Caregiver survey. The first column presents these statistics for the average PC in the overall sample. Columns 3 and 5 show the average scores for the sample of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between VMs relative to non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *, ** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size. A few missing observations for religion and literacy. The SRQ-20 contains a total of 20 items. Guidelines on how the Rosenberg self-esteem scale is scored can be found on <https://www.wvnorton.com/college/psych/psychsci/media/rosenberg.htm>.

Table 17. House environment (Family Caregiver Indicators, FCI)

	All	N	Non-VM	N	VM	N
Learning and play materials						
Number of adult books at home	2.35	2206	2.24	1769	2.82	437
Number of non religious books at home	1.52	2206	1.43	1769	1.9	437
Number of children's books at home	0.029	2206	0.028	1769	0.03	437
Any play materials at home	0.65	2206	0.65	1769	0.64	437
Number of different play materials	1.16	2206	1.14	1769	0.03	437
Number of homemade toys	2.79	2206	2.78	1769	2.85	437
Number of bought toys	0.92	2206	0.9	1769	0.98	437
Activities with the target child (tc)						
Any play activities in last 3 days (tc) (%)	13.2	2206	13.3	1769	12.6	437
Read/look at picture books together (tc)	30.1	655	30.4	520	28.9	135
Tell stories to the child(tc)	20.8	655	21.7	520	17	135
Sing to (tc)	8.09	655	8.85	520	5.19	135
Go out outside the house (tc)	1.68	655	1.92	520	0.74	135
Play together with his/her toys (tc)	1.53	655	1.92	520	0**	135
Make drawings/paintings/writing (tc)	0.61	655	0.77	520	0**	135
Name objects/colours/count objects (tc)	0.15	655	0.19	520	0	135
Activities with the younger sibling (ys)						
Any play activities in last 3 days (ys) (%)	9.81	214	9.64	166	10.4	48
Read/look at picture books together (ys)	5.14	214	4.82	166	6.25	48
Tell stories to the child(ys)	4.67	214	5.42	166	2.08	48
Sing to (ys)	2.34	214	2.41	166	2.08	48
Go out outside the house (ys)	0	214	0	166	0	48
Play together with his/her toys (ys)	0.47	214	0.6	166	0	48
Make drawings/paintings/writing (ys)	0.47	214	0.6	166	0	48
Name objects/colours/count objects (ys)	0	214	0	166	0	48
Activities with the older sibling (os)						
Any play activities in last 3 days (os) (%)	47.4	325	47.8	249	46.1	76
Read/look at picture books together (os)	36.3	325	36.1	249	36.8	76
Tell stories to the child(os)	17.2	325	19.3	249	10.5	76
Sing to (os)	5.85	325	6.43	249	3.95	76
Go out outside the house (os)	1.23	325	0.8	249	2.63	76
Play together with his/her toys (os)	1.23	325	1.2	249	1.32	76
Make drawings/paintings/writing (os)	0.62	325	0.4	249	1.32	76
Name objects/colours/count objects (os)	0	325	0	249	0	76

Notes: Statistics are obtained from the baseline PC survey. First column presents statistics for average PC in overall sample. Columns 3 and 5 show average scores for sample of non-VMs and VMs, respectively. Statistical significance of difference in characteristics between VMs relative to non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size. A few missing observations for religion and literacy.

Table 18. Primary Caregiver beliefs about child development

	All	N	Non-VM	N	VM	N
How strongly do you agree or disagree with (scale 1-4):						
Parents play an important role in child's development	3.64	2206	3.64	1769	3.64	437
Knowing how to read and write is important for children	3.63	2206	3.63	1769	3.61	437
Parents can support child's development at home	3.58	2206	3.59	1769	3.54	437
Children can learn a lot of skills by playing games	3.38	2206	3.39	1769	3.36	437
Parents can engage children in games while working	3.25	2206	3.26	1769	3.23	437
It is important to praise the child when learning	3.54	2206	3.54	1769	3.55	437

Notes: All statistics are obtained from the baseline Primary Caregiver survey. Scale ranges from 4 = strongly agree; 3 = agree; 2 = disagree; 1 = strongly disagree; The first column presents these statistics for the average PC in the overall sample. Columns 3 and 5 show the average scores for the sample of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between VMs relative to non-VMs is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Table 19. Parental practices

	Mean	Standard Deviation
Punishing behaviours (observed) (%)		
Called child dumb, lazy, etc.	7.68	26.64
Explained why the behaviour was wrong	13.53	34.22
Hit the child on the bottom or elsewhere with something	1.28	11.24
Took away privileges, forbade something they like	24.77	43.19
Shook	5.85	23.48
Shouted, yelled or screamed to the child	19.47	39.61
Hit or slapped the child on the face, head, ears	1.55	12.37
Spanked, hit, slapped the child on the bottom with bare hand	3.29	17.85
Punishing behaviours (self-reported) (%)		
Called child dumb, lazy, etc.	24.77	43.19
Explained why the behaviour was wrong	38.57	48.70
Hit the child on the bottom or elsewhere with something	22.67	41.89
Took away privileges, forbade something they like	43.05	49.54
Shook	19.65	39.76
Shouted, yelled or screamed to the child	45.80	49.85
Hit or slapped the child on the face, head, ears	8.04	27.21
Spanked, hit, slapped the child on the bottom with bare hand	25.23	43.45
Encouraging behaviours (observed) (%)		
Praised the child with words such as well done, good work	44.70	49.74
Rewarded the child by giving them something that is desired	1.83	13.40
Showed physical signs of affection e.g. kiss, cuddle	20.93	40.70
Encouraging behaviours (self-reported) (%)		
Praised the child with words such as well done, good work	61.43	48.70
Rewarded the child by giving them something that is desired	26.51	44.16
Showed physical signs of affection e.g. kiss, cuddle	34.64	47.61
Observations	1094	

Notes: All statistics are obtained from the endline Primary Caregiver survey (this information was not captured at baseline). Sample restricted to the control group only, to capture the trend in the absence of the LM intervention. Observed behaviour represents behaviour as observed by the interviewer during the interview, whereas self-reported behaviour is behaviour as reported by the PC herself.

Table 20. Primary Caregiver beliefs about school quality

	Mean	Standard Deviation
Most important feature of a good KG		
Motivated teachers (%)	4.21	20.09
Teachers' attendance (%)	5.25	22.31
Overall reputation of the school (%)	5.25	22.31
Quality of teachers (experience or instruction) (%)	55.37	49.72
Results on tests (%)	4.16	19.98
Focus on play-based learning (%)	3.31	17.9
Good infrastructure (%)	3.93	19.43
Children well behaved and disciplined (%)	7.1	25.69
Good food for children (%)	2.93	16.88
English as language of instruction (%)	1.51	12.22
Observations		2113
How important (on a scale of 1 to 5) is it that KG teachers:		
Know about children's needs as they grow and develop	4.52	0.68
Encourage children to recognize letters or words	4.56	0.59
Encourage children to recognize numbers or shapes	4.54	0.62
Work with families to set individual plans for children	4.36	0.77
Provide materials for play and learning	4.42	0.75
Measure children's development overtime	4.51	0.66
Help children to build relationships with peers and adults	4.45	0.73
Help children learn to control their behaviour	4.49	0.65
Encourage children to express thoughts and feelings	4.42	0.73
Help children resolve conflicts with other children	4.44	0.71
Teach discipline and behaviour guidance that match with parents	4.39	0.79
Consider parents' goals ideas and suggestions when caring for the child	4.27	0.91
Be willing to work with parents about their work schedules	4.09	1.02
Include families in decision-making for child's education	4.2	0.94
Care about the entire family not just the child	3.59	1.35
Connect families to outside community resources	3.68	1.32
Observations		2206

Notes: All statistics are obtained from the baseline Primary Caregiver survey. There are several missing observations in the upper panel, as a result of 'don't know' or refusal responses or coding errors.

Table 21. Presence of Target Child siblings in the same household

	All	N	Non-VM	N	VM	N
Sibling (%)	84.3	2206	83.6	1769	87.4*	437
Siblings older than TC (%)	72.2	2206	70.7	1769	78.3***	437
Siblings older than TC but below 16 (%)	70.5	2206	68.9	1769	76.9***	437
Younger siblings (%)	50	2206	49.9	1769	50.6	437
Number of Siblings younger than TC	0.7	2206	0.71	1769	0.64*	437
Number of Siblings older than TC	2.02	2206	1.98	1769	2.18	437
Number of Siblings older than TC and under 16	1.54	2206	1.51	1769	1.69**	437
Younger sibling interviewed (%)	44.2	2206	43.8	1769	46.2	437

Notes: All statistics are obtained from the baseline household survey. The first column presents these statistics for the average household in the overall sample. Columns 3 and 5 show the average scores for the sample of households of non-VMs and VMs, respectively. Statistical significance of the difference in characteristics between VM households relative to non-VM households is obtained by running a regression of the outcome of interest on VM status, controlling for strata and district fixed effects. *,** and *** represent statistical significance of the difference at the 10, 5 and 1 % significance level, respectively. N = Sample size.

Appendix E Baseline balance between control and treatment groups

With one exception, all tables in this Appendix present average baseline statistics separately for the control group and the treatment group. As explained in Section 4, we do not have baseline data available for half of the endline teacher sample. Therefore, in the final table in this section we present statistics on time invariant endline characteristics of teachers in the endline sample to demonstrate balance of the teacher sample used in our analysis of impacts on teacher outcomes. Asterisks *,** and *** represent statistical significance of the difference in means between the two groups at the 10, 5 and 1 % significance level, respectively. N stands for sample size.

Balance checks relevant for the analysis of impacts on Target Child and Primary Caregiver outcomes

Table 1. Community characteristics

	Control	Treatment	p-value	N
Number of inhabitants	1715.33 (1925.61)	2379.37 (2042.64)	0.08*	2206
Number of households	249.04 (274.88)	332.98 (333.68)	0.28	2206
Daily agricultural wage for men (GHS)	11.72 (3.38)	13.63 (9.22)	0.23	2206
Daily agricultural wage for women (GHS)	11.33 (6.37)	14.33 (13.53)	0.24	2206
Daily non-agricultural wage for men (GHS)	13.29 (5.30)	12.99 (8.90)	0.69	2206
Daily non-agricultural wage for women (GHS)	11.65 (5.42)	11.27 (6.41)	0.61	2206

Table 2. Community facilities

	Control	Treatment	p-value	N
Bank nearby (%)	46.43 (49.90)	44.70 (49.74)	0.81	2206
Bookshop (%)	38.55 (48.69)	26.57 (44.19)	0.38	2206
Microfinance institution (%)	15.38 (36.10)	12.57 (33.16)	0.78	2206
Shop to buy children's clothes (%)	39.38 (48.88)	10.77 (31.02)	0.00***	2206
Shop to buy children's toys and games (%)	14.10 (34.82)	8.08 (27.26)	0.39	2206
Weekly market (%)	52.29 (49.97)	55.57 (49.71)	0.67	2206
General market (%)	43.77 (49.63)	40.66 (49.14)	0.79	2206
Main water source is borehole (%)	50.18 (50.02)	57.81 (49.41)	0.30	2206
Main water source is surface water (%)	39.74 (48.96)	32.05 (46.69)	0.18	2206
Any public toilet in use (%)	34.52 (47.57)	35.28 (47.81)	0.77	2206
Open defecation is common (%)	77.56 (41.74)	80.97 (39.27)	0.54	2206
At least one electricity connection (%)	68.59 (46.44)	64.09 (47.99)	0.87	2206
Family planning facility (%)	20.88 (40.66)	34.47 (47.55)	0.11	2206
Pharmacy (%)	69.14 (46.21)	67.24 (46.96)	0.96	2206
Private clinic (%)	12.91 (33.55)	15.26 (35.98)	0.90	2206
Private hospital (%)	4.30 (20.30)	3.05 (17.21)	0.84	2206
Public clinic (%)	53.02 (49.93)	60.05 (49.00)	0.54	2206
Public hospital nearby (%)	21.61 (41.18)	23.70 (42.54)	0.88	2206
Private school (%)	11.17 (31.52)	16.16 (36.82)	0.22	2206
Public school (%)	91.12 (28.46)	87.52 (33.06)	0.29	2206

Table 3. Community problems experienced in the last four years

	Control	Treatment	p-value	N
Drought (%)	65.29 (47.63)	42.73 (49.49)	0.04**	2206
Erosion or landslide (%)	8.33 (27.65)	5.12 (22.04)	0.54	2206
Flood (%)	51.10 (50.01)	49.01 (50.01)	0.98	2206
Fire (%)	8.15 (27.37)	13.46 (34.15)	0.35	2206
Pests (crops) (%)	47.62 (49.97)	41.83 (49.35)	0.75	2206
Pests (humans) (%)	6.23 (24.18)	9.61 (29.48)	0.87	2206
Pests (animals) (%)	45.97 (49.86)	37.07 (48.32)	0.55	2206
Overflow of river (%)	3.85 (19.24)	0.00 (0.00)	0.10*	2206
Strong wind (%)	71.61 (45.11)	54.58 (49.81)	0.16	2206

Table 4. School characteristics.

	Contro l	Treatment	p-value	N
Average no. Pupils per KG class	61.01 (34.85)	55.31 (21.97)	0.58	2206
Desks per pupil	0.09 (0.15)	0.14 (0.17)	0.06*	2206
Has electricity (%)	5.86 (23.50)	14.00 (34.72)	0.22	2206
Has a blackboard (%)	93.41 (24.83)	92.82 (25.83)	0.56	2206
School has a toilet (%)	33.52 (47.23)	43.00 (49.53)	0.19	2206
Books available (%)	41.30 (49.26)	53.95 (49.87)	0.31	2206
School has a major safety hazard (%)	44.41 (49.71)	45.96 (49.86)	0.96	2206
Preschool teaching in English (%)	25.27 (43.48)	28.37 (45.10)	0.47	2206
Preschool teaching in main local language (%)	67.95 (46.69)	65.71 (47.49)	0.74	2206

Table 5. Teacher characteristics

	Control	Treatment	p-value	N
Male (%)	50.27 (35.31)	49.37 (44.00)	0.89	2206
Religion: Christian (%)	48.49 (47.63)	44.17 (48.54)	0.77	2206
Religion: Islam (%)	50.05 (48.42)	55.83 (48.54)	0.65	2206
Teacher has tertiary education (%)	82.74 (26.60)	90.57 (20.68)	0.23	2206
Working hours at school and working (%)	69.73 (35.72)	77.21 (31.61)	0.31	2206
Total years of experience	3.32 (2.81)	4.17 (3.51)	0.42	2206
Total years in current KG	1.83 (1.46)	2.42 (1.77)	0.06*	2206
Married (%)	78.94 (31.69)	89.86 (20.12)	0.06*	2206
Standardized values of (wealth index)	0.07 (0.74)	0.11 (0.74)	0.71	2206
Teacher born in community (%)	6.73 (17.07)	10.46 (25.62)	0.11	2206
Live in community (%)	16.16 (23.40)	17.10 (28.38)	0.27	2206
Moved community for the job (%)	54.21 (34.92)	38.91 (35.98)	0.05**	2206
At school and working (%)	22.04 (7.88)	22.11 (5.27)	0.90	2206
Hours of preparation	7.25 (4.67)	6.47 (5.02)	0.48	2206
Teacher salary per month (GHS)	938.09 (248.55)	998.76 (315.78)	0.67	2206
How frequently are you paid on time (%)	380.86 (87.51)	376.30 (84.78)	0.65	2206
Do you work outside teaching (%)	4.17 (18.25)	6.33 (20.28)	0.85	2206
Teacher Literacy knowledge	3.86 (1.10)	4.55 (1.04)	0.01**	2206

Table 6. Teaching style

	Control	Treatment	p-value	N
Deals with bad behaviour aggressively (%)	46.98 (38.12)	51.15 (40.16)	0.31	2206
How often do you praise a child on a normal day	2.58 (0.39)	2.78 (0.30)	0.00***	2206
Teachers depression Score	3.93 (2.94)	3.90 (2.08)	0.82	2206
External control raw score	16.59 (2.85)	16.49 (2.80)	0.74	2206
Motivation raw score	37.56 (3.48)	37.73 (4.27)	0.98	2206
Job satisfaction raw score	89.42 (5.82)	88.57 (5.90)	0.42	2206
Burnout raw score	78.15 (10.36)	76.64 (8.19)	0.29	2206
Parents are not sufficiently interested (%)	80.04 (24.50)	66.89 (35.39)	0.02**	2206
Parents are not sufficiently actively involved (%)	82.60 (25.86)	70.66 (35.99)	0.18	2206
Lack of financial resources (%)	96.43 (12.88)	93.18 (20.84)	0.53	2206
Classes are large (%)	60.44 (35.71)	60.82 (35.37)	0.98	2206

Table 7. Household characteristics

	Control	Treatment	p-value	N
Farming own land (%)	43.96 (49.66)	44.43 (49.71)	0.50	2206
Waged work (%)	27.93 (44.89)	27.65 (44.75)	0.54	2206
Profits from small enterprise (%)	19.14 (39.36)	17.68 (38.17)	0.93	2206
Livestock (%)	3.57 (18.57)	3.14 (17.45)	0.88	2206
Wealth Index (PCA factor score) (%)	-5.88 (204.77)	5.24 (214.74)	0.61	2206
Household Size	10.37 (6.71)	10.09 (5.95)	0.63	2206
Number of male adults (18 or older)	2.35 (1.59)	2.32 (1.58)	0.71	2206
Number of female adults (18 or older)	2.52 (2.09)	2.54 (1.97)	1.00	2206
Number of children (16 or under)	4.50 (3.78)	4.22 (3.31)	0.44	2206
Number of children (6 or under)	1.64 (1.88)	1.54 (1.74)	0.55	2206
Christian (%)	21.34 (40.99)	24.15 (42.82)	0.48	2206
Muslim (%)	52.56 (49.96)	52.78 (49.94)	0.81	2206
Traditional African religion (%)	25.64 (43.69)	22.53 (41.80)	0.76	2206
Polygamous household (%)	24.73 (43.16)	23.34 (42.32)	0.63	2206

Table 8. Household shocks

	Control	Treatment	p-value	N
A fire (%)	5.31 (22.44)	6.10 (23.95)	0.23	2206
Severe flood (%)	18.68 (38.99)	17.06 (37.63)	0.87	2206
Severe drought (%)	35.71 (47.94)	32.14 (46.72)	0.46	2206
Decrease in food (%)	33.70 (47.29)	32.23 (46.76)	0.97	2206
Livestock died (%)	40.57 (49.12)	38.87 (48.77)	0.85	2206
Crops failed (%)	38.92 (48.78)	37.61 (48.46)	0.70	2206
Livestock stolen (%)	12.27 (32.83)	12.03 (32.54)	0.85	2206
Crops stolen (%)	3.39 (18.10)	3.05 (17.21)	0.95	2206
Death/reduction in household members (%)	17.49 (38.01)	18.49 (38.84)	0.48	2206
Job loss/loss source of income (%)	4.85 (21.50)	4.85 (21.49)	0.87	2206
Severe illness or injury (%)	25.18 (43.43)	25.13 (43.40)	0.68	2206
Victim of crime (%)	1.83 (13.41)	1.89 (13.61)	0.98	2206
Divorced or separated (%)	2.20 (14.67)	2.06 (14.23)	0.80	2206
Birth/new household member (%)	17.49 (38.01)	20.47 (40.36)	0.24	2206
Paying for child's education (%)	14.38 (35.10)	16.43 (37.07)	0.16	2206
Moved/migrated/fled (%)	4.49 (20.71)	4.49 (20.71)	0.81	2206
Empirical Bayes means for Theta (factor)	0.01 (0.86)	-0.01 (0.85)	0.95	2206

Table 9. Primary caregiver characteristics

	Control	Treatment	p-value	N
Primary caregiver age	37.02 (12.25)	35.96 (11.98)	0.06*	2206
Christian (%)	33.61 (47.26)	35.19 (47.78)	0.65	2206
Muslim (%)	54.85 (49.79)	54.85 (49.79)	0.80	2206
Some education (%)	19.78 (39.85)	20.47 (40.36)	0.67	2206
Illiterate (%)	92.49 (26.37)	92.01 (27.12)	0.80	2206
Primary carer since birth (%)	84.34 (36.36)	88.15 (32.33)	0.07*	2206
Lived in community since birth (%)	29.95 (45.82)	31.60 (46.51)	0.39	2206
Born outside district (%)	15.29 (36.01)	17.15 (37.71)	0.31	2206
Ravens score	43.54 (17.01)	42.85 (15.88)	0.73	2206
Is biological mother (%)	76.83 (42.21)	80.16 (39.90)	0.29	2206

Table 10. Primary caregiver's time use

	Control	Treatment	p-value	N
Sleeping	8.00 (1.19)	8.10 (1.25)	0.52	2206
Care for other household members	4.43 (2.08)	4.39 (2.36)	0.76	2206
Domestic tasks	4.41 (1.55)	4.36 (1.51)	0.55	2206
Leisure time	3.27 (1.47)	3.19 (1.41)	0.26	2206
Tasks on family farm	1.95 (1.45)	1.94 (1.41)	1.00	2206
Paid (remunerated) work or activities outside the household	1.12 (2.15)	1.17 (2.28)	0.50	2206
Play with small children in the household	0.63 (0.96)	0.62 (0.90)	0.66	2206
Taking /picking up a household member to/from school	0.19 (0.52)	0.24 (0.64)	0.16	2206

Table 11. Mental health and long-term problems

	Control	Treatment	p-value	N
Raw Self-esteem (Rosenberg) score	17.99 (3.71)	17.97 (3.53)	0.97	2206
Raw Depression (SRQ-20) Score	8.10 (5.29)	8.43 (5.28)	0.46	2206
Depressed (SRQ greater than 7) (%)	50.82 (50.02)	52.60 (49.95)	0.46	2206
Raw Rumination Scale Score	10.67 (2.86)	10.91 (2.91)	0.64	2206
Have you experienced weight loss in the last 12m? (%)	62.55 (48.99)	60.68 (49.23)	0.46	2206
Have you suffered from a prolonged fever in the last 12m? (%)	43.86 (50.20)	44.34 (50.42)	0.70	2206
Have you had chronic diarrhea in the last 12m? (%)	19.14 (39.82)	21.54 (42.00)	0.45	2206
Do you usually smoke cigarettes? (%)	0.92 (9.53)	0.90 (9.44)	0.72	2206
Do you usually use other form of tobacco? (%)	3.21 (17.62)	3.05 (17.21)	0.62	2206

Table 12. Primary caregiver's community involvement and support

	Control	Treatment	p-value	N
Number of community groups you are an active member	1.10 (1.05)	1.04 (1.07)	0.63	2206
Number of community groups discussing ECD	0.25 (0.52)	0.26 (0.57)	0.82	2206
Number of people you can rely on for emotional support	2.51 (1.90)	2.36 (1.70)	0.21	2206
Number of people you would lend GHS100 or more	2.94 (3.64)	2.90 (2.67)	0.55	2206
Number of people you could borrow GHS100 or more	2.55 (1.83)	2.62 (2.32)	0.62	2206

Table 13. Primary caregiver's social networks.

	Control	Treatment	p-value	N
Social network: Number of women who know	6.31 (2.53)	5.86 (2.66)	0.12	2206
Social network: Number of women whom are close to	3.10 (2.83)	2.85 (2.72)	0.50	2206
Social network: Number of women whom talk about children with	2.91 (2.71)	2.58 (2.64)	0.29	2206
Parents play an important role in child's development	3.62 (0.53)	3.65 (0.51)	0.15	2206
Knowing how to read and write is important for child's development	3.62 (0.52)	3.63 (0.50)	0.36	2206
Parents can support child's development at home	3.59 (0.55)	3.56 (0.55)	1.00	2206
Children can learn a lot of skills by playing games	3.40 (0.67)	3.36 (0.65)	0.89	2206
It is possible to engage children while working	3.29 (0.71)	3.22 (0.76)	0.42	2206
Praising children when he/she tries to do something new is important	3.55 (0.56)	3.54 (0.56)	0.78	2206

Table 14. Target child's characteristics

	Control	Treatment	p-value	N
Age (months)	56.21 (9.51)	56.32 (9.55)	0.81	2206
Female (%)	50.00 (50.02)	48.56 (50.00)	0.73	2206
Ever attended school (%)	78.48 (41.12)	75.85 (42.82)	0.59	2206
Currently attending pre-school (%)	75.92 (42.78)	75.76 (42.87)	0.84	2206
Had a caregiver outside household members (%)	4.85 (21.50)	7.36 (26.13)	0.15	2206

Table 15. Target child's health

	Control	Treatment	p-value	N
3 or more loose or watery stools? (%)	31.14 (46.33)	32.14 (46.72)	0.66	2206
Blood in his/her stools? (%)	8.61 (28.06)	9.34 (29.11)	0.51	2206
High fever (%)	32.05 (46.69)	36.36 (48.12)	0.17	2206
Cough (%)	52.20 (49.97)	57.18 (49.50)	0.02**	2206
Very fast or difficult breathing? (%)	15.75 (36.44)	18.58 (38.91)	0.11	2206
Vomiting everything? (%)	29.40 (45.58)	32.14 (46.72)	0.02**	2206
Stomach pain (%)	47.99 (49.98)	53.50 (49.90)	0.01***	2206
Serious loss of appetite? (%)	42.67 (49.48)	43.45 (49.59)	0.56	2206
Skin rashes? (%)	19.41 (39.57)	18.31 (38.69)	0.33	2206
Sores on feet and legs? (%)	16.48 (37.12)	16.16 (36.82)	0.52	2206
Convulsions? (%)	4.85 (21.50)	4.13 (19.91)	0.86	2206
Unusual tiredness? (%)	8.24 (27.51)	10.23 (30.32)	0.54	2206
Unconsciousness? (%)	1.37 (11.64)	1.62 (12.61)	0.75	2206
Extreme lethargy (e.g. extremely weak/listless)? (%)	4.40 (20.51)	5.66 (23.11)	0.20	2206
Empirical Bayes means for Theta	-0.05 (0.87)	0.05 (0.86)	0.04**	2206

Table16. Target child's long-term problems

	Control	Treatment	p-value	N
Physical disability (%)	0.73 (8.53)	0.99 (9.89)	0.88	2206
Mental disability (%)	0.37 (6.04)	0.36 (5.98)	0.68	2206
Fits/epilepsy/convulsions (%)	1.65 (12.74)	1.53 (12.26)	0.86	2206
Skin problems (%)	6.96 (25.46)	6.91 (25.38)	0.95	2206
Asthma/respiratory problems (%)	1.65 (12.74)	1.97 (13.92)	0.52	2206
Anaemia (%)	2.93 (16.87)	2.15 (14.53)	0.48	2206
HIV/AIDS (%)	0.09 (3.03)	0.18 (4.24)	0.47	2206
Congenital illness (%)	0.82 (9.05)	0.90 (9.44)	0.90	2206
Stomach ache/abdominal problems (%)	14.29 (35.01)	16.16 (36.82)	0.23	2206
Other (specify) (%)	1.01 (9.99)	1.53 (12.26)	0.23	2206

Table 17. Target child's time use

	Control	Treatment	p-value	N
Sleeping	10.06 (1.06)	10.11 (1.02)	0.64	2206
Time spent playing with other hh members	4.61 (1.97)	4.73 (2.09)	0.83	2206
At school/nursery/kg	4.49 (2.50)	4.40 (2.49)	0.84	2206
General leisure time other than playing	3.29 (1.52)	3.23 (1.55)	0.37	2206
Caring for others	0.46 (0.89)	0.41 (0.86)	0.89	2206
Studying at home	0.35 (0.70)	0.41 (0.70)	0.10*	2206
Domestic tasks	0.38 (0.77)	0.35 (0.75)	0.85	2206
Tasks on family farm/cattle herding	0.35 (0.80)	0.37 (0.95)	0.62	2206

Table 18. Target child's hygiene practices (WASH)

	Control	Treatment	p-value	N
Number of different times for hand washing stated	1.98 (0.99)	1.95 (0.99)	0.86	2206
Water is needed (%)	86.90 (33.75)	85.01 (35.71)	0.96	2206
Soap is needed (%)	98.44 (12.39)	98.03 (13.92)	0.95	2206
Example of why hand washing is important (%)	97.89 (14.37)	98.65 (11.53)	0.22	2206
Number of different times for hand washing stated	0.70 (0.78)	0.70 (0.79)	0.86	2206

Balance checks relevant for heterogeneity analysis by VM status

This Appendix shows similar balance tables as the ones in the previous section, but now restricted to the VM sample (those that are PCs in our sample).

Table 19. Community characteristics

	Control	Treatment	p-value	N
Number of inhabitants	1397.48 (1127.52)	2372.59 (2132.34)	0.01***	437
Number of households	197.67 (240.93)	297.32 (309.89)	0.10	437
Daily agricultural wage for men (GHS)	11.75 (3.69)	13.66 (10.06)	0.40	437
Daily agricultural wage for women (GHS)	12.40 (8.15)	15.05 (15.60)	0.51	437
Daily non-agricultural wage for men (GHS)	13.71 (4.97)	13.27 (8.49)	0.91	437
Daily non-agricultural wage for women (GHS)	12.02 (5.40)	11.88 (6.55)	0.61	437

Table 20. Community facilities

	Control	Treatment	p-value	N
Bank nearby (%)	49.46 (50.13)	45.45 (49.89)	0.75	437
Bookshop (%)	43.48 (49.71)	28.06 (45.02)	0.25	437
Microfinance institution (%)	14.13 (34.93)	11.46 (31.92)	0.92	437
Shop to buy children's clothes (%)	44.02 (49.78)	9.49 (29.36)	0.00***	437
Shop to buy children's toys and games (%)	16.85 (37.53)	7.11 (25.76)	0.22	437
Weekly market (%)	59.24 (49.27)	56.52 (49.67)	0.82	437
General market (%)	39.67 (49.06)	38.34 (48.72)	1.00	437
Main water source is borehole (%)	42.39 (49.55)	56.92 (49.62)	0.18	437
Main water source is surface water (%)	46.74 (50.03)	31.23 (46.43)	0.10	437
Any public toilet in use (%)	30.98 (46.37)	35.97 (48.09)	0.56	437
Open defecation is common (%)	73.37 (44.32)	71.54 (45.21)	0.89	437
At least one electricity connection (%)	55.98 (49.78)	53.75 (49.96)	0.99	437
Family planning facility (%)	20.65 (40.59)	37.15 (48.42)	0.04**	437
Pharmacy (%)	72.83 (44.61)	69.96 (45.93)	0.87	437
Private clinic (%)	9.78 (29.79)	14.23 (35.00)	0.74	437
Private hospital (%)	7.61 (26.59)	3.56 (18.56)	0.45	437
Public clinic (%)	53.80 (49.99)	66.80 (47.19)	0.29	437
Public hospital nearby (%)	27.17 (44.61)	23.32 (42.37)	0.78	437
Private school (%)	9.24 (29.04)	13.83 (34.59)	0.10	437
Public school (%)	91.30 (28.25)	84.58 (36.18)	0.16	437

Table 21. Community problems experienced in the last four years

	Control	Treatment	p-value	N
Drought (%)	61.41 (48.81)	47.04 (50.01)	0.27	437
Erosion or landslide (%)	3.26 (17.81)	2.37 (15.25)	0.82	437
Flood (%)	48.91 (50.12)	48.62 (50.08)	0.82	437
Fire (%)	6.52 (24.76)	14.23 (35.00)	0.23	437
Pests (crops) (%)	52.17 (50.09)	42.29 (49.50)	0.46	437
Pests (humans) (%)	7.07 (25.69)	9.88 (29.90)	0.71	437
Pests (animals) (%)	46.74 (50.03)	33.99 (47.46)	0.30	437
Overflow of river (%)	3.80 (19.18)	0.00 (0.00)	0.21	437
Strong wind (%)	72.83 (44.61)	60.08 (49.07)	0.24	437

Table 22. School characteristics.

	Control	Treatment	p-value	N
Average no. Pupils per KG class	61.37 (35.19)	51.82 (21.70)	0.24	437
Desks per pupil	0.08 (0.14)	0.16 (0.19)	0.05**	437
Has electricity (%)	5.98 (23.77)	11.46 (31.92)	0.47	437
Has a blackboard (%)	94.57 (22.73)	89.72 (30.43)	0.92	437
School has a toilet (%)	32.07 (46.80)	41.11 (49.30)	0.34	437
Books available (%)	38.04 (48.68)	52.96 (50.01)	0.41	437
School has a major safety hazard (%)	39.67 (49.06)	45.45 (49.89)	0.34	437
Preschool teaching in English (%)	22.83 (42.09)	24.11 (42.86)	0.90	437
Preschool teaching in main local language (%)	76.09 (42.77)	73.91 (44.00)	0.75	437

Table 23. Teacher characteristics

	Control	Treatment	p-value	N
Male (%)	51.90 (35.59)	47.43 (44.13)	0.81	437
Religion: Christian (%)	44.02 (46.95)	43.08 (48.81)	0.81	437
Religion: Islam (%)	55.43 (47.31)	56.92 (48.81)	0.77	437
Teacher has tertiary education (%)	76.90 (29.97)	90.91 (23.07)	0.10*	437
Working hours at school and working (%)	72.01 (34.94)	79.05 (31.12)	0.43	437
Total years of experience	3.25 (2.20)	3.76 (2.84)	0.54	437
Total years in current KG	2.03 (1.68)	2.31 (1.66)	0.28	437
Married (%)	75.82 (34.27)	89.33 (20.53)	0.05**	437
Standardized values of (wealth index)	0.05 (0.76)	-0.02 (0.79)	0.41	437
Teacher born in community (%)	7.61 (18.01)	12.65 (27.78)	0.27	437
Live in community (%)	13.86 (22.44)	18.77 (29.77)	0.22	437
Moved community for the job (%)	53.26 (38.63)	40.51 (38.69)	0.15	437
At school and working	21.92 (7.44)	21.81 (6.00)	0.62	437
Hours of preparation	7.18 (4.74)	6.36 (4.86)	0.32	437
Teacher salary per month (GHS)	934.16 (258.67)	970.12 (320.78)	0.96	437
How frequently are you paid on time (%)	391.30 (83.99)	382.02 (84.06)	0.63	437
Do you work outside teaching (%)	5.71 (19.77)	5.34 (19.95)	0.71	437
Teacher Literacy knowledge	3.85 (1.16)	4.45 (0.98)	0.06*	437

Table 24. Teaching style

	Control	Treatment	p-value	N
Deals with bad behaviour aggressively (%)	45.11 (41.86)	48.42 (40.55)	0.48	437
How often do you praise a child on a normal day	2.63 (0.38)	2.74 (0.33)	0.16	437
Teachers depression Score	4.26 (3.22)	4.00 (2.03)	0.67	437
External control raw score	17.10 (2.56)	16.50 (2.55)	0.62	437
Motivation raw score	37.34 (3.92)	37.42 (4.65)	0.56	437
Job satisfaction raw score	88.79 (5.87)	88.58 (6.15)	0.62	437
Burnout raw score	78.90 (10.18)	77.88 (7.81)	0.47	437
Parents are not sufficiently interested (%)	82.07 (24.05)	66.40 (38.55)	0.02**	437
Parents are not sufficiently actively involved (%)	85.33 (25.11)	69.17 (36.43)	0.09*	437
Lack of financial resources (%)	96.47 (12.85)	94.07 (19.53)	0.75	437
Classes are large (%)	60.05 (37.25)	50.20 (36.32)	0.22	437

Table 25. Household characteristics

	Control	Treatment	p-value	N
Farming own land (%)	47.83 (50.09)	47.43 (50.03)	0.62	437
Waged work (%)	21.74 (41.36)	24.11 (42.86)	0.73	437
Profits from small enterprise (%)	17.93 (38.47)	17.00 (37.63)	0.90	437
Livestock (%)	4.35 (20.45)	3.56 (18.56)	0.85	437
Wealth Index (PCA factor score) (%)	1.87 (196.11)	-19.65 (200.85)	0.53	437
Household Size	10.48 (6.15)	9.95 (5.65)	0.49	437
Number of male adults (18 or older)	2.24 (1.59)	2.29 (1.51)	0.85	437
Number of female adults (18 or older)	2.55 (1.94)	2.44 (1.87)	0.57	437
Number of children (16 or under)	4.69 (3.50)	4.22 (3.12)	0.34	437
Number of children (6 or under)	1.63 (1.74)	1.52 (1.69)	0.82	437
Christian (%)	13.04 (33.77)	26.48 (44.21)	0.01***	437
Muslim (%)	58.70 (49.37)	55.34 (49.81)	0.66	437
Traditional African religion (%)	28.26 (45.15)	17.39 (37.98)	0.07*	437
Polygamous household (%)	26.09 (44.03)	23.32 (42.37)	0.65	437

Table 26. Household shocks

	Control	Treatment	p-value	N
A fire (%)	4.89 (21.63)	5.14 (22.12)	0.74	437
Severe flood (%)	22.83 (42.09)	17.00 (37.63)	0.34	437
Severe drought (%)	38.59 (48.81)	34.39 (47.59)	0.69	437
Decrease in food (%)	37.50 (48.54)	36.36 (48.20)	1.00	437
Shocks: Livestock died (%)	40.22 (49.17)	39.92 (49.07)	0.60	437
Shocks: Crops failed (%)	42.93 (49.63)	41.50 (49.37)	0.83	437
Shocks: Livestock stolen (%)	8.70 (28.25)	11.86 (32.39)	0.29	437
Shocks: Crops stolen (%)	2.17 (14.62)	3.56 (18.56)	0.12	437
Shocks: Death/reduction in household members (%)	17.93 (38.47)	13.04 (33.74)	0.48	437
Shocks: Job loss/loss source of income (%)	5.98 (23.77)	2.77 (16.43)	0.17	437
Shocks: Severe illness or injury (%)	26.63 (44.32)	23.72 (42.62)	0.58	437
Shocks: Victim of crime (%)	1.09 (10.40)	2.37 (15.25)	0.55	437
Shocks: Divorced or separated (%)	1.63 (12.70)	1.98 (13.95)	0.72	437
Shocks: Birth/new household member (%)	21.20 (40.98)	19.37 (39.60)	0.77	437
Shocks: Paying for child's education (%)	17.39 (38.01)	19.76 (39.90)	0.41	437
Shocks: Moved/migrated/fled (%)	5.98 (23.77)	4.35 (20.43)	0.76	437
Empirical Bayes means for Theta	0.08 (0.88)	0.04 (0.82)	0.74	437

Table 27. Primary caregiver's characteristics

	Control	Treatment	p-value	N
Primary caregiver age	37.12 (10.98)	36.26 (10.52)	0.24	437
Christian (%)	33.70 (47.40)	36.36 (48.20)	0.62	437
Muslim (%)	59.78 (49.17)	54.94 (49.85)	0.56	437
Some education (%)	21.20 (40.98)	18.58 (38.97)	0.66	437
Illiterate (%)	90.76 (29.04)	94.07 (23.66)	0.19	437
Primary carer since birth (%)	88.04 (32.53)	90.51 (29.36)	0.31	437
Lived in community since birth (%)	29.35 (45.66)	31.62 (46.59)	0.44	437
Born outside district (%)	15.76 (36.54)	17.39 (37.98)	0.75	437
Ravens score	42.71 (17.73)	41.96 (15.24)	0.42	437
Is biological mother (%)	83.15 (37.53)	82.61 (37.98)	0.57	437

Table 28. Primary caregiver's time use

	Control	Treatment	p-value	N
Sleeping	7.92 (1.17)	8.06 (1.39)	0.53	437
Care for other household members	4.66 (1.88)	4.67 (2.28)	0.82	437
Domestic tasks	4.52 (1.45)	4.44 (1.46)	0.98	437
Leisure time	3.22 (1.40)	3.18 (1.37)	0.81	437
Tasks on family farm	2.08 (1.46)	1.98 (1.38)	0.83	437
Paid (remunerated) work or activities outside the household	0.95 (1.92)	0.94 (2.08)	0.92	437
Play with small children in the household	0.53 (0.87)	0.51 (0.78)	0.59	437
Taking /picking up a household member to/from school	0.13 (0.38)	0.22 (0.60)	0.05*	437

Table 29. Mental health and long-term problems

	Control	Treatment	p-value	N
Raw Self-esteem (Rosenberg) score	17.93 (3.98)	17.77 (3.39)	0.75	437
Raw Depression (SRQ-20) Score	8.60 (5.39)	8.53 (5.16)	0.95	437
Depressed (SRQ greater than 7) (%)	53.80 (49.99)	54.15 (49.93)	0.93	437
Raw Rumination Scale Score	10.46 (2.69)	10.91 (2.65)	0.24	437
Have you experienced weight loss in the last 12m? (%)	64.67 (49.06)	61.66 (48.72)	0.73	437
Have you suffered from a prolonged fever in the last 12m? (%)	50.00 (50.14)	47.43 (51.59)	0.53	437
Have you had chronic diarrhea in the last 12m? (%)	23.91 (44.03)	22.13 (42.54)	0.97	437
Do you usually smoke cigarettes? (%)	0.54 (7.37)	0.40 (6.29)	0.71	437
Do you usually use other form of tobacco? (%)	1.09 (10.40)	2.77 (16.43)	0.27	437

Table 30. Primary caregiver's community involvement and support

	Control	Treatment	p-value	N
Number of community groups you are an active member	1.33 (0.97)	1.25 (1.26)	0.60	437
Number of community groups discussing ECD	0.33 (0.57)	0.39 (0.75)	0.45	437
Number of people you can rely on for emotional support	2.78 (2.18)	2.28 (1.51)	0.02**	437
Number of people you would lend GHS100 or more	2.90 (1.85)	2.79 (1.87)	0.41	437
Number of people you could borrow GHS100 or more	2.85 (1.79)	2.70 (2.54)	0.52	437

Table 31. Primary caregiver's social networks.

	Control	Treatment	p-value	N
Social network: Number of women who know	7.02 (2.39)	6.13 (2.80)	0.03**	437
Social network: Number of women whom are close to	3.60 (2.98)	3.11 (2.86)	0.45	437
Social network: Number of women whom talk about children with	3.33 (2.94)	3.05 (2.91)	0.89	437
Parents play an important role in child's development	3.60 (0.58)	3.68 (0.50)	0.15	437
Knowing how to read and write is important for child's development	3.55 (0.59)	3.64 (0.50)	0.08*	437
Parents can support child's development at home	3.57 (0.60)	3.52 (0.56)	0.33	437
Children can learn a lot of skills by playing games	3.40 (0.71)	3.34 (0.64)	0.21	437
It is possible to engage children while working	3.30 (0.69)	3.18 (0.75)	0.18	437
Praising children when he/she tries to do something new is important	3.53 (0.63)	3.56 (0.56)	0.74	437

Table 32. Target child's characteristics

	Control	Treatment	p-value	N
Age (months)	55.91 (9.39)	55.91 (9.83)	0.97	437
Female (%)	51.63 (50.11)	47.04 (50.01)	0.15	437
Ever attended school (%)	84.24 (36.54)	81.03 (39.29)	0.48	437
Currently attending pre-school (%)	79.89 (40.19)	82.21 (38.32)	0.50	437
Had a caregiver outside household members (%)	4.89 (21.63)	7.11 (25.76)	0.50	437

Table 33. Target child's health

	Control	Treatment	p-value	N
3 or more loose or watery stools? (%)	33.70 (47.40)	32.41 (46.90)	0.72	437
Blood in his/her stools? (%)	9.24 (29.04)	9.88 (29.90)	0.98	437
High fever (%)	34.24 (47.58)	40.32 (49.15)	0.21	437
Cough (%)	50.54 (50.13)	58.10 (49.44)	0.20	437
Very fast or difficult breathing? (%)	16.85 (37.53)	21.34 (41.05)	0.44	437
Vomiting everything? (%)	27.72 (44.88)	35.57 (47.97)	0.04**	437
Stomach pain (%)	45.11 (49.90)	56.92 (49.62)	0.02**	437
Serious loss of appetite? (%)	44.02 (49.78)	46.25 (49.96)	0.54	437
Skin rashes? (%)	18.48 (38.92)	17.79 (38.32)	0.49	437
Sores on feet and legs? (%)	18.48 (38.92)	17.00 (37.63)	0.60	437
Convulsions? (%)	5.98 (23.77)	5.93 (23.66)	0.98	437
Unusual tiredness? (%)	8.15 (27.44)	11.86 (32.39)	0.29	437
Unconsciousness? (%)	2.72 (16.30)	0.79 (8.87)	0.06*	437
Extreme lethargy (e.g. extremely weak/listless)? (%)	5.43 (22.73)	3.95 (19.52)	0.52	437
Empirical Bayes means for Theta	-0.03 (0.89)	0.12 (0.88)	0.14	437

Table 34. Target child's long-term problems

	Control	Treatment	p-value	N
Physical disability (%)	1.09 (10.40)	0.40 (6.29)	0.37	437
Mental disability (%)	0.54 (7.37)	0.40 (6.29)	0.78	437
Fits/epilepsy/convulsions (%)	1.63 (12.70)	2.37 (15.25)	0.25	437
Skin problems (%)	9.78 (29.79)	8.30 (27.64)	0.38	437
Asthma/respiratory problems (%)	1.09 (10.40)	4.35 (20.43)	0.05**	437
Anaemia (%)	1.63 (12.70)	2.77 (16.43)	0.22	437
HIV/AIDS (%)	0.54 (7.37)	0.00 (0.00)	0.34	437
Congenital illness (%)	1.63 (12.70)	2.37 (15.25)	0.38	437
Stomach ache/abdominal problems (%)	15.22 (36.02)	17.00 (37.63)	0.60	437
Other (specify) (%)	1.09 (10.40)	1.98 (13.95)	0.30	437

Table 35. Target child's time use

	Control	Treatment	p-value	N
Sleeping	9.93 (1.07)	10.07 (1.10)	0.47	437
Time spent playing with other hh members	4.38 (2.02)	4.58 (2.05)	0.57	437
At school/nursery/kg	4.83 (2.23)	4.69 (2.24)	0.87	437
General leisure time other than playing	3.26 (1.55)	3.09 (1.45)	0.34	437
Caring for others	0.40 (0.76)	0.39 (0.87)	0.76	437
Studying at home	0.41 (0.77)	0.48 (0.82)	0.31	437
Domestic tasks	0.41 (0.81)	0.34 (0.73)	0.28	437
Tasks on family farm/cattle herding	0.39 (0.89)	0.36 (0.93)	0.61	437

Table 36. Target child's hygiene practices (WASH)

	Control	Treatment	p-value	N
Number of different times for hand washing stated	1.98 (0.97)	2.05 (0.95)	0.80	437
Water is needed (%)	86.41 (34.36)	87.35 (33.31)	0.48	437
Soap is needed (%)	99.46 (7.37)	99.21 (8.87)	0.79	437
Example of why hand washing is important (%)	97.83 (14.62)	98.81 (10.85)	0.66	437
Number of different times for hand washing stated	0.78 (0.87)	0.76 (0.78)	0.80	437

Balance checks relevant for analysis of impacts on KG teacher outcomes

Table 37. Community characteristics

	Control	Treatment	p-value	N
Community size (people)	1792.39 (2352.85)	2538.80 (2206.56)	0.14	160
Community size (households)	246.89 (272.11)	357.88 (358.90)	0.14	160
Daily agricultural wage for men (GHS)	11.71 (3.63)	13.82 (8.81)	0.16	160
Daily agricultural wage for women (GHS)	11.14 (6.19)	14.34 (13.12)	0.17	160
Daily non-agricultural wage for men (GHS)	13.71 (5.23)	14.10 (10.56)	0.56	160
Daily non-agricultural wage for women (GHS)	11.95 (5.47)	12.18 (7.71)	0.56	160
Commercial facilities				
Bank nearby (%)	42.50 (49.75)	45.00 (50.06)	0.86	160
Bookshop (%)	37.50 (48.72)	27.50 (44.93)	0.42	160
Microfinance institution (%)	15.00 (35.93)	12.50 (33.28)	0.87	160
Shop to buy children's clothes (%)	40.00 (49.30)	8.75 (28.43)	0.00***	160
Shop to buy children's toys and games (%)	17.50 (38.24)	7.50 (26.51)	0.22	160
Weekly market (%)	52.50 (50.25)	55.00 (50.06)	0.72	160
General market (%)	42.50 (49.75)	41.25 (49.54)	0.94	160
Other facilities				
Family planning facility (%)	20.00 (40.25)	35.00 (48.00)	0.11	160
Pharmacy (%)	66.25 (47.58)	67.50 (47.13)	0.95	160
Private clinic (%)	11.25 (31.80)	15.00 (35.93)	0.77	160
Private hospital (%)	5.00 (21.93)	2.50 (15.71)	0.54	160
Public clinic (%)	53.75 (50.17)	60.00 (49.30)	0.45	160
Public hospital nearby (%)	22.50 (42.02)	20.00 (40.25)	0.92	160
Private school (%)	10.00 (30.19)	17.50 (38.24)	0.19	160
Public school (%)	88.75 (31.80)	87.50 (33.28)	0.69	160
Sanitation				

Main water source is borehole (%)	50.00 (50.32)	57.50 (49.75)	0.40	160
Main water source is surface water (%)	40.00 (49.30)	32.50 (47.13)	0.28	160
Any public toilet in use (%)	28.75 (45.55)	35.00 (48.00)	0.39	160
Open defecation is common (%)	75.00 (43.57)	80.00 (40.25)	0.75	160
At least one electricity connection (%)	65.00 (48.00)	65.00 (48.00)	0.79	160
Main communities problems in the last 4 years				
Drought (%)	65.00 (48.00)	45.00 (50.06)	0.05**	160
Erosion or landslide (%)	7.50 (26.51)	5.00 (21.93)	0.66	160
Flood (%)	60.00 (49.30)	50.00 (50.32)	0.53	160
Fire (%)	7.50 (26.51)	12.50 (33.28)	0.44	160
Pests (crops) (%)	51.25 (50.30)	42.50 (49.75)	0.47	160
Pests (humans) (%)	7.50 (26.51)	7.50 (26.51)	0.88	160
Pests (animals) (%)	48.75 (50.30)	40.00 (49.30)	0.49	160
Overflow of river (%)	5.00 (21.93)	0.00 (0.00)	0.10	160
Strong wind (%)	73.75 (44.28)	57.50 (49.75)	0.14	160

Table 38. School characteristics

	Control	Treatment	p-value	N
Average no. Pupils per KG class	59.23 (34.78)	54.89 (22.56)	0.48	160
Desks per pupil	0.10 (0.16)	0.16 (0.19)	0.16	160
Has electricity (%)	5.00 (21.93)	15.00 (35.93)	0.15	160
Has a blackboard (%)	95.00 (21.93)	95.00 (21.93)	0.79	160
School has a toilet (%)	28.75 (45.55)	47.50 (50.25)	0.08*	160
Books available (%)	42.50 (49.75)	52.50 (50.25)	0.50	160
School has a major safety hazard (%)	43.75 (49.92)	48.75 (50.30)	0.64	160
Preschool teaching in English (%)	18.75 (39.28)	30.00 (46.11)	0.28	160
Preschool teaching in main local language (%)	78.75 (41.17)	65.00 (48.00)	0.24	160

Table 39. Teachers' (time invariant) characteristics captured at endline

	Control	Treatment	p-value	N
Age	33.05 (6.69)	33.08 (6.29)	0.73	160
Male (%)	53.75 (50.17)	55.00 (50.06)	0.60	160
Religion: Christian (%)	46.25 (50.17)	42.50 (49.75)	0.75	160
Religion: Islam (%)	52.50 (50.25)	56.25 (49.92)	0.76	160
Teacher has tertiary education (%)	90.00 (30.19)	92.50 (26.51)	0.74	160
Formally trained in ECD (%)	72.50 (44.93)	72.50 (44.93)	0.82	160
Years of KG teaching experience	0.64 (12.30)	2.34 (2.31)	0.47	81
Total years in current KG	0.83 (0.63)	1.13 (1.16)	0.06*	81
Taught in study KG in last term	87.50 (33.28)	90.00 (30.19)	0.73	160
Married (%)	77.50 (42.02)	85.00 (35.93)	0.31	160
Teacher born in community (%)	3.75 (19.12)	3.75 (19.12)	0.77	160
Live in community (%)	19.15 (39.77)	26.67 (44.72)	0.34	92
Moved community for the job (%)	51.06 (50.53)	40.00 (49.54)	0.29	92
Father has at least some formal education (%)	18.99 (39.47)	7.59 (26.66)	0.01**	158
Mother has at least some formal education (%)	15.00 (35.93)	2.53 (15.81)	0.00***	159
Is proficient in speaking English (%)	58.75 (49.54)	60.00 (49.30)	0.63	160
Is proficient in writing English (%)	65.00 (48.00)	60.00 (49.30)	0.77	160
Is proficient in speaking local language (%)	96.25 (19.12)	91.25 (28.43)	0.22	160
Is proficient in writing local language	58.75 (49.54)	52.50 (50.25)	0.66	160