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

Salifu Amadu
Orazio Attanasio
Bet Caeyers
Sarah Cattan
Lina Cardona Sosa
Sonya Krutikova
Peter Leighton
Lise Masselus
Mubarik Yakubu

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

Salifu Amadu[υ], Orazio Attanasio[⋆,†], Bet Caeyers[⋆], Sarah Cattan[⋆], Lina Cardona Sosa[⋆,ϵ], Sonya Krutikova[⋆], Peter Leighton[⋆], Lise Masselus[υ,γ], Mubarik Yakubu[υ]

[υ]Innovations for Poverty Action, Ghana

[⋆]The Institute for Fiscal Studies, London, UK

[†]University College London, UK

[ϵ] Central Bank of Colombia

[γ] Wageningen University, Netherlands

Written by

The Institute for Fiscal Studies
in collaboration with Innovations for Poverty Action, Ghana

7 Ridgmount Street
London WC1E 7AE
Tel: +44 (0) 20-7291 4800
Fax: +44 (0) 20-7323 4780
Email: mailbox@ifs.org.uk
Website: http:/ww.ifs.org.uk

# Preface

This report presents a detailed overview of the baseline data collection activities as part of the project “Improving early childhood development in rural Ghana through scalable low-cost community-run play schemes”. The project is a collaboration between The Institute for Fiscal Studies (UK), Lively Minds (UK, Ghana) and Innovations for Poverty Action (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 the others individuals or institutions mentioned here, including the Institute for Fiscal Studies (IFS), which has no corporate view. Data was collected in collaboration with Innovations for Poverty Action (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. We acknowledge the 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).

The authors would like to thank Madeleen Husselman, Ishmail Azindoo Baako, and Joyce Jumpah of IPA, as well as countless field staff, for their valuable contributions during the design and implementation of baseline 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 cooperation from the very beginning of this project, and their commitment to the implementation of the intervention. We are grateful for useful suggestions and collaboration from Sharon Wolf and Jere Behrman, and for excellent research assistance from Sam Crossman. Any errors and all views expressed are those of the authors.

# 1 Executive Summary

Early childhood care and education (ECCE) is critical to children's development and their success in adult life. Ghana has shown substantial commitment to improving ECCE. It has one of the highest Early Childhood Education (ECE) enrolment rates in Sub-Saharan Africa (SSA) and recognition that improving quality of kindergarten (KG) education is central to improving early childhood development and to reducing inequalities in learning outcomes, as demonstrated for instance through emphasis on teaching practice and learning environment in the 2012 Programme to Scale-up Quality KG Education of the Ghana Education Service (GES) - the implementing arm of the Ministry of Education.

However, especially in remote rural communities, low standard of kindergarten teaching, poverty, and low levels of education and empowerment of mothers, remain important barriers to ensuring adequate ECCE provision. Over the last nine years, Lively Minds, an award winning NGO, has been running a programme in rural communities in northern Ghana and Uganda which aims to overcome these barriers. It focuses on unlocking the potential of caregivers, both volunteer mothers and KG 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 programme targets those in rural kindergarten, aged 4 and 5. A key feature of the programme is its scalability; it is low cost and requires only locally available human and physical resources and infrastructure for implementation. Having developed, trialled and refined the programme content and training materials, Lively Minds are focusing on adapting implementation to move to a training of trainers model in which the Ghana Education Service (GES) and KG teachers take key roles in ensuring the success of the programme.

The Institute for Fiscal Studies (IFS) has partnered with Innovations for Poverty Action in Ghana (IPA) to implement a rigorous evaluation of the impact of the Lively Minds programme on the targeted children, their siblings and caregivers, volunteer mothers who run the play-schemes and teachers who train the volunteer mothers. For this purpose we are conducting a Randomised Controlled Trial (RCT) of 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. Further, we aim to generate evidence which will contribute more broadly to the state of knowledge on development and scaling of ECCE interventions in low-income remote rural contexts. To this end we have three overarching objectives:

1. Evaluate direct impacts of the Lively Minds programme on targeted children and caregivers, as well as indirect impacts on teachers and siblings.

2. Identify the mechanisms that determine the impacts of ECCE interventions on child development.

3. Increase knowledge on how to effectively scale up ECCE programmes at sustainable costs, using local resources and infrastructure, in very low-income rural contexts.

This report presents findings from the baseline stage of the RCT. The validity of our evaluation approach hinges on achieving successful randomisation so that there are no ex-ante differences between treatment and control communities that might hinder our ability to identify full impacts of the Lively Minds Programme. The key objective of this report is, therefore, to utilise data collected as part of the baseline before the start of programme implementation to test whether the randomisation was successful and the sample is balanced on key observable characteristics between Treatment and Control communities. In addition, the report provides an overview of the Lively Minds programme, our evaluation design, baseline data-collection, as well as implementation challenges and risks. The baseline data provide a rare glimpse of a remote and understudied part of Ghana. We, therefore, also present a snapshot of some key characteristics of the environment and practices that the target children experience.

In total, 80 schools were chosen to be part of the evaluation. These included 38 schools in the Bongo District (Upper East Region), and 42 Schools in the Tolon District (Northern Region). Half of these schools were randomly assigned to receive the Lively Minds intervention, and the other half to a control group - who will receive the intervention a year later. As part of the baseline data collection, we surveyed 2407 target children (aged 3-5), as well as their primary caregiver, household head and older and/or younger sibling if they have one. Observations were also conducted on all 80 schools and surveys administered to the 151 kindergarten teachers in these schools.

We implemented balance tests on 303 individual variables, and of these only 25 (8.3%) showed statistically significant differences between treatment and control groups at the 10% level; this is less than would be expected by chance. In addition there is balance along our main outcomes of interest; including IDELA scores, primary caregiver mental health, teacher well-being and parental investments. This is vital as it shows that the randomisation was successful in creating study groups that are comparable on observable dimensions, prior to the intervention.

The majority of outcome measures, including tests of child development and primary caregiver well being, performed well. Certain items such as primary caregiver knowledge of childhood development were less successful and will be adapted for endline. Summary statistics presented provide an interesting picture of the key barriers to improved child outcomes which highlights the value of the Lively Minds approach in targeting not only schooling, but also empowering teachers and mothers themselves. The main findings are:

1. *Schools are under resourced and learning outcomes are low* ; Average class sizes in the kindergarten are extremely large at 58, and schools, particularly in Tolon, often lack books and other important resources for children. These problems extend into later ages, with only a quarter of 6-10 year olds being able to do simple addition or read even a single letter.

2. *Parental "investments" in their children are low* : The majority of primary caregivers have not conducted any form of play activity with their child in the last 3 days, and less than half have any form of play materials, even those that are homemade. In part, this may be due to a low availability of materials; only 12% of communities have a shop that stocks toys nearby, and even these are largely restricted to dolls.

3. *Parental involvement in school is low* : Less than half of primary caregivers know the name of their child's kindergarten teacher, and many teachers report lack of parental involvement as a significant issue.

4. *Parental educational levels are low* : Only 20% of primary caregivers in our sample have ever attended formal schooling.

5. *Maternal mental health issues are highly prevalent* : By any prevailing cut-off used for the SRQ-20 measure, the risk of depression is high among primary caregivers in this sample.

6. *Low socio-economic status*: Agriculture is the main source of income for the majority of households in the sample, and the reported average daily agricultural wage is only £1.81. Ownership of basic durables is low, with less than a half owning a bed, or table and chair. In addition, wealth indices constructed from

these assets show that the study districts are relatively deprived compared to the national average.

Although it is too early to conclude a great deal about the potential scalability of the intervention, baseline results offer some interesting lessons to consider. On average, the data confirm the validity of several contextual assumptions on the basis of which the scalability of the model is based. However, there are large differences seen between the two study districts in terms of culture, socio-economic status, religion and household structure, but also interestingly in developmental outcomes for children and mental health of primary caregivers. These regional differences raise the question of whether the same intervention model is equally applicable and effective across the country or whether context-specific adaptations will need to be made when going to scale.

# Contents

# 2 Introduction and background

## 2.1 Introduction

Early childhood care and education (ECCE) is critical to children's development and their success in adult life. Good quality ECCE can help children be healthier, do better and stay longer in school, and have better trajectories in adult life (Cawley, Heckman and Vytlacil, 2001; Heckman, Stixrud and Urzua, 2006; Van der Gaag, 2010). Children living in the 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 0.0 years in Northern Region and 0.6 years in Upper East Region, compared to a national median of 4.4 years (Ghana Statistical Service - GSS, Ghana Health Service - GHS and ICF International, 2015).

Overall the evidence available in the literature suggests that interventions that improve the quality of interactions between caregivers (parents and teachers) and children can have positive impacts on child development. The robust evidence that exists on impacts of pre-schools is limited to Lower and Middle Income Country (LMIC) settings and suggests that impacts on child development are mixed (Berlinski, Galiani and Gertler, 2009; Bernal and Fernández, 2013; Rosero and Oosterbeek, 2011). Evidence from urban Ghana suggests that pre-school involvement serves as the primary mediating mechanism between household socioeconomic status and child school readiness (Wolf and McCoy, 2017). Pedagogical experience is considered crucial to pre-school efficacy and interventions that focus on improving this dimension have been found to be more effective at raising the impacts of pre-school than those targeting structural quality only (e.g. physical infrastructure, quality of furnishings and space; Bernal and Fernández (2013); Attanasio et al. (2016)). On the other hand, a well-known set of studies suggests that home-visiting interventions that aim to promote children's stimulation by encouraging mothers to teach skills and concepts in daily routine activities can achieve impressive long-term impacts on children's cognitive and socio-emotional development, success at school and even wages (Heckman et al., 2010; Barnett and

Masse, 2007; Grantham-McGregor et al., 1991; Gertler et al., 2014). Importantly, however, the evidence on such programmes is by no means conclusive and there are many instances where early impacts fade out (see overview by Bailey et al. (2017)).

As far as we are aware, there is no evidence in the literature on the efficacy of interventions that target improvements in both parental practice and quality of pre-schools. Over the last nine years, Lively Minds, an award winning NGO, has been running an innovative programme in highly deprived communities in Ghana and Uganda which aims to unlock the childcare potential of both mothers and GK teachers. LM's model is based on training and empowering uneducated marginalised volunteer mothers and GK teachers with the knowledge, skills and confidence to run free educational Play Schemes 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, Lively Minds are now trialling a more scalable and cost-effective implementation model in which the play schemes are integrated into the Government KG system and delivered through the Ghana Education System (GES) using a training of trainers approach, making use of GES infrastructure and resources, such as classrooms.

Lively Minds' own monitoring and evaluation of initial pilot testing shows high rates of compliance and suggestive evidence of wide-ranging positive outcomes for teachers, mothers and children. The Institute for Fiscal Studies (IFS) is now implementing a large-scale randomised controlled trial (RCT) to conduct a rigorous evaluation of the full impacts of the scalable model on the targeted children, their siblings and caregivers, volunteer mothers who run the play-schemes and teachers who train the volunteer mothers. The study will build on the available evidence in a number of important ways.

First, the literature continues to be dominated by studies of High Income Countries (HIC) such as the US, UK and Canada. Furthermore, even among studies set in LMIC's very few are in rural deprived settings that are comparable to our study context. At the same time, on the whole, children from more deprived backgrounds have been found to benefit the most from ECCE interventions so the promise of such programmes may be especially great in settings such as ours.

Second, existing studies currently focus on "siloed" interventions which either target parents or nursery/pre-school teachers and rarely measure spillovers on other children such as siblings. Our study will contribute novel evidence on effectiveness of targeting both home and pre-school environments at the same time and provide a more complete assessment of impact by evaluating not only the target children but also their siblings.

Third, some of the most widely studied ECCE interventions, such as the High/Scope Perry School and the Abecedarian project, were implemented at very small scale and relying on highly skilled professionals. Translating these findings into programmes delivered at scale in low resource settings requires vital further research into best practice in adaptation of pilot successes to implementation at scale and cost-effective strategies for monitoring and evaluating this process. A key objective of this study is to identify key mechanisms underlying any impacts so that going forward process data can be used to verify to what extent these survive with going to scale. 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.

Finally, the project contributes to the ECCE database by collecting rare, large-scale detailed information on child development outcomes and potential determinants, of a remote and under-studied part of Ghana. It will also contribute to the general effort to adapt and apply existing measures of child development and measures of early childhood environment and will assess the performance of these measures in remote settings about which only very little is known to date.

The focus of this report is on the activities carried out up until and including baseline data collection. The remainder of Section 2 covers the policy context, details about the intervention, and a description of partners involved in the study. An outline of all activities carried out is provided in Section 3, including details on randomisation, sampling, survey instruments and data collection; more precise details are given in the Appendix. The purpose of analysing the baseline data is three fold; Firstly, it is used to check the validity of our study by testing any systematic underlying differences between the treatment and control group (see Section 4). Secondly, it is useful in providing an interesting snapshot of the study area, as shown through a range of descriptive statistics in Section 5. Finally, the baseline data also provide valuable information on the performance of our main outcome measures, which is highlighted where relevant in the same section. These data are also used to provide some tentative conclusions for the intervention and the scale-up for the Lively Minds project (see Section 6).

## 2.2 Policy context

The Government of Ghana's adoption of the National Early Childhood Care and Development Policy in 2004 highlights access to quality kindergarten education as central to improving early childhood development and to reducing inequalities in learning outcomes. In 2007, Ghana became the first Sub-Saharan African (SSA) country to expand to 2 years of pre-primary education (kindergarten) and currently has one of the highest SSA early childhood education (ECE) enrollment rates. The 2012 Government Kindergarten (KG) Situational Report concludes that the curriculum established in 2004 is sound but efforts are needed to improve teacher behaviour to reect the pedagogy. The GES 2012 Programme to Scale-up Quality KG Education sets three critical pathways to scaling up quality Kindergarten education;

- To provide access to KG for all 4 and 5 year olds
- To transform teaching practice and learning environments in order to deliver activity based learning
- To define and measure a set of outcomes

The LM innovation aligns with this strategic framework and has the potential to improve the quality and outcomes of KG education in Ghana by tackling specific goals outlined in the plan, the second goal in particular. By targeting KGs in deprived rural Ghana with limited resources and inadequate infrastructure the programme aims to transform the KG classroom and learning outcomes of children through teacher training, parental capacity building and stimulation of KG children.

## 2.3 The Intervention

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) for 9 years. The intervention focuses on unlocking the potential of caregivers, both volunteer 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 programme targets those in rural kindergarten, aged 4 and 5, and is structured as follows:

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

2. **Teacher training**: Two Kindergarten 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 Headteacher and PTA representative attend two of these days. The training is facilitated by Lively Minds staff, and supported by GES officials.

3. **Training of Volunteer Mothers**: The trained Kindergarten Teachers then train 30-40 Volunteer Mothers (VMs) in their community using a scripted curriculum. To maintain quality, Teachers are supervised and supported by high performing Kindergarten Teachers from schools with existing Play Schemes. District officials and Lively Minds staff also provide some monitoring support. 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 kindergarten age, child-friendly teaching. In addition, VMs are taught how to install simple handwashing devices (tippy-taps) at home.

4. **Play Schemes run**: VMs are divided into 4 four groups, and each group is given a different day where they teach at the kindergarten for one hour. 25 kindergarten children are arranged in small groups (maximum of 5) and rotate around the following 5 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. The remaining kindergarten children play outdoor games, led by VMs. The Kindergarten teachers supervise the sessions. Children have to handwash with soap before using the Scheme, sensitising them to this vital practice.

5. **Ongoing support**: GES officials and Lively Minds staff have monthly meetings to track progress of the Play Schemes and identify corrective measures. GES officials monitor the Play Schemes as part of their normal supervisory functions and Lively Minds conduct some unannounced monitoring visits. Once a month "top-up" training workshops are held for Kindergarten Teachers where they discuss problems, share successes and are trained to provide the VMs with a monthly Parenting Workshop.

   VMs are given monthly parenting workshops by the Kindergarten 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, inclusive education. Over time, Play Scheme Committees are established in each school to manage their Play Scheme, made up of 4 VMs, PTA representative and Kindergarten Teachers and Head Teacher.

## 2.4 Intervention timeline

Play schemes in treatment communities are running throughout most of the 2017/2018 school year, starting in October 2017, soon after the start of the school year in September 2017, and ending in July 2018. Table 1 provides a timeline for the intervention activities starting from the beginning of intervention activities to the start of the play schemes.

Table 1: Intervention timeline

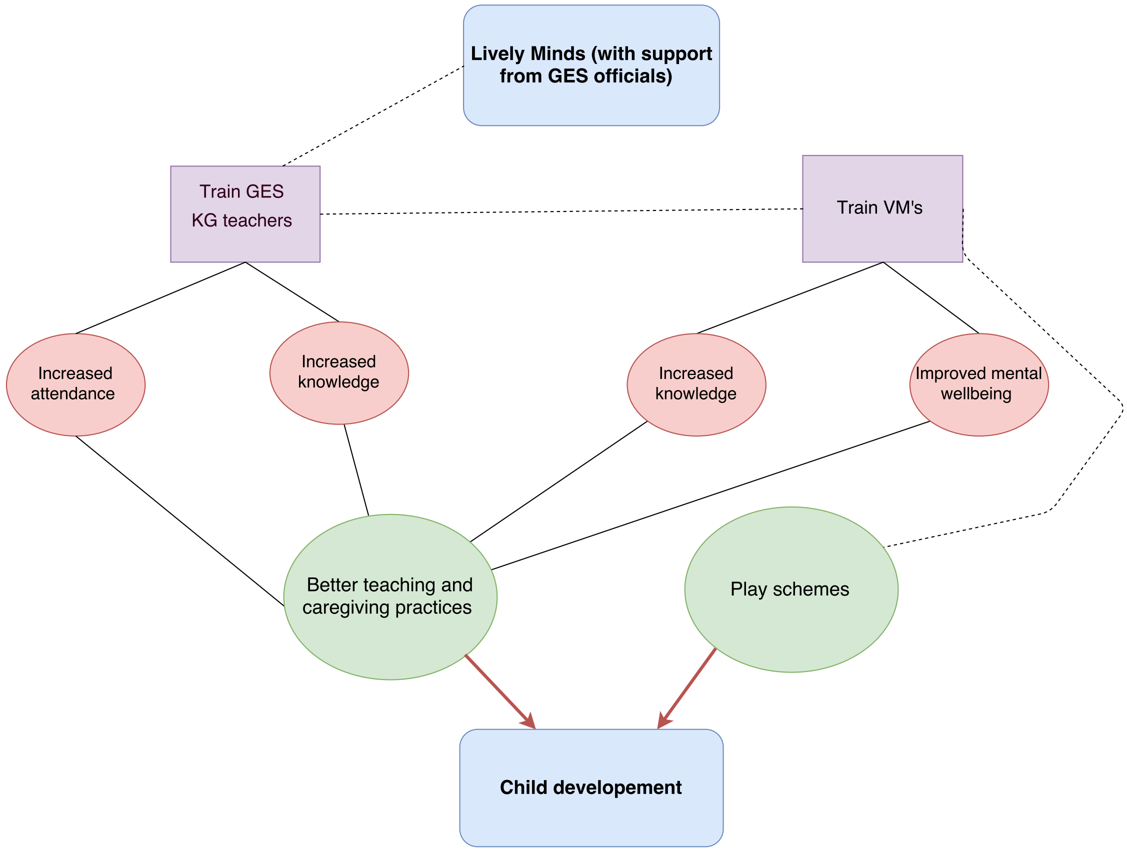
| INTERVENTION TIMELINE | September | | | | October | | | | | | | | | | | | December | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Activity** | **5** | **20** | **25** | **27-29** | **2-3** | **4** | **5** | **11-13** | **16** | **18** | **20** | **23** | **25** | **27** | **30** | **31** | **1-30** | **1-16** |
| Play scheme training team workshop | | | | | | | | | | | | | | | | | | |
| Mobilisation of KG teachers, Head teachers, PTA, CSs for Training of Teachers (ToT) in 40 Treatment communities | | | | | | | | | | | | | | | | | | |
| Facilitator training | | | | | | | | | | | | | | | | | | |
| Trainer of trainers (TOT) | | | | | | | | | | | | | | | | | | |
| Community meetings to inform about the programme, responsibilities and next recruitment dates | | | | | | | | | | | | | | | | | | |
| VM recruitment meeting | | | | | | | | | | | | | | | | | | |
| VM training and graduation | | | | | | | | | | | | | | | | | | |
| Top-Up 1 (will review programme set up, quality control, etc) | | | | | | | | | | | | | | | | | | |
| VM Graduation | | | | | | | | | | | | | | | | | | |
| Play scheme starts (until winter break) | | | | | | | | | | | | | | | | | | |

## 2.5 Theory of change

Our 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 context. These are: (1) low teacher training, knowledge and motivation; (2) risk of mental health problems for volunteer mothers as well as lack of information on early childhood development ; and (3) a lack of play based learning. 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 Play Schemes. 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) 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 - both in terms of materials and time; (iii) Children will benefit directly from the Play Schemes through exposure to play based learning which will improve their cognitive, motor, and socio-emotional skills.

This theory of change encapsulates the direct effects of the intervention. We hypothesise that there could be a range of additional spillover effects. Firstly, although only volunteer mothers directly participate in the intervention, there could be beneficial impacts upon other women in the community. These could set in through interaction with volunteer mothers and teachers, or with their children who are in the play scheme. Moreover there could be benefits for siblings of children attending a Play Scheme even if they do not attend. This could happen through interaction with the target child 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 target child.

Figure 1: Theory of change




## 2.6 Partnerships

This project is a collaboration between the IFS, Innovations for Poverty Action (IPA), Ghana, and Lively Minds. The main roles are as follows:

- *IFS* : Lead on the overall design of the RCT and analysis of results. This includes study design, randomisation, designing survey, data analysis, and academic and policy publications. IFS staff have also been present throughout large parts of the field work to assist in monitoring the data collection, logistics, and training.

- *IPA*: Are primary responsible for the collection of high quality data from which the RCT analysis is based. This includes the managing, hiring and training of field staff, ensuring the collection of four main waves of data collection (census, baseline, midline, endline), and conducting a range of quality checks on the data. IPA are also responsible for providing feedback on proposed survey instruments, and offering contextual knowledge and recommendations through extensive piloting.

- *Lively Minds*: Are primarily responsible for the implementation of the play scheme intervention in the 40 treatment schools. Lively Minds are also collecting a range of monitoring data which will aid the analysis of key mechanisms underlying the programme.

New partnerships have also been formed with researchers on related ECCE projects completed and ongoing in urban Ghana; they include Sharon Wolf (University of Pennsylvania), Jere Behrman (University of Pennsylvania) and Lawrence Aber (New York University). Going forward we plan to collaborate with this group of researchers to study geographic variation in developmental trajectories and ECCE experience of children living in Ghana. Where appropriate we are using the same measures of child development, as well as home and school environments. We plan to start with a comparative analysis of the IDELA assessment in our two samples.

# 3 Evaluation

## 3.1 The evaluation problem

The core evaluation problem is that in order to measure the true impact of Lively Minds on a set of outcomes, we would ideally want to compare the outcomes for those who receive the intervention (the "treatment" group) across two states of the world ; (i) where they receive the Lively Minds intervention, (ii) where the same group does not receive the intervention. However, as it is not possible to observe outcomes for the same group of children in two states of the world, the core evaluation problem is to come up with the best possible estimate for (ii), the *counterfactual*.

One way of estimating this counterfactual is through allocating the intervention randomly, through a randomised controlled trial (RCT). The random assignment means that there should be no systematic differences, prior to the intervention, between the treatment and control groups. As a result, measures of outcomes in the control group are a good approximation to what outcomes in the treatment group would have been, had the programme not been implemented. Therefore, in the context of an RCT, a simple difference between the means of a given outcome in the treatment and control groups provides a valid estimate of the causal effect of the intervention. This is the approach we adopt to evaluate the impact of the Lively Minds programme.

## 3.2 Evaluation Design

### 3.2.1 Evaluation outcomes and instruments

As outlined in the theory of change, there are a number of groups of individuals that we will be studying in the evaluation. These are defined as follows;

1. *Target children (TC)* are the key subjects of interest in the study. They were subject to the following eligibility criteria; (i) aged between 3 and 5 years old as of the start of the school term on the 11th September 2017, (ii) intending to attend or currently attending one of the 80 study schools.

2. *Younger Sibling (YS)* are children who meet the following criteria; (i) Aged less than 3 years at the time of baseline, (ii) share the same primary caregiver as the TC within their household. If multiple children satisfy these criteria, the oldest eligible sibling in the household was chosen as the YS. The younger "siblings" need not be biological siblings of the TC, but must share the same primary caregiver. Younger siblings were sampled in order to assess indirect or "spillover" effects of the intervention. With perhaps a few exceptions of volunteer mothers taking the YS to the play scheme, younger siblings are unlikely to have directly participated in the Lively Minds intervention but could have still benefited in some way due to interaction with the TC and other household members.

3. *Older Sibling (OS)* are children who meet the following criteria; (i) Aged greater than 5 and less than 11 (the age of moving from primary school to junior secondary school) at the time of baseline, (ii) share the same primary caregiver as the TC within their household. If multiple children satisfy these criteria, the youngest eligible sibling in the household was chosen as the OS. Similarly to younger siblings, older siblings were sampled in order to assess potential spillover effects of the intervention.

4. *Primary Caregivers* (PC) are of key interest in this study due to their importance in the development of the TC and their potential direct involvement in the intervention as volunteer mothers. They are defined as the person that spent most time caring for the child (i.e. more than half the time) during the last 6 months when the child was not in formal pre-school child care.

5. *Households* are defined as a person or a group of persons who: (i) live together under the same roof or compound recognising one person as the head,(ii) eat from

the same pot, (iii) have lived in the household for at least 30 days consecutively/60 days non-consecutively in the last twelve months.

6. *Volunteer mothers* (VMs) are women who volunteer to be a part of the Lively Minds intervention, and complete all of the required training to do so. These are not all mothers of TCs, some are grandparents or other relations, and some may have no relation to a child in the play scheme but volunteer nevertheless.

7. *Kindergarten Teachers* (KGTs) are GES staff recruited to be kindergarten teachers in the academic years 2017/2018

8. *Communities* are geographic areas of settlement, defined as an area under the rule of one chief. This is a not necessarily the same level as the school. Some communities have multiple schools, whereas other communities do not have any schools. In the total sample of 80 schools, 78 communities were sampled. For a given school, the community chosen for the baseline survey, was the one reported to be the "main" community served by the school.

We are interested in how the Lively Minds intervention affects a range of primary and secondary outcomes. The main aim is to assess the impact of the intervention on a broad set of children's developmental outcomes - these constitute the primary outcomes. However, a key aim of the study is to disentangle some of the mechanisms underlying intervention impacts. Therefore, we are collecting additional data that will allow us to estimate impacts on secondary outcomes, including on volunteer mothers and teachers through whom we expect the impact on the primary outcomes to occur. Finally, we will study whether there are any spillover effects on siblings. The data from baseline will allow us to assess the performance of different outcome measures, so that they can be changed and adjusted for endline if needed.

*Primary outcome*

1. Target children's cognitive and socio-emotional developmental outcomes: this will be measured primarily using the International Development and Early Learning Assessment (IDELA) tool, developed by Save the Children. This provides measures of development along 5 core domains; emergent numeracy, emergent literacy, socio-emotional skills, motor skills, and executive function (Pisani, Borisova and Dowd, 2015). Some small adaptions were made after piloting to fit the study context (see Section A.3 for full details). We will also measure socio-emotional development using the Strength and Difficulties Questionnaire (Goodman, 1997).

*Secondary outcomes*

1. Target children's health outcomes: this will be measured through data on the incidence of diarrhea, fever and respiratory infections using the definitions of the WHO as measures of morbidity, as well as measured arm circumference.

2. Developmental outcomes of siblings: the development of younger siblings will be measured through the Caregiver-Reported Early Development Index (CREDI) short form (McCoy et al., 2017), and of older siblings through Ravens progressive matrices (Raven, 1936) and tests of basic literacy and numeracy.

3. Child pre-school attendance and participation in the Play Schemes: this will be collected from self reports in the household survey

4. Maternal knowledge of child stimulation and care practices: this will be collected through information on the mother's knowledge of stimulation and care practice, 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).

5. Psychological well-being and empowerment of primary caregivers: this will be measured through the SRQ-20 screening tool for common mental disorders (Goodman, 1997), the Rosenberg measure of Self-esteem (Rosenberg, 1965) and reported influence over different household decisions.

6. Investment in the child within the households: we will use the Family Care Indicators instrument, developed by UNICEF (Frongillo, Sywulka and Kariger, 2003) to capture availability of toys and learning materials in the household, parental involvement with the child, the child's routines and organisation of the child's time inside and outside their home.

7. Target children's hygiene knowledge: we will 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 why is hand-washing important. These questions are taken from the IDELA tool (Pisani, Borisova and Dowd, 2015).

8. Pre-school engagement of primary caregivers: we will measure how often primary caregivers visit the child's school and whether they know the teacher's name

9. Teacher wellbeing, teaching practices and knowledge: this will be measured in part using an instrument developed in a previous study in Ghana (Behrman, Wolf and Aber, 2017) which captures teacher presence, practices, burnout and job satisfaction. The SRQ-20 will be administered to teachers to assess their mental wellbeing. Given the small sample size of teachers in our study, however, we may not have sufficient power to detect significant impacts on this outcome.

Data on all of these outcomes measures were collected during the baseline. This allows us to study whether there were any significant differences in these between the treatment and control groups before the start of the intervention (see Section 4). In addition, we will be able to control for these in the evaluation analysis, increasing the precision with which we can estimate treatment effects on the primary and secondary outcomes. Furthermore, baseline data includes measures of a rich set of background characteristics of the household, school and primary caregiver. These will be used as control variables in the evaluation analysis to increase precision, as well as to investigate heterogeneity in impacts. Table 2 outlines the details of the main survey modules used during baseline.

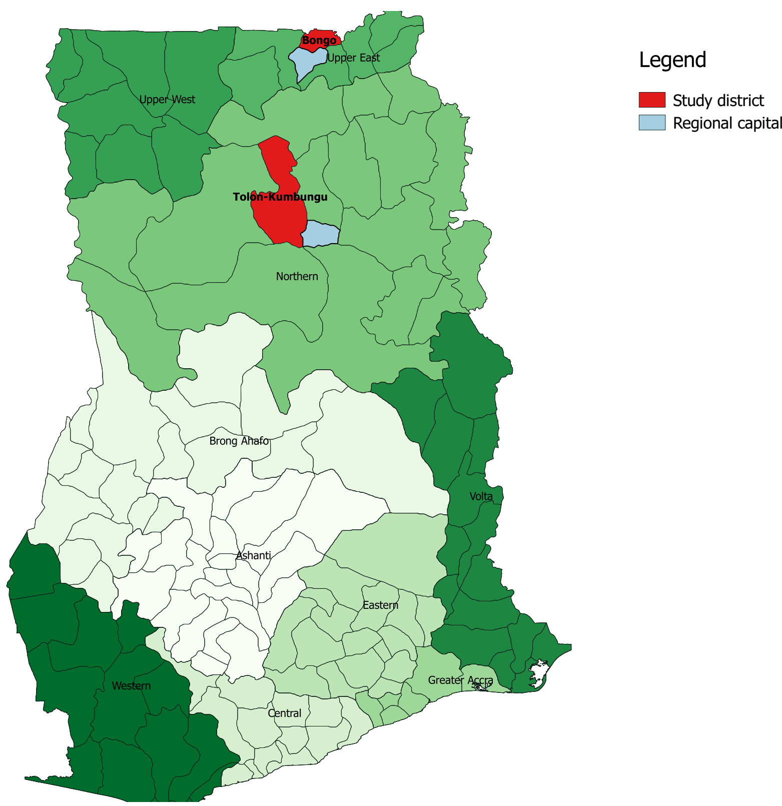
Table 2: Baseline survey modules

| **Module** | **Description** | **Respondent** | **Average duration** |
|---|---|---|---|
| Household | (i) Household roster: Age verification, respondent identification and information on each household member, covering gender, age, biological mother and father, ability to read and write, school attendance, highest grade completed, pre-school attendance, reason for attending/not attending school. (ii) Main household: covers household assets, dwelling, land and livestock ownership, income, consumption and expenditures, savings and credit, shocks, travel to pre-school | Household head or most knowledgeable person | 78 minutes |
| Primary Caregiver | (i) Information on the primary caregiver including basic demographics, time use, social capital, knowledge on child development, hygiene, physical health, mental health, self-esteem, empowerment, school engagement, home environment (FCI), Ravens matrices. (ii) Information on the target child including; health, time use, socio-emotional skills, languages spoken. | Primary caregiver of TC | 64 minutes |
| Target child | Measurement of development using IDELA tool | TC | 41 minutes |
| Younger sibling | Maternal report of development using CREDI | PC | 9 minutes |
| Older sibling | Test comprising of backwards and forwards digit span, literacy, numeracy, ravens matrices. | OS | 24 minutes |
| Teacher survey | Background characteristics, economic situation, social capital, work conditions, mental health, motivation, job satisfaction, burnout, reading knowledge, teaching practices, knowledge of early childhood development. | KGT | 92 minutes |
| Community survey | General characteristics, education and childcare services, health services, local economy and wages, local prices, water and sanitation, shocks, social protection, migration. | Community elders | 147 minutes |
| Market survey | Prices of items not found in the community from the major local markets. | Observation from IPA enumerator | |
| School survey | Observation on basic amenities, state of physical building, learning materials and environment, toilets and sanitation | Observation from IPA enumerator | |

### 3.2.2 Randomisation

In total, 80 schools were chosen to be part of the evaluation. These include 38 schools in the Bongo District (Upper East Region), and 42 Schools in the Tolon District (Northern Region). Figure 2 displays the location of study districts. It can be seen that both districts are neighbouring the respective regional capitals, and are therefore less rural than the average community where Lively Minds works. Despite neighbouring each other, these regions differ in a number of aspects, including language, religion, culture and schooling quality (see Section 5.10 for further information). The inclusion of both regions in the analysis will, therefore, help to evaluate the success of the Lively Minds intervention across different contexts. For further details on how the two study districts were chosen see Section A.1.

Figure 2: Study Districts




*Note:* Tolon-Kumbungu district was split into two districts; Tolon and Kumbungu, in 2012. However it was not possible to locate shapefiles with these new boundaries. The present day Tolon district is a smaller segment of the red area displayed.

The unit of randomisation was chosen to be the school, since the Lively Minds intervention takes places at the school level. Each of the 80 schools were randomly allocated to one of two groups:

1. Treatment (40 schools) : schools to receive the LM intervention starting in October 2017

2. Control (40 schools) : schools to receive the LM intervention in September 2018

The randomisation was conducted using two levels of stratification: circuit, and school size. A circuit is a geographical cluster of around 10 schools which 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 (both KG1 and KG2) within the strata, (ii) "large schools" ; defined as having above the median number of total KG children. The size of school is likely to be highly relevant for the efficacy of the intervention, hence this stratification was used to ensure strong balance on this variable. Figures 3 and 4, maps the randomisation of schools, revealing that treatment and control schools are well spread out geographically.

Figure 3: Randomisation of schools (Bongo)

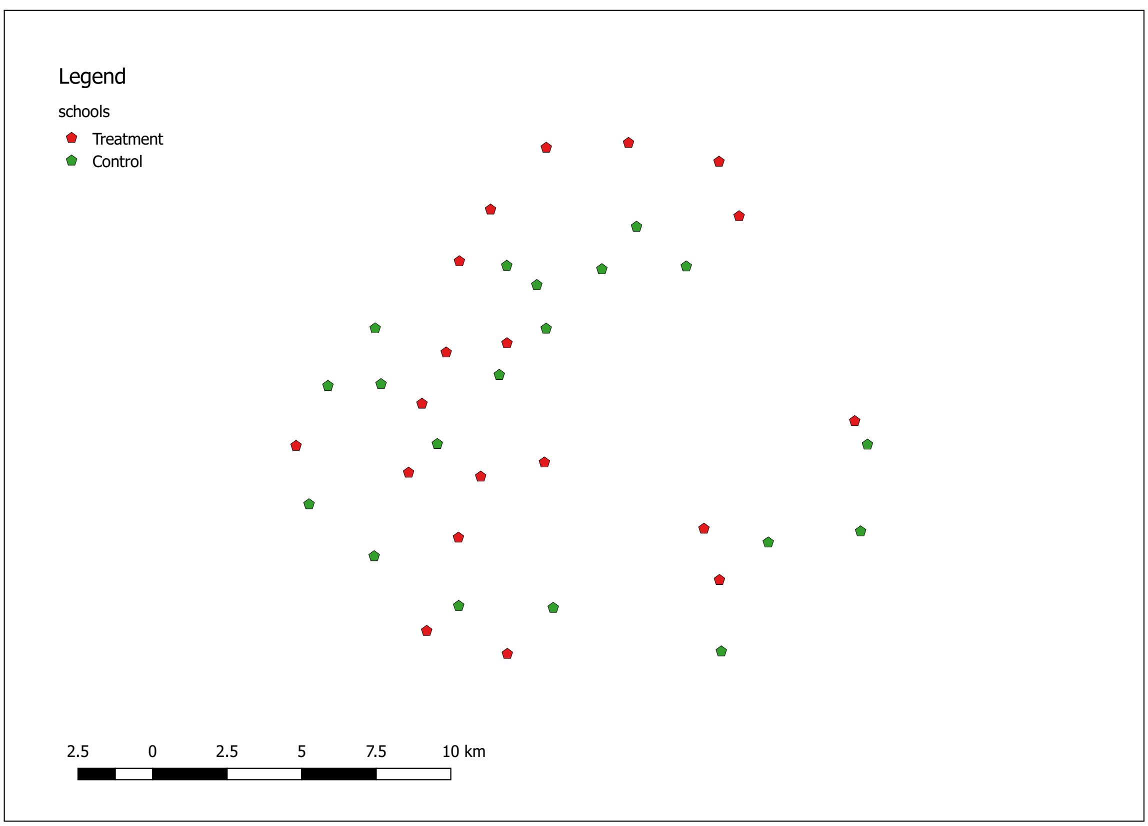
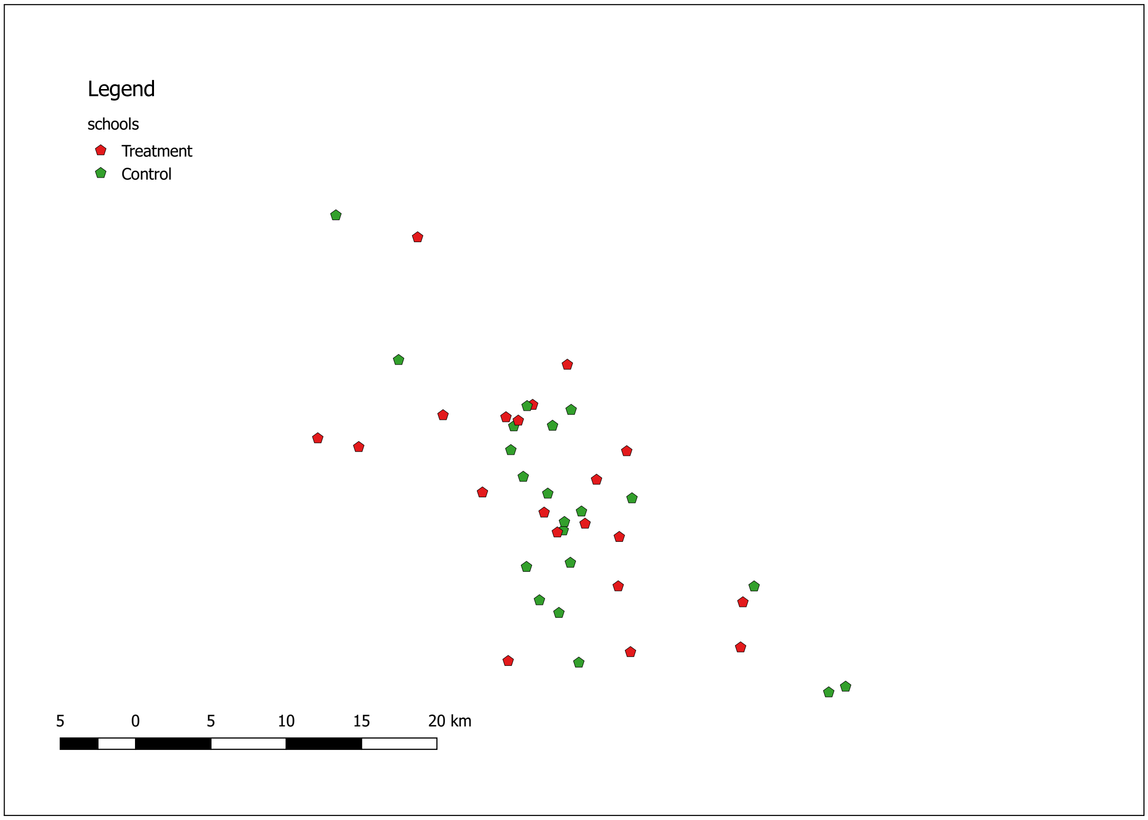


Figure 4: Randomisation of schools (Tolon)




As explained in Section 3.1, the RCT allows for the estimation of the "causal" effect of the programme, as the control group provides a valid counterfactual for what would have happened to the treatment group in the absence of the Lively Minds programme. This is true when estimating impacts on children, siblings, or teachers. However, within our design this does not hold for VMs. Volunteer Mothers are not a randomly selected sample of women in communities randomly allocated to treatment or control - these are women who chose to volunteer based on specific characteristics or personality traits. In order to form an appropriate counterfactual for this group it is necessary to look at a group of mothers who *would have volunteered had they been in the treatment group*. In order to identify this group, mothers were mobilised in the control group, in a way that followed the strategy in the treatment communities as closely as possible. Mothers in the control group were called to a community meeting and informed about the intervention that would be coming in a year's time. They were then offered the chance to sign up to the programme, with the names of those signing up being recorded. This group of mothers are a valid counterfactual for the treatment group under two main assumptions: (i) those who are willing to sign up to a programme starting in a year's time are not different to those who would sign up to a programme starting immediately in any way that is related to outcomes of interest; (ii) mobilisation of mothers in control schools was done in the same way as treatment schools. Although not the first best situation, we believe that these assumptions are reasonable and that therefore the use of the quasi "control group" will allow for a valid estimate of the effect of the programme on volunteer mothers.

### 3.2.3 Power calculations

Power calculations were originally done at the proposal 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 of between 19% and 30% of a standard deviation. Given that we now have baseline data, these power calculations have been updated using more accurate information. The new calculations, done by simulation as outlined in Section A.2, show a minimal detectable effect size of between 11% and 12% of a standard deviation, on the primary outcome (IDELA school readiness score). This is well within the acceptable range given the measured impact of previous early childhood interventions.

## 3.3 Data collection

### 3.3.1 Sampling

*Census*

As there are no existing representative surveys of the population in the study areas, a census survey was conducted in July 2017 in order to provide a sampling frame from which a random sample of target children could then be drawn. The survey included a few basic questions designed to ascertain whether children eligible for the Lively Minds programme (within the target age-range) resided in the household. These questions included basic roster of the household head, caregivers, and all children under the age of 10, alongside the school that they are currently attending or intended to go to, if any, the following September.

With the intervention and randomisation taking place at the school level, the sampling frame needed to provide a sample of children for each school. Following exploratory work in the field, however, it became clear that there was no simple association between which communities an individual resides in and which school they attend or are planning to attend. In order to provide a suitable sample of children for each school in the face of these challenges, the following strategy was implemented:

1. The closest households to each school, up to a maximum of 150 households per school, were surveyed during the census. Given information from Lively Minds that children often attend nearby schools, this was done to ensure that the majority of children who are likely to be attending a given school, were included in the sampling frame. The methods for achieving this are outlined in Section 3.5.2.

2. Individuals with children of pre-school age were asked: (i) which school they were attending at the time of the census; or (ii) if not attending school, whether they intend to go to school in September and if so, to which school they do. Children were then classed as being part of a given school, if they reported attending that school, or intending to do so in September.

This approach can be seen visually in Figures 5 and 6; where for each school the surrounding households were enumerated as part of the census. This also highlights the differing dispersion of households across the two study districts; households within a community are clustered close to one another in Tolon, whereas in Bongo they tend to be spread further apart. Target children were then defined according to two criteria: (i) aged 3-5 as of 11th September 2017 (the start of the school term); and (ii) currently attending or intending to attend one of the 80 study schools.

The census survey took place between 11th-31st July 2017. This resulted in a total of 9503 household surveys, capturing all 80 schools. Table 3 displays some basic statistics from the census survey. In total the census covered 6,446 target children (TCs) living in 4,486 households.

Figure 5: Example census schools (Tolon)

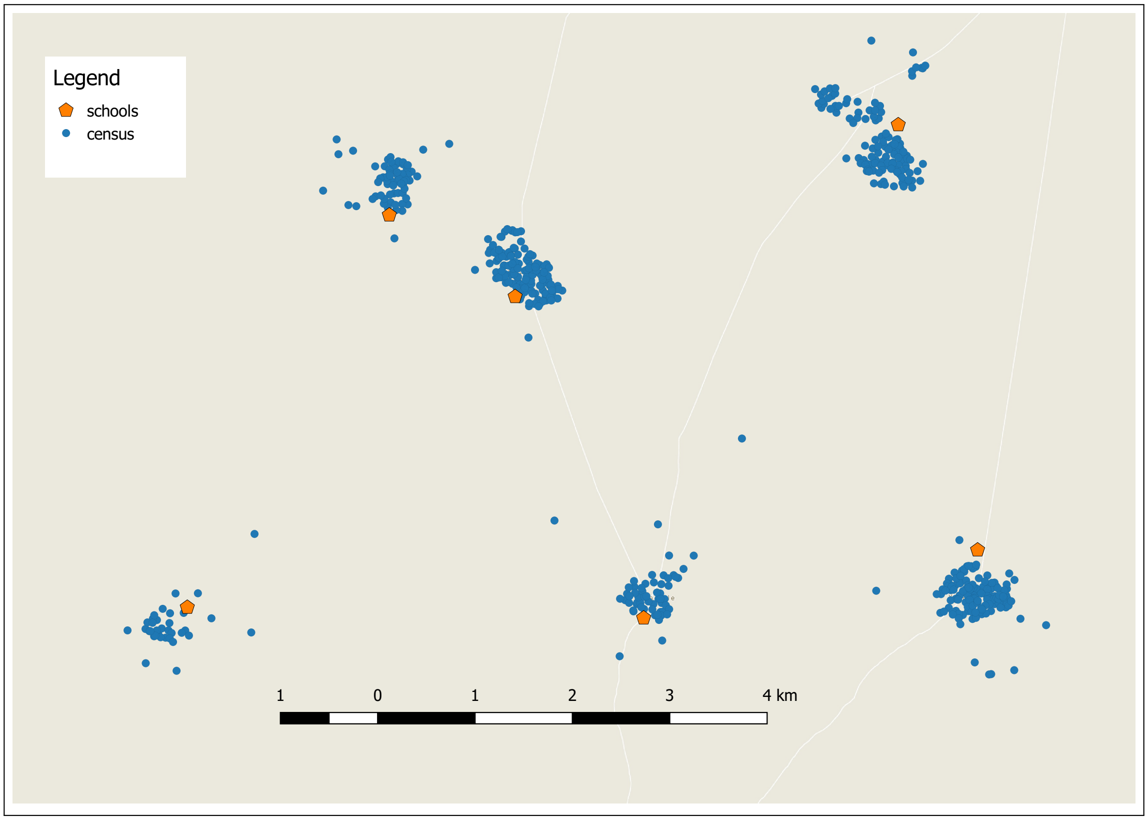


Figure 6: Example census schools (Bongo)

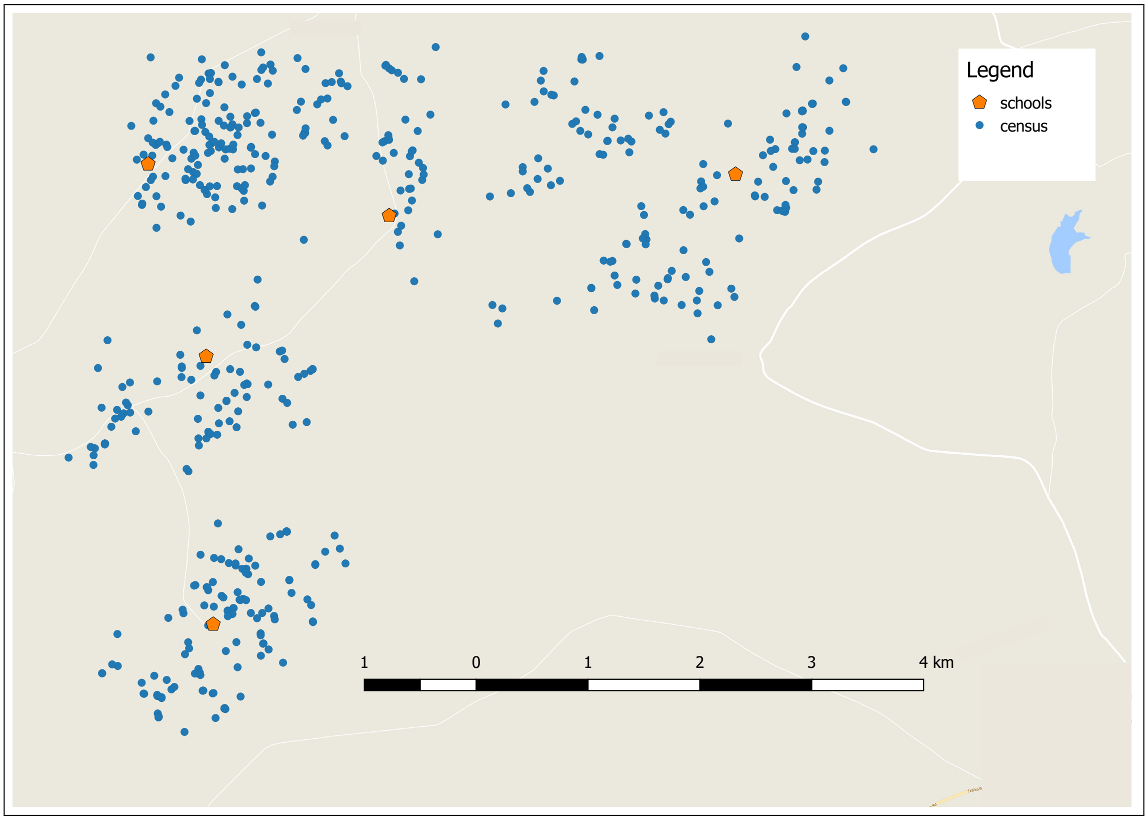



Table 3: Census statistics

| | Tolon | Bongo | All |
|---|---|---|---|
| Households | 4,200 | 5,303 | 9,503 |
| Average Household size | 11.86 | 6.25 | 8.73 |
| Households with children under 10 | 3,935 | 3,982 | 7,917 |
| Total children under 10 | 15,739 | 7,977 | 23,716 |
| Households With TC (age) | 2,887 | 2,032 | 4,919 |
| Total TC (age) | 4,980 | 2,399 | 7,379 |
| Households with TC (Age + KG) | 2,534 | 1,952 | 4,486 |
| Total TC (Age + KG) | 4,154 | 2,292 | 6,446 |

*Note:* TC age criteria between 3 and 5 years, KG criteria is that they either are currently attending or intend to attend one of the 80 study schools

*Sampling*

The final sample of 2,400 TCs as well as their primary caregivers and siblings was drawn from the census sample of 4,486 households with TCs. While the original plan was to sample 30 children per school, during the census it became clear that there was large heterogeneity in the number of students across the schools. Despite an average of around 80 TCs per school, there were some with as few as 11. In order to maximise power, we developed a sampling strategy which ensured that the number of children sampled in each school were as similar as possible. The sample of target children was chosen as follows:

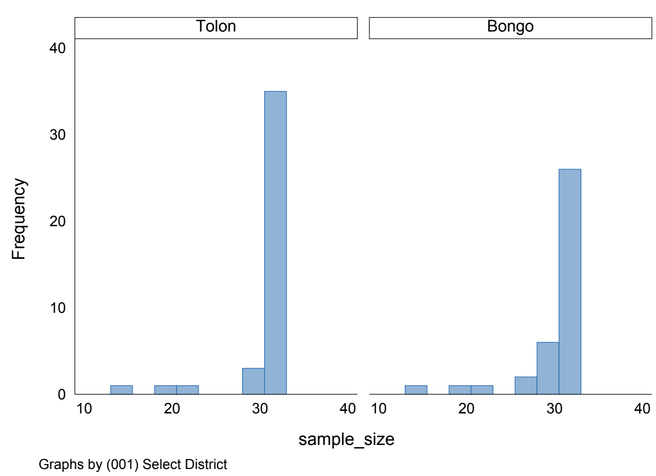
1. For all households with multiple TCs, one of these children was randomly chosen. This is to ensure that all the sampled TCs came from different households, ensuring greater variation and therefore higher power.

2. For all schools with 30 or less TCs (after implementing (1)), all TCs within the school were assigned to be part of the baseline sample.

3. The remaining sample (2400 children minus the amount sampled in step (2)) was split equally among the remaining schools.

As displayed in Figure 7 this led to a final sample of 31 or 32 in most schools. This to make up for the smaller sample sizes in some of the other schools, to keep the average at 30 per school.

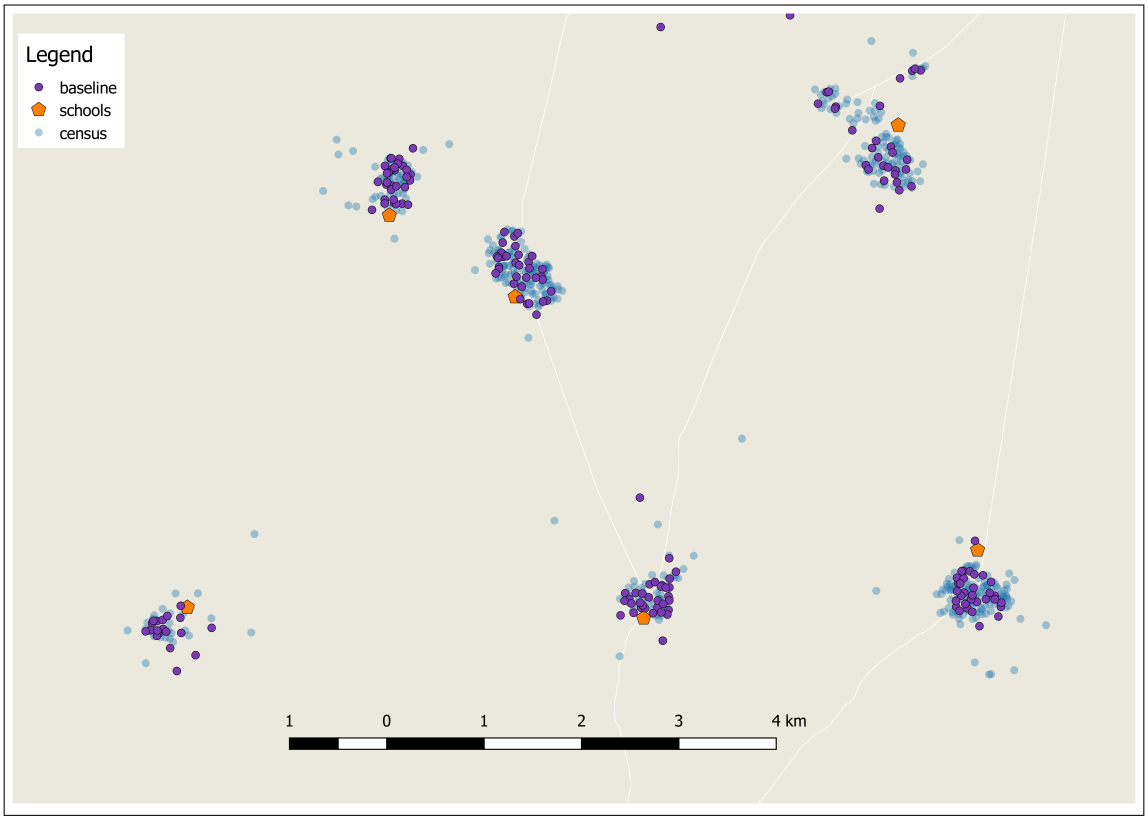
Figure 7: Original distribution of cluster size



A visual representation of how the baseline sample was drawn from the original sample frame is given in Figure 8; the baseline sample provides a representative sample of the full population of TC households from the census.

For every sampled TC, the household they live in was surveyed, as well as their primary caregiver, and older and/or younger sibling if they have one.

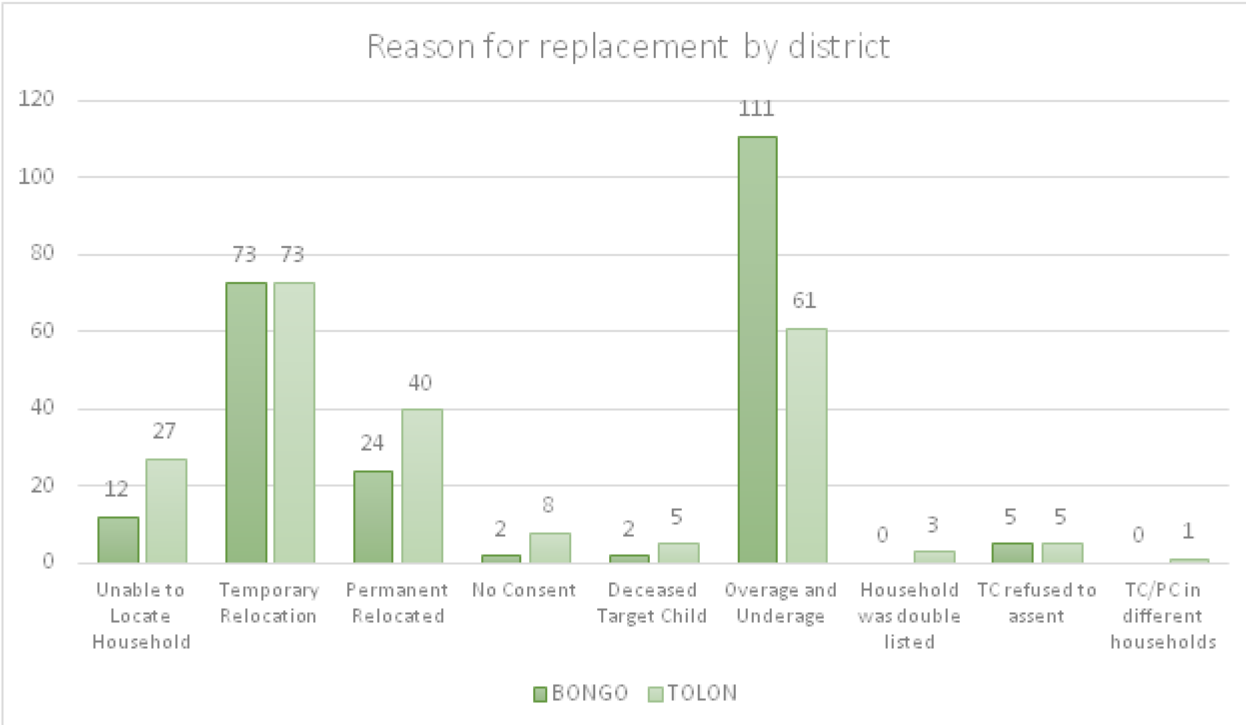
Figure 8: Example baseline sample




### 3.3.2 Replacements

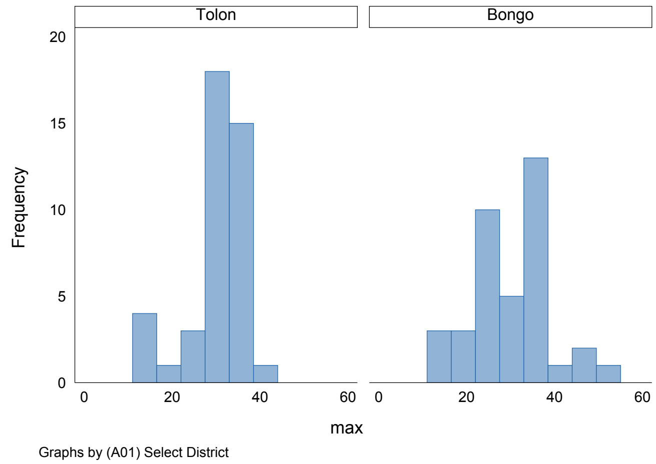
It was not possible to survey all of the original baseline sample. In total, 451 replacements were made. The main reasons for this, as displayed in Figure 9, were temporary relocation or the TC not being in the required age range. The latter occurred due to misreporting in the census, with many individuals not knowing the accurate ages of the children. Parents were asked to provide birth certificates to confirm their child's age, however this was only available for two thirds of the children, leading to some uncertainty.

Figure 9: Reasons for sample replacements



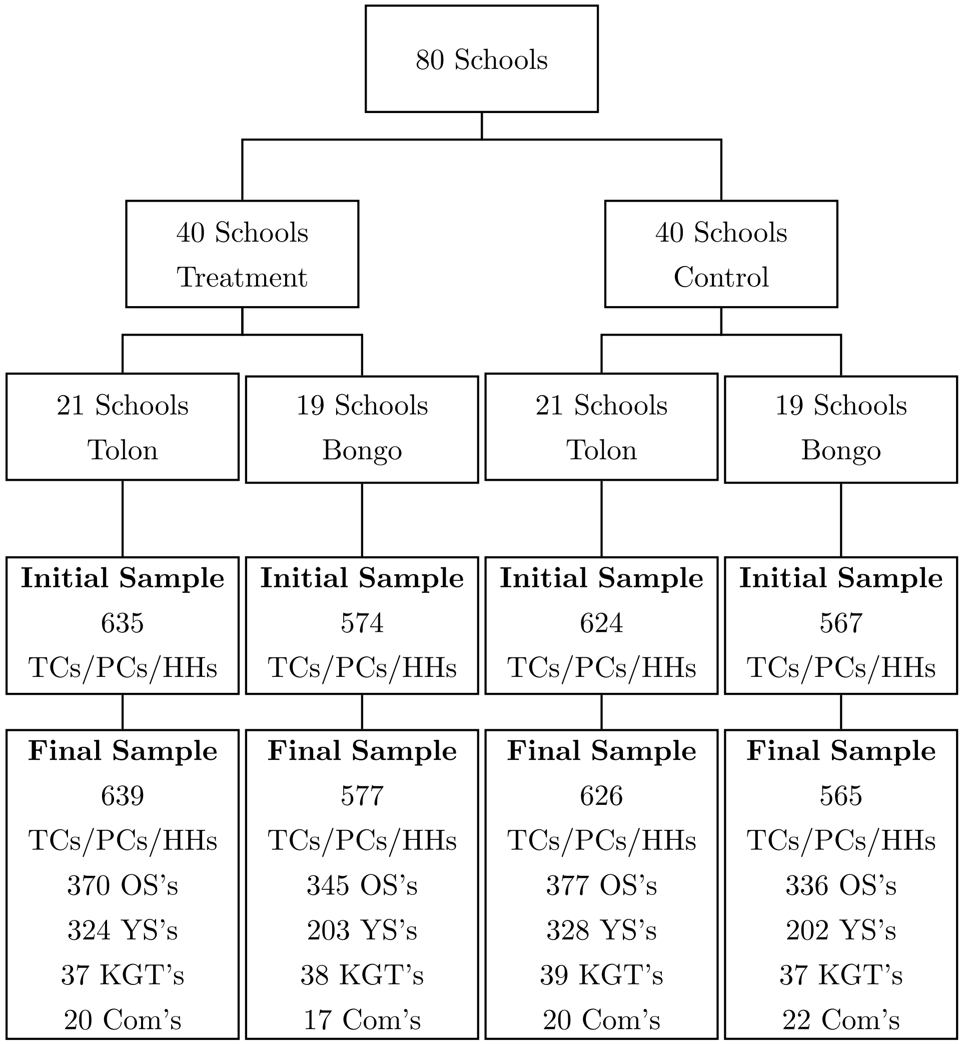

Replacement households were taken from the same school where possible. If there were no replacement households within the same school, one was randomly added from a school within the same treatment group and district. As shown in Figure 10, the need for replacements led to greater heterogeneity in final sample size across schools than shown in Figure 7. This was because some of the schools did not have many households for replacement, so that replacement households had to be drawn from the sub-set of schools with more available replacements. This will lead to a slight reduction in power compared to what would have occurred under equal sample size across clusters.

Figure 10: Final distribution of cluster size




The sampling procedure and final sample numbers are represented in Figure 11

Figure 11: Baseline Sampling




*Note:* TC; Target child, YS; Younger sibling, OS; Older sibling, HHs; Household, KGT's; Kindergarten teachers, Com's ; Communities. See Section 3.2.1 for more information.

### 3.3.3 Timeline

Baseline data collection took place between the 18th of September and 3rd of December 2017. This included surveying a total of 2,407 households, primary caregivers, and target children, 1,428 older siblings, 1,057 younger siblings, 78 communities, 80 schools, and 151 teachers.

Baseline data collection had to be completed within a short rigid time-frame, with start of the school year and start of the VM training sessions constraining the start and end dates, respectively. It was vital to complete the baseline before VM engagement with Lively Minds which would likely 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. 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 LM engagement commenced). Within the two treatment categories, schools were surveyed in a random order. Please see Appendix A.4 for a full timeline of data collection activities.

The household, primary caregiver, IDELA, and sibling surveys were all administered at the homes of the respondents. We decided early on that it was not feasible to survey TCs at school as not all children attend school every day or even at all (as children were classified as TCs on the basis of whether they intended to attend school in September, not whether they actually were) and we did not want there to be systematic differences in the survey environment between kids who did and did not attend school on the day of the survey. Conducting the assessments at the home ensured a consistent testing environment for all TCs. The community surveys were conducted by the team leaders with community elders, whilst the rest of the team were in the community conducting the other surveys. All kindergarten teachers from the 80 study schools were called to a central location in each district to be surveyed. These meetings took place on the 19th September in Bongo and the 22nd September in Tolon, and led to a total of 129 teachers being surveyed. An additional 22 teachers who were not in attendance were surveyed separately at later dates.

See Appendix A.5 for full details of training, staff arrangements, and data quality procedures.

## 3.4 Monitoring data

In addition to the data described above, we are currently collecting monitoring data in order to check the fidelity of intervention implementation and assess the mechanisms underlying its impacts. The data being collected are as follows:

1. *Teacher registers* : teachers in each of the 40 treatment schools keep daily registers. This includes information on whether the play scheme ran, and the attendance of mothers and children.

2. *Lively Minds visits* : throughout the full school year, Lively Minds and GES staff conduct unannounced monthly spot checks on play schemes. During these visits, they record information including whether the Play Scheme took place, the number of children present, the number of volunteer mothers present, whether certain activities took place (hand-washing, outdoor games, mats placed correctly), and subjective assessments of the quality of the play scheme.

3. *IPA visits* : in a similar vein to Lively Minds, IPA are also conducting unannounced spot checks of play schemes. In some cases this will be at the same time as Lively Minds staff and in other cases, IPA will visit alone. The data collected include whether the Play Scheme took place, the number of children present, the number of volunteer mothers present, and whether certain activities took place (hand-washing, outdoor games, mats placed correctly). Each school will be visited at least once over the course of the intervention.

The collection of monitoring data from IPA will allow us to validate the accuracy of reporting from both teachers and Lively Minds.

## 3.5 Challenges and risks

### 3.5.1 Implementation challenges

Lively Minds have had to adapt their implementation approach in order to accommodate the RCT design in ways that present some risks to implementation quality. Specific issues include:

1. *Selection of less remote districts*: Experience has shown that LM is likely to be most effective in remote rural communities, where the need is greatest. However, we were limited in the choice of districts for the RCT as (i) districts needed to have a sufficient number of schools; (ii) schools needed to be within 90 minutes drive from the Lively Minds offices for logistical reasons; and (iii) there was a requirement from one of the implementation funders (UNICEF) to work in the Tolon district. This resulted in selection of two districts (Bongo and Tolon) that are relatively close to the capital in their respective regions and have a far higher proliferation of NGOs than districts that LM are used to operating in.

2. *Need for within circuit randomisation*: The standard LM implementation approach is to include all schools from a circuit (a group of roughly 10 schools)

in the same implementation cycle which is overseen by one circuit supervisor. However, it was not possible to randomise at the circuit level as we would have needed a much larger sample to achieve the required level of power. Schools within a circuit were, therefore, split between treatment and control groups which means that circuit supervisors have had to put in a large amount of effort for just a few schools in their circuit, and monitoring and top-up training have been much more complicated to implement.

3. *Limited attendance and punctuality enforcement capability*: It is normal for Lively Minds to impose a rule that if teachers are absent twice, the school is disqualified from participation in the programme. However, this was not possible in the RCT as we did not have a sufficient number of replacement schools. Therefore the RCT sample of schools could contain schools with low levels of commitment to the programme, which would not exist under the normal implementation.

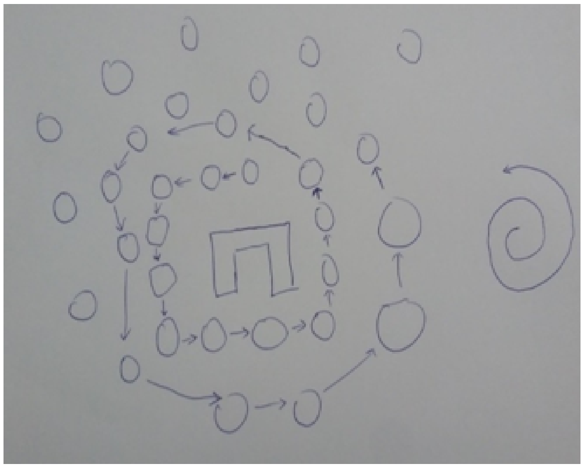
This demonstrates the importance of having an in-depth understanding of the implementation features that are pivotal for sustaining fidelity of the intervention. It is also important to understand if any of these issues will remain throughout the full scale-up of the programme.

### 3.5.2 Defining communities and schools

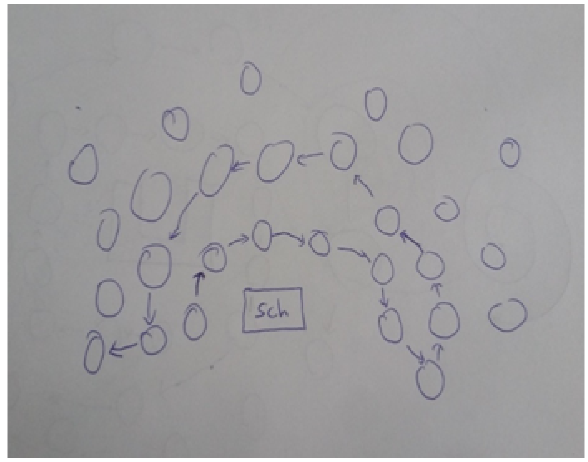
One of the main challenges that was encountered during the baseline data collection was the relationship between schools and communities. It was originally believed that there was a simple mapping between schools and communities with one school per community and all (or great majority) of the children in the community going to that school. If that had been the case, then census sampling could be done at the community level; simply surveying all those living in the 80 communities associated with the 80 study schools.

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 utilising one of four approaches, depending on placement of a school within a community (see Figure 12).

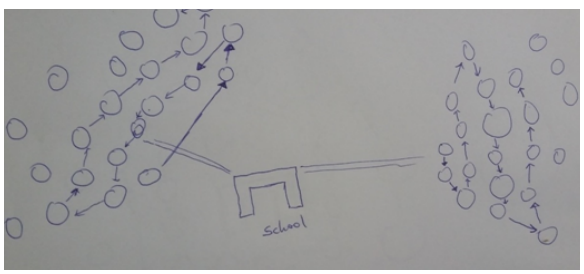
Figure 12: Census sampling techniques



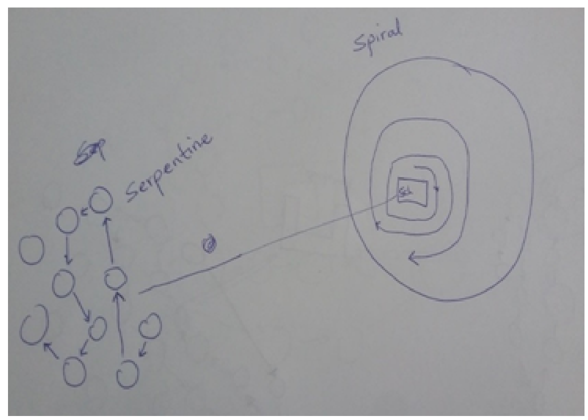
(a) Case 1: Spiral method




(b) Case 2: Serpentine method




(c) Case 3: Multiple serpentine method




(d) Case 4: Combined method

### 3.5.3 Volunteer Mother matching

Although impacts on the TCs is the focus of the evaluation, we are also interested in impacts on the VMs. In order to make sure that we have the necessary statistical power for this analysis we had originally planned to match lists with names of VMs (in treatment schools) and potential VMs [1] (in control schools) to the census list and sample in a way that ensured we included the desired number of VMs. This proved to not be possible as in this context individuals tend to have many different names and VM lists could not be matched to the census list. To overcome this, IPA have returned to the

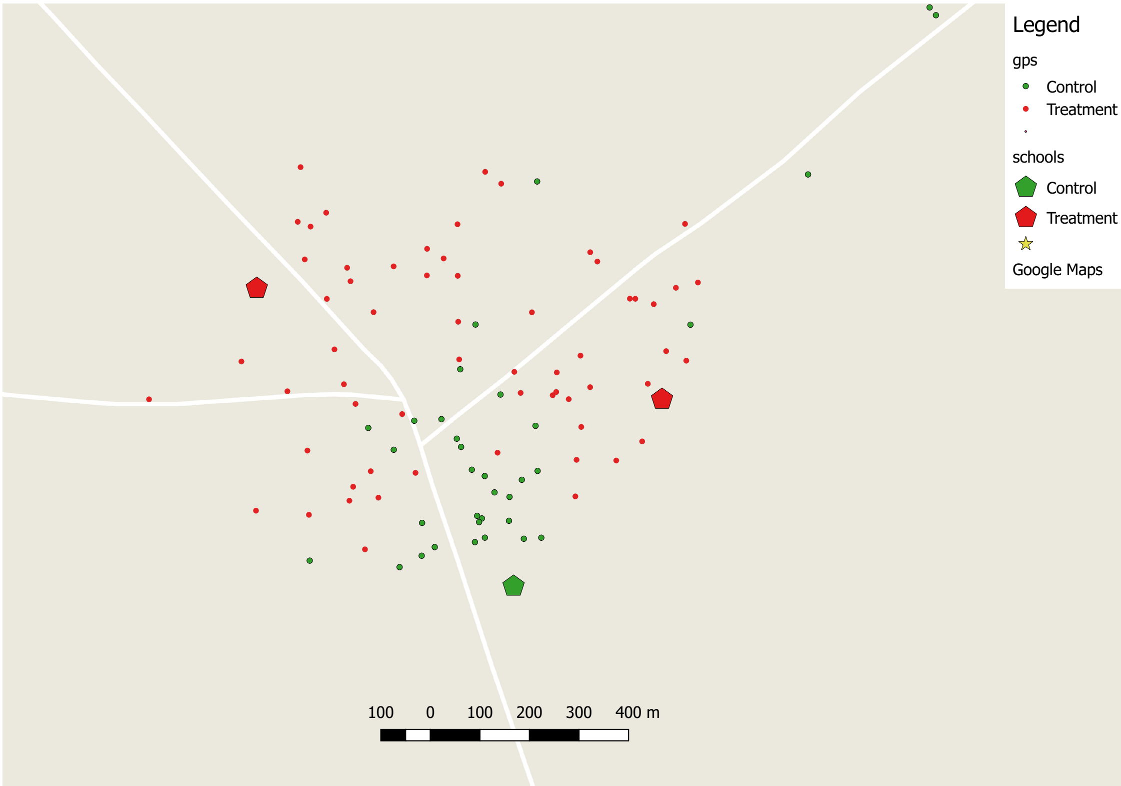
[1] As highlighted in Section 3.2.2, in control communities women who said they would be interested in volunteering in a year's time are defined as potential VMs and suitable comparison group for the actual VMs in treatment communities

field, organised village meetings with the VMs themselves and are conducting manual matching during these meetings. As the matching is ongoing, we do not know at this stage whether enough of the women in the baseline sample are also VMs and, therefore, whether we have the power to identify VM specific impacts. If, once the matching is complete, we do not have a sufficiently large sample, we will consider options for sampling additional VMs during the end-line.

#### 3.5.4 Contamination risk

There is a concern over the potential for contamination between treatment and control groups. As displayed in Figure 13, there are certain (few) cases where treatment and control schools are extremely close which may result in TCs who are attending a treatment school living in extremely close proximity to those in a control school. There could therefore be spillovers from intervention impacts either through direct contact between treatment and control TCs or VMs sharing the knowledge they acquired from the Lively Minds scheme with mothers in the control group. In addition, children who we originally believed to be attending control schools could switch to treatment schools following the intervention, or vice versa. Such contamination would introduce a downward bias to estimates of intervention impacts, through improving the outcomes of the control group. This could not be avoided due to the limited number of schools in the district. However, we do not expect this downward bias to be large as it is an issue for a minority of communities and it is unclear that the intervention has strong potential for spillovers. As a robustness check we will present all main estimates excluding schools where contamination concerns are high.

Figure 13: Contamination concerns




### 3.5.5 Human Resource issues

In general, relationships between the main project partners (IPA, IFS and Lively Minds) have been very good and productive. There has been no substantial staff turnover within either the IFS or the project partners during this time, which has allowed for strong continuity and minimal disruptions to the project.

# 4 Balance tests

## 4.1 Methodology

As explained in Section 3.1, the key assumption that allows for the identification of a causal impact is that in the absence of the Lively Minds programme the treatment and control groups would have identical outcomes. While this assumption cannot be directly tested in the baseline data, we can provide supportive evidence of its validity by testing if there are any significant differences in a range of observable variables between the treatment and control groups. If the randomisation was successful, we would expect to see no evidence of any systematic differences between the treatment and control groups.

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

$$y_{i,j} = \alpha + \beta T_j + s_j + \epsilon_{i,j} \tag{1}$$

Where $y_{i,j}$ is an indicator of interest for individual i in school j, and $T_j$ is an indicator variable equal to 1 if the individual is in a treatment school, $s_j$ is a fixed effect for the randomisation strata 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 (i.e. $\beta = 0$ ). 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. Therefore, the higher the p-value the more similar our study groups are, statistically speaking, for that particular outcome. In this analysis we allow for arbitrary correlation of the errors, $\epsilon_{i,j}$ , at the level of the school. For certain outcomes with large outliers, the means are reported having dropped observations above the 95th percentile.

## 4.2 Results

Overall the balance analysis clearly indicates that the randomisation was successful. Balance tests were ran on 303 individual variables (reported in Section A.7), and of these only 25 (8.3%) were statistically significantly different between treatment and control groups at the 10% level. As this is fewer than would be expected by chance, this provides strong evidence that the treatment and control groups are balanced and therefore that the randomisation has been successful which is a pre-requisite for successful implementation of an RCT. In addition there is balance along our main outcomes of interest; including IDELA scores, primary caregiver mental health, teacher well-being and parental investments. All of the variables that are not balanced (displayed in Table A1) will be included as controls in the endline estimation specification in order to mitigate any remaining concerns and improve the precision of our impact estimates.

# 5 Description of the baseline data

We now proceed to a description of the baseline data. The main aim of this section is to highlight key characteristics of the environment in which the evaluation is taking place.

This is a remote part of Ghana which little is known about and on which there are few data. The data collected at baseline are therefore valuable not only for the evaluation analysis (e.g. to check balance) but also to provide some insight into the specific study context.

## 5.1 Communities

The average community in the sample is reported to have around 1,700 inhabitants and 300 households. Communities are fairly remote, with limited access to public services and institutions; 42% have a bank nearby, and only 22% have a public hospital nearby. The sanitation conditions are fairly poor, just under a third of communities have any form of public toilet in use, and open defecation is common in the vast majority (77%) of communities. Communities are exposed to a large amount of risk of environmental shocks; over half have experienced a drought and/or a flood in the last 4 years. The average daily agricultural wage for men is 11.5 GHS (£1.81).

Table 4: Summary Statistics : Communities

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Number of inhabitants | 2101.26 | 1670.53 | 2532.00 | 0.08* | 80 |
| | (2174.93) | (2061.97) | (2224.77) | | |
| Number of households | 300.11 | 244.15 | 356.07 | 0.16 | 80 |
| | (323.21) | (272.48) | (361.86) | | |
| Bank nearby (%) | 45.00 | 45.00 | 45.00 | 0.93 | 80 |
| | (50.06) | (50.38) | (50.38) | | |
| Public hospital nearby (%) | 21.25 | 22.50 | 20.00 | 0.93 | 80 |
| | (41.17) | (42.29) | (40.51) | | |
| Any public toilet in use (%) | 32.50 | 30.00 | 35.00 | 0.47 | 80 |
| | (47.13) | (46.41) | (48.30) | | |
| Open defecation is common (%) | 77.50 | 75.00 | 80.00 | 0.73 | 80 |
| | (42.02) | (43.85) | (40.51) | | |
| Drought in last 4 years (%) | 53.75 | 62.50 | 45.00 | 0.09* | 80 |
| | (50.17) | (49.03) | (50.38) | | |
| Flood in last 4 years (%) | 53.75 | 57.50 | 50.00 | 0.70 | 80 |
| | (50.17) | (50.06) | (50.64) | | |
| Average male agricultural wage (GHS) | 11.47 | 11.66 | 11.25 | 0.69 | 66 |
| | (4.24) | (3.85) | (4.72) | | |
| Average female agricultural wage (GHS) | 12.69 | 11.01 | 14.52 | 0.25 | 69 |
| | (11.20) | (6.54) | (14.60) | | |

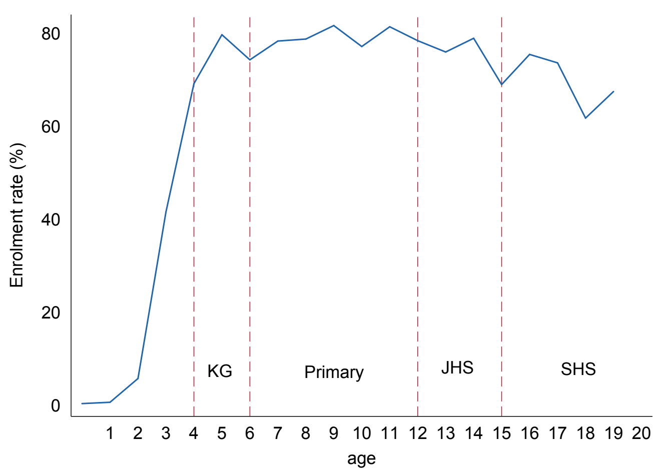
Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.2 School enrolment and facilities

Figure 14 displays the proportion of children of a given age who have attended school in the last 12 months. This shows that for the majority of schooling levels, enrolment is around 80%. For those of KG age; enrolment is 69% for 4 years olds, and 80% for 5 year olds. While official school starting age is 4, there is a substantial number of under aged children enrolled in school: for instance, 42% of 3 year olds attended school in the last 12 months.

Kindergarten education is made up of two grades, KG1 and KG2. Within our sample, there are two teachers in the majority (86%) of cases, and in 68% of schools there are two KG classrooms. Average class sizes are large at 58 children, however there is large variation with class sizes in the sample ranging between 18 and 220. In addition, there is on average only one desk for every 8 pupils. Based on school observations from surveyors; over half of the schools do not provide access to a toilet, or possess any books for KG children, and nearly a half of schools have a major safety hazard [2] Parental reports suggest that two thirds of schools teach in the dominant local language, and the rest in English.

Figure 14: School enrolment by age




*Source:* Information from full household roster. Classified as “enrolled” if attended school in the last 12 months.

[2]Defined as: anything that could cause injury or illness to a child, or adult in the KG. Examples of safety hazards: loose electric cords, medicines, cleaning materials and other substances not locked away, unprotected hot stove or fireplace, play areas in front of doors, 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.

Table 5: Summary Statistics : Schools

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Average no. Pupils per KG class | 57.51 | 60.45 | 54.58 | 0.44 | 80 |
| | (30.16) | (36.16) | (22.74) | | |
| Desks per pupil | 0.12 | 0.09 | 0.16 | 0.09* | 80 |
| | (0.18) | (0.16) | (0.19) | | |
| School has a toilet (%) | 38.75 | 30.00 | 47.50 | 0.10 | 80 |
| | (49.03) | (46.41) | (50.57) | | |
| School has a major safety hazard (%) | 45.00 | 42.50 | 47.50 | 0.71 | 80 |
| | (50.06) | (50.06) | (50.57) | | |
| Books accesible for children (%) | 47.50 | 42.50 | 52.50 | 0.49 | 80 |
| | (50.25) | (50.06) | (50.57) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.3 Teachers

Compared to others in our sample, teachers are highly educated; 86% have a diploma/HND or higher, and 74% have some kind of ECCE training. Teachers are on average 33 years old and only possess 4 years of experience. This suggests high turnover within the profession. Teachers do not have particularly strong ties to to the communities in which they teach; fewer than one in ten were born there and only 17% live there even now. Teachers work 29 hours a week (22 in school and 7 preparing), and report to be paid 959 GHS (£151) per month on average. The main problems cited by teachers in their jobs are a lack of school financial resources and large class sizes. The majority of teachers also report insufficient parental interest and involvement. This highlights the importance of Lively Minds in encouraging greater involvement of volunteer mothers in the running of schools. Depression seems fairly low, with an average SRQ score of 3.72, and less than 10% having a score above 8 (see Section 5.6 for a more detailed discussion of the SRQ measure). A range of other teacher outcomes such as job satisfaction, burnout and motivation were also collected (see Table A17). These are all subjective scales, and do not have any real intuitive interpretation at baseline. However, these measures show good variation so will be used to assess the impacts of the programme at endline.

Table 6: Summary Statistics : Teachers

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Age | 33.05 | 32.61 | 33.49 | 0.60 | 151 |
| | (7.15) | (6.90) | (7.41) | | |
| Tertiary education (%) | 86.09 | 82.89 | 89.33 | 0.31 | 151 |
| | (34.72) | (37.91) | (31.08) | | |
| Some ECD training (%) | 73.51 | 69.74 | 77.33 | 0.21 | 151 |
| | (44.27) | (46.24) | (42.15) | | |
| Total years of experience | 3.72 | 3.34 | 4.09 | 0.36 | 151 |
| | (4.43) | (3.94) | (4.88) | | |
| Born in community (%) | 9.27 | 6.58 | 12.00 | 0.11 | 151 |
| | (29.10) | (24.96) | (32.71) | | |
| Live in community (%) | 17.22 | 15.79 | 18.67 | 0.34 | 151 |
| | (37.88) | (36.71) | (39.23) | | |
| Hours worked (per week) | 29.27 | 29.86 | 28.67 | 0.64 | 150 |
| | (13.13) | (13.24) | (13.09) | | |
| Monthly salary (GHS) | 958.71 | 948.53 | 969.03 | 0.96 | 151 |
| | (364.10) | (374.42) | (355.57) | | |
| Raw Depression (SRQ-20) Score | 3.72 | 3.79 | 3.64 | 0.78 | 151 |
| | (3.55) | (3.80) | (3.30) | | |
| *% agreeing that the following is a problem* | | | | | |
| Parents are not sufficiently interested | 72.19 | 78.95 | 65.33 | 0.02** | 151 |
| | (44.96) | (41.04) | (47.91) | | |
| Parents are not sufficiently actively involved | 77.48 | 82.89 | 72.00 | 0.17 | 151 |
| | (41.91) | (37.91) | (45.20) | | |
| Lack of financial resources | 94.70 | 96.05 | 93.33 | 0.61 | 151 |
| | (22.47) | (19.60) | (25.11) | | |
| Classes are large | 58.28 | 57.89 | 58.67 | 0.95 | 151 |
| | (49.47) | (49.70) | (49.57) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.4 Household economic situation

Households are large and contain on average; 9 total members, 4 children under the age of 17 and 2 children under 6. The main source of income for households is agriculture; 44% of households gain their main income from farming of their own land, whereas 28% earn their living through wage work, largely as labourers on the land of others. Only about a fifth of individuals have some form of small enterprise as their main source of income.

It is very difficult to accurately measure incomes in these contexts. The measures of income collected will, therefore, mainly be used for the purposes of comparing richer and poorer individuals, rather than making any statements on absolute levels of income in the study area. Importantly, these measures of income and expenditure are balanced between treatment and control groups at baseline. Another useful measure of relative wealth is the ownership of assets. The full list of assets and their ownership can be seen in Table A.7.5. These are used to form a wealth index, through a principal component analysis. This is our preferred measure of household socio-economic status, and is also balanced across treatment and control. Using comparable data from the Demographic and Health Survey (DHS) from 2016, socio-economic status as proxied by such a wealth index, is significantly lower in our study sample then Ghana as a whole.[3]

[3] Asset index created using assets surveyed in both baseline and DHS data sets. Principal factor loadings created on baseline data and then applied to DHS data to ensure comparable weights. This leads to wealth index scores of 114 in our sample, compared to 161 for Ghana as a whole.

Table 7: Summary Statistics : Households

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Household Size | 9.20 | 9.14 | 9.25 | 1.00 | 2291 |
| | (4.52) | (4.62) | (4.43) | | |
| Number of children (16 or under) | 3.88 | 3.93 | 3.84 | 0.65 | 2316 |
| | (2.67) | (2.78) | (2.57) | | |
| Number of children (6 or under) | 1.58 | 1.61 | 1.54 | 0.59 | 2314 |
| | (1.54) | (1.59) | (1.50) | | |
| Farming own land (%) | 43.66 | 43.66 | 43.67 | 0.45 | 2407 |
| | (49.61) | (49.62) | (49.62) | | |
| Waged work (%) | 28.00 | 27.62 | 28.37 | 0.40 | 2407 |
| | (44.91) | (44.73) | (45.10) | | |
| Livestock (%) | 3.41 | 3.78 | 3.04 | 0.66 | 2407 |
| | (18.14) | (19.08) | (17.18) | | |
| Profits from small enterprise (%) | 18.70 | 19.40 | 18.01 | 0.96 | 2407 |
| | (39.00) | (39.56) | (38.44) | | |
| Use any source of savings (%) | 51.35 | 56.09 | 46.71 | 0.01** | 2407 |
| | (49.99) | (49.65) | (49.91) | | |
| Use any source of loans (%) | 34.77 | 35.77 | 33.80 | 0.88 | 2407 |
| | (47.64) | (47.95) | (47.32) | | |
| Wealth Index (PCA factor score) | -0.00 | -0.06 | 0.06 | 0.55 | 2394 |
| | (2.11) | (2.05) | (2.17) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

*Note:* Outliers in household size and number of children trimmed

## 5.5 Investments in children

Past research has shown that the measures of parental investments captured by the FCI questionnaire are strong determinants of child development outcomes (Attanasio et al., 2015). However, in the sample households measures of time and material investments captured by the FCI are very low. Only 65% of households have any type of play materials (including household objects that children play with) for their children. Less than a half of households have any homemade toys and less than a quarter have any bought toys. Only 13% of households have conducted any form of play activities with the target child in the last 3 days.

Availability of play materials is also low in these communities; only 12% have a shop in the area which stocks children's toys. There are some toys at the major regional markets, but these are largely restricted to dolls, which are available for an average of 2 GHS (32p).

Table 8: Summary Statistics: Parental investments

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Number of different play materials | 1.15 | 1.07 | 1.22 | 0.15 | 2407 |
| | (1.18) | (1.12) | (1.23) | | |
| Have any homemade toys (%) | 47.20 | 46.26 | 48.11 | 0.49 | 2407 |
| | (49.93) | (49.88) | (49.98) | | |
| Have any bought toys (%) | 23.51 | 21.58 | 25.41 | 0.53 | 2407 |
| | (42.42) | (41.15) | (43.55) | | |
| Number of play activities (TC) | 0.19 | 0.17 | 0.21 | 0.14 | 2407 |
| | (0.59) | (0.55) | (0.62) | | |
| Any play activities in last 3 days (%) | 13.13 | 12.01 | 14.23 | 0.10* | 2407 |
| | (33.78) | (32.52) | (34.95) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.6 Primary caregivers

Primary caregivers are on average 36 years of age[4], and are the biological mother of the target child in 78% of cases. The vast majority (80%) of primary caregivers have never attended any formal schooling, and only 7% ever completed primary school. Primary

[4]It should be noted that as the majority of primary caregivers do not possess birth certificates, so age reports are likely to be inaccurate. Also most primary caregivers have a child older than the target child. Hence they are older then would expect of women who's first child is now between 3 and 5 years of age.

caregivers have relatively low involvement in their child's school: only 41% know the teacher's name (or report any name at all even if incorrect), and 39% visited the school in the last month. With no validation of the SRQ-20 mental health measure for the study area, it is difficult to accurately choose a cut off level to classify someone as at risk of depression. However, in order to give some form of informal indication, cut-offs found in other similar contexts in developing countries can be applied (Chipimo and Fylkesnes, 2010; Scholte et al., 2011; Tuan, Harpham and Huong, 2004). This suggests that a high proportion of caregivers are at risk of depression (around half have an SRQ-20 score of 8 or greater). The SRQ-20 measures also performed well at baseline, with good variation and a high measure of internal consistency. We had hoped to measure parental knowledge of childhood development, through the use of scales developed in other studies. However we found that these did not perform well in these contexts, as there was very little variation in responses to the questions (see Table A33). For endline we will pilot alternative instruments to try and more accurately measure parental knowledge.

Table 9: Summary Statistics: Primary caregiver

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Primary caregiver age | 36.46 | 37.00 | 35.93 | 0.07* | 2407 |
| | (12.26) | (12.41) | (12.09) | | |
| Some education (%) | 20.61 | 20.40 | 20.81 | 0.73 | 2407 |
| | (40.46) | (40.32) | (40.61) | | |
| Is biological mother (%) | 77.61 | 75.90 | 79.28 | 0.29 | 2407 |
| | (41.70) | (42.79) | (40.55) | | |
| Knows teachers name (%) | 41.47 | 38.75 | 44.13 | 0.12 | 1818 |
| | (49.28) | (48.75) | (49.68) | | |
| Visited school last month (%) | 39.14 | 38.20 | 40.07 | 0.48 | 1819 |
| | (48.82) | (48.61) | (49.03) | | |
| Raw Depression (SRQ-20) Score | 8.19 | 8.06 | 8.32 | 0.58 | 2407 |
| | (5.27) | (5.29) | (5.26) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.7 Target children

Target children are on average 56 months old: 78% were at school at the time of the baseline. Health problems were reported for many: 55% were reported to have had a cough in the 30 days before the baseline, half had had stomach pain and a third a fever.

Table 10 shows summary statistics for performance on the IDELA - our main measure of child development - the primary outcome. The IDELA was scored following the guidance of Save the Children, who developed this assessment. Firstly the percentage of questions scored correctly on each item was calculated. Then the items were aggregated to the 5 sub-domains; through calculating an average of the relevant item percentage scores for each domain. Finally, following Behrman, Wolf and Aber (2017), a combined "school readiness" score was calculated which averaged the percentage scores along four domains; socio-emotional, emergent numeracy, emergent literacy and executive function. Overall the IDELA test performed well; there is substantial variation in both the individual domains and most of the specific items. A few of the items were too hard, with average scores less than 10%; items 5 (number identification), 10 (emotional awareness), 17 (letter identification), and 18 (letter sounds). This was largely expected; skills such as number and letter awareness and the understanding of English words is something that can only be learnt in school and is unlikely to have been covered sufficiently in the kindergarten. However performance on these questions could increase by the end of the intervention, so will be kept in for endline in order to pick up such changes. Although it is hard to interpret IDELA scores, as described in Section 2.6, ongoing work is looking to place developmental outcomes in the study regions in the context of previous studies in Ghana.

Table 10: Target child characteristics

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| TC age in months | 56.22 | 56.12 | 56.31 | 0.68 | 2405 |
| | (9.53) | (9.50) | (9.56) | | |
| Currently in school | 0.78 | 0.78 | 0.78 | 0.91 | 2407 |
| | (0.41) | (0.41) | (0.42) | | |
| *% with the following in the last 30 days:* | | | | | |
| Cough | 54.49 | 52.44 | 56.50 | 0.04** | 2406 |
| | (49.81) | (49.96) | (49.60) | | |
| Stomach pain | 50.50 | 48.28 | 52.67 | 0.02** | 2406 |
| | (50.01) | (49.99) | (49.95) | | |
| High fever | 34.03 | 32.07 | 35.94 | 0.18 | 2407 |
| | (47.39) | (46.70) | (48.00) | | |

Table 11: IDELA : Subdomain scores (%)

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Socio-emotional | 27.71 | 27.28 | 28.13 | 0.78 | 2407 |
| | (17.35) | (16.75) | (17.92) | | |
| Emergent Numeracy | 27.89 | 27.61 | 28.17 | 0.91 | 2407 |
| | (14.13) | (13.53) | (14.69) | | |
| Emeregent Literacy | 19.43 | 19.53 | 19.34 | 1.00 | 2407 |
| | (13.13) | (12.69) | (13.54) | | |
| Motor Skills | 29.31 | 28.80 | 29.81 | 0.66 | 2407 |
| | (26.49) | (26.13) | (26.85) | | |
| Executive Function | 36.14 | 36.52 | 35.76 | 0.63 | 2407 |
| | (26.33) | (25.38) | (27.23) | | |
| School readiness | 27.79 | 27.74 | 27.85 | 0.96 | 2407 |
| | (14.61) | (13.91) | (15.27) | | |

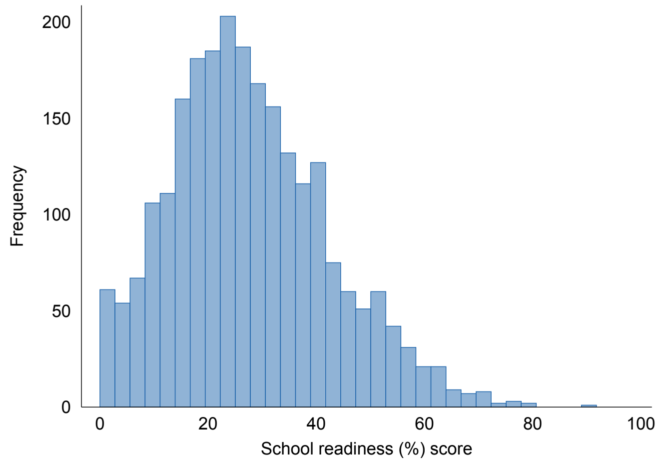
Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: IDELA : Individual item scores (%)

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| *Socio-emotional* | | | | | |
| Self-awareness | 58.16 | 57.46 | 58.83 | 0.72 | 2407 |
| | (26.80) | (26.53) | (27.05) | | |
| Number of friends | 31.17 | 31.44 | 30.91 | 0.84 | 2407 |
| | (26.29) | (26.64) | (25.95) | | |
| Emotional awareness | 6.51 | 5.98 | 7.03 | 0.54 | 2407 |
| | (20.27) | (19.27) | (21.20) | | |
| Empathy/perspective taking | 14.31 | 13.37 | 15.23 | 0.49 | 2407 |
| | (25.73) | (24.49) | (26.87) | | |
| Solving conflict | 28.40 | 28.13 | 28.66 | 0.90 | 2407 |
| | (36.39) | (35.71) | (37.06) | | |
| *Emergent Numeracy* | | | | | |
| Comparison by size and length | 78.99 | 79.93 | 78.06 | 0.27 | 2407 |
| | (26.65) | (26.00) | (27.26) | | |
| Sorting and classification | 24.10 | 23.22 | 24.96 | 0.71 | 2407 |
| | (29.91) | (29.00) | (30.77) | | |
| Shape identification | 29.12 | 28.38 | 29.85 | 0.64 | 2407 |
| | (24.83) | (24.18) | (25.44) | | |
| Number identification | 4.34 | 3.98 | 4.68 | 0.15 | 2407 |
| | (9.87) | (7.89) | (11.48) | | |
| Counting | 16.49 | 16.26 | 16.72 | 0.88 | 2407 |
| | (24.18) | (23.78) | (24.57) | | |
| Addition and subtraction | 29.68 | 30.03 | 29.33 | 0.45 | 2407 |
| | (27.75) | (27.71) | (27.79) | | |
| Puzzle completion | 12.54 | 11.49 | 13.56 | 0.04** | 2407 |
| | (14.88) | (14.14) | (15.52) | | |
| *Emergent Literacy* | | | | | |
| Expressive vocabulary | 25.69 | 25.76 | 25.61 | 0.90 | 2407 |
| | (20.98) | (21.11) | (20.87) | | |
| Print awareness | 25.90 | 26.22 | 25.58 | 0.88 | 2407 |
| | (30.49) | (30.80) | (30.20) | | |
| Letter identification | 2.33 | 1.57 | 3.08 | 0.00*** | 2407 |
| | (9.65) | (6.28) | (12.03) | | |
| First letter sounds | 9.51 | 9.74 | 9.29 | 0.98 | 2407 |
| | (18.94) | (18.94) | (18.95) | | |
| Emergent writing | 13.50 | 13.75 | 13.26 | 0.95 | 2407 |
| | (23.18) | (23.02) | (23.34) | | |
| Oral comprehension | 39.66 | 40.12 | 39.21 | 0.41 | 2407 |
| | (32.61) | (32.30) | (32.92) | | |
| *Motor Skills* | | | | | |
| Copying a shape | 35.71 | 35.81 | 35.61 | 0.89 | 2407 |
| | (41.07) | (40.62) | (41.52) | | |
| Drawing a person | 26.48 | 26.36 | 26.59 | 0.98 | 2407 |
| | (31.04) | (31.01) | (31.08) | | |
| Folding paper | 25.74 | 24.22 | 27.22 | 0.28 | 2407 |
| | (26.08) | (25.37) | (26.69) | | |
| *Executive Function* | | | | | |
| Short term memory | 53.94 | 55.21 | 52.69 | 0.27 | 2407 |
| | (30.42) | (30.02) | (30.77) | | |
| Inhibitory control | 32.13 | 32.12 | 32.14 | 0.93 | 2407 |
| | (40.34) | (39.54) | (41.12) | | |
| Pencil tapping task | 22.34 | 22.24 | 22.45 | 0.92 | 2407 |
| | (33.00) | (32.65) | (33.34) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 15: Distribution of IDELA school readiness score




## 5.8 Siblings

Literacy and numeracy outcomes are particularly poor among older siblings, aged between 6 and 10 years of age. Only around one quarter of surveyed older siblings could correctly solve simple addition problems, such as 7 + 4, or read the single letter "r". Only just over 5% could read a single English word such as "us" and less than 2% could read a simple sentence in English. A range of other outcomes were also collected for older and younger siblings, these can be seen in Table A41.

Table 13: Summary Statistics: Older siblings

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Addition: 7 + 4 | 24.72 | 26.83 | 22.63 | 0.16 | 1428 |
| | (43.15) | (44.34) | (41.87) | | |
| Multiplication: 4 * 3 | 22.20 | 21.49 | 22.91 | 0.76 | 1428 |
| | (41.57) | (41.10) | (42.05) | | |
| Read : r | 28.08 | 26.40 | 29.75 | 0.26 | 1428 |
| | (44.96) | (44.11) | (45.75) | | |
| Read : Us | 5.60 | 5.34 | 5.87 | 0.53 | 1428 |
| | (23.00) | (22.49) | (23.52) | | |
| Read : John is sick | 1.96 | 1.83 | 2.09 | 0.44 | 1428 |
| | (13.87) | (13.40) | (14.33) | | |

*Note:* % of sample who answer questions correctly

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.9 W.A.S.H practices

Understanding of handwashing practices is very good among primary caregivers. Nearly all could state a time when handwashing is needed, that water and soap are needed for washing hands, and could provide a reason of why hand-washing is important. In comparison, knowledge of hand-washing is a lot poorer among target children; only around half of children can state a good time to wash their hands, or recognise that water or soap is needed, with only 22% correctly understanding why handwashing is important.

Table 14: W.A.S.H outcomes

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| *Primary caregiver* | | | | | |
| Water is needed | 86.59 | 87.32 | 85.88 | 0.97 | 2387 |
| | (34.08) | (33.29) | (34.84) | 0.72 | 1769 |
| Soap is needed | 98.58 | 98.82 | 98.35 | 0.81 | 2398 |
| | (11.83) | (10.80) | (12.75) | 0.59 | 1793 |
| Example of why handwashing is important | 98.46 | 98.23 | 98.68 | 0.35 | 2401 |
| | (12.32) | (13.19) | (11.41) | 0.96 | 1215 |
| A time when handwashing is needed | 88.91 | 89.08 | 88.73 | 0.65 | 2407 |
| | (31.41) | (31.20) | (31.63) | 0.95 | 2407 |
| *Target child* | | | | | |
| Water is needed | 59.98 | 57.06 | 62.94 | 0.97 | 2387 |
| | (49.01) | (49.53) | (48.32) | 0.72 | 1769 |
| Soap is needed | 73.84 | 72.73 | 75.00 | 0.81 | 2398 |
| | (43.96) | (44.56) | (43.33) | 0.59 | 1793 |
| Example of why handwashing is important | 22.14 | 20.92 | 23.38 | 0.35 | 2401 |
| | (41.54) | (40.70) | (42.36) | 0.96 | 1215 |
| A time when handwashing is needed | 52.22 | 52.06 | 52.38 | 0.65 | 2407 |
| | (49.96) | (49.98) | (49.96) | 0.95 | 2407 |

*Note:* Variables report the % correctly stating the following:

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 5.10 District differences

Despite lying in two neighbouring regions of Northern Ghana, the context in the two study districts; Bongo and Tolon, is very different across a number of domains. It should be noted that we do not possess representative samples of the populations in these two districts[5], and there are likely some differences in the composition of the sample across districts.[6] As a result the specific numbers should not be taken as representative or directly comparable across districts, however the size of the differences across districts highlights interesting contextual factors to be aware of when assessing the likely impacts of the intervention across different areas and cultures.

The first main difference between Bongo and Tolon is in the predominant religion. In Tolon the vast majority of households (92%) are Muslim, with 6% traditional African, and 2% Christian. In comparison, in Bongo 46% are Christian, 45% are of traditional African religion and 8% are Muslim. The household structure is also very different. In Tolon the average household size is 13, whereas in Bongo it is only 7. There are three main reasons for this difference; (i) 42% of households are polygamous in Tolon, whereas only 3% are in Bongo, (ii) household have an average of 5 children in Tolon and only 3 in Bongo, (iii) other family members such as the households heads siblings are more likely to live in the household in Tolon than Bongo.

In terms of economic status, the baseline data suggest that those in Tolon are on average richer than those in Bongo. Reported average monthly expenditures are 12% higher in Tolon, and the ownership of most assets, and the constructed wealth index, significantly higher. However due to the large measurement error in assessing income and expenditure, it is difficult to accurately quantify these income differentials.

The relative empowerment, education and psychological wellbeing of women also appears to show large differences across the two regions. In Tolon only 7% of primary caregivers have ever attended school, whereas in Bongo 36% have. In Bongo, 78% women report to being involved in schooling decisions for their child, and 65% report to being involved in the decision to become pregnant with another child. In Tolon the corresponding figures are as low as 42% and 34% respectively. The mental health of women also seems to be better in Bongo then Tolon, with an average SRQ-20 score of 6 in Bongo compared to 10 in Tolon.

The educational situation is also largely different across districts. In terms of resources, schools in Bongo are almost four times more likely to have books accessible for KG children. In addition, teachers report to working substantially more hours, are far

[5] We took a representative sample of households with target children, living within a close radius to the school; this need not be representative of the population in these areas as a whole

[6] As shown in Section 3.3.1, the census sample was a lot greater in Tolon than it was in Bongo. This is despite schools being of a similar size, and therefore the school age population being similar. This suggests that the proportion of children covered in our sample in Tolon is higher than in Bongo. In particular, these households could be missing in Bongo in some non random way, which would change the composition of the sample.

more likely to be born in the community and have lower depression scores. The developmental outcomes of children seems to show significant variation, with IDELA scores 5-10 percentage points higher in Bongo along all domains. Suggested explanations from Lively Minds staff in Ghana for this fact include the activities of the Catholic church and NGOs to improve schooling, and the discipline, attendance, and regulation of district GES officials. The findings on the differing psychological outcomes and empowerment of women provide another potential explanation for these differences.

Therefore, despite being an overall poorer district, both children and mothers in Bongo have a higher baseline level on the main outcomes of interest. This combined with the differing cultural context means that the effects seen at endline across districts could potentially be very different.

Table 15: District differences : household members

| | Total | Tolon | Bongo | p-value | N |
|---|---|---|---|---|---|
| Muslim (%) | 52.01 | 91.78 | 7.97 | 0.00*** | 2407 |
| | (49.97) | (27.48) | (27.09) | | |
| Christian (%) | 23.22 | 2.37 | 46.32 | 0.00*** | 2407 |
| | (42.23) | (15.22) | (49.89) | | |
| Traditional African religion (%) | 24.30 | 5.85 | 44.75 | 0.00*** | 2407 |
| | (42.90) | (23.48) | (49.74) | | |
| Household Size | 9.20 | 11.75 | 6.62 | 0.00*** | 2291 |
| | (4.52) | (4.70) | (2.38) | | |
| Polygamous household (%) | 31.48 | 46.40 | 4.49 | 0.00*** | 1814 |
| | (46.46) | (49.89) | (20.72) | | |
| Number of children (16 or under) | 3.88 | 5.29 | 2.44 | 0.00*** | 2316 |
| | (2.67) | (2.85) | (1.43) | | |
| Number of children (6 or under) | 1.58 | 2.40 | 0.74 | 0.00*** | 2314 |
| | (1.54) | (1.65) | (0.81) | | |
| Total monthly exp (GHS) | 166.39 | 175.05 | 156.82 | 0.01** | 2405 |
| | (183.26) | (205.29) | (154.82) | | |
| Total monthly income (GHS) | 176.59 | 186.89 | 165.56 | 0.27 | 2355 |
| | (565.29) | (421.00) | (687.07) | | |
| Wealth Index (PCA factor score) | -0.00 | 0.85 | -0.94 | 0.00*** | 2394 |
| | (2.11) | (2.02) | (1.79) | | |
| *Primary Caregivers* | | | | | |
| Some education (%) | 20.61 | 6.88 | 35.81 | 0.00*** | 2407 |
| | (40.46) | (25.32) | (47.97) | | |
| Involved in schooling decisions (%) | 59.08 | 41.90 | 78.11 | 0.00*** | 2407 |
| | (49.18) | (49.36) | (41.37) | | |
| Involved in pregnancy decisions (%) | 48.36 | 33.75 | 64.54 | 0.00*** | 2407 |
| | (49.98) | (47.31) | (47.86) | | |
| Raw Depression (SRQ-20) Score | 8.19 | 9.85 | 6.36 | 0.00*** | 2407 |
| | (5.27) | (5.37) | (4.50) | | |
| Ravens score | 5.21 | 4.77 | 5.69 | 0.00*** | 2407 |
| | (1.98) | (1.93) | (1.93) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: District differences: education

| | Total | Tolon | Bongo | p-value | N |
|---|---|---|---|---|---|
| Socio-emotional (%) score | 27.71 | 25.03 | 30.68 | 0.00*** | 2407 |
| | (17.35) | (16.31) | (17.99) | | |
| Emergent Numeracy (%) score | 27.89 | 23.99 | 32.21 | 0.00*** | 2407 |
| | (14.13) | (12.37) | (14.69) | | |
| Emeregent Literacy (%) score | 19.43 | 17.78 | 21.26 | 0.00*** | 2407 |
| | (13.13) | (12.14) | (13.91) | | |
| Motor Skills (%) score | 29.31 | 23.82 | 35.39 | 0.00*** | 2407 |
| | (26.49) | (24.18) | (27.60) | | |
| Executive Function (%) score | 36.14 | 32.26 | 40.43 | 0.00*** | 2407 |
| | (26.33) | (25.89) | (26.14) | | |
| Total (%) score | 27.79 | 24.77 | 31.14 | 0.00*** | 2407 |
| | (14.61) | (13.72) | (14.85) | | |
| Average no. Pupils per KG class | 58.75 | 61.02 | 56.23 | 0.26 | 2407 |
| | (30.67) | (31.91) | (29.03) | | |
| Schools with books accesible for children (%) | 47.86 | 21.03 | 77.58 | 0.00*** | 2407 |
| | (49.96) | (40.77) | (41.72) | | |
| School has a toilet | 38.01 | 33.20 | 43.35 | 0.48 | 2407 |
| | (48.55) | (47.11) | (49.58) | | |
| Teacher has tertiary education (%) | 87.27 | 85.06 | 89.71 | 0.36 | 2407 |
| | (23.71) | (26.27) | (20.22) | | |
| Teachers depression Score | 3.87 | 4.66 | 2.99 | 0.00*** | 2407 |
| | (2.56) | (2.70) | (2.08) | | |
| Total work hours by teacher | 28.88 | 24.54 | 33.67 | 0.00*** | 2407 |
| | (10.29) | (8.46) | (10.01) | | |
| Teacher born in community (%) | 8.81 | 1.30 | 17.12 | 0.00*** | 2407 |
| | (22.18) | (7.97) | (28.90) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

# 6 Conclusion and Policy Recommendations

## 6.1 Key constraints and policy recommendations

The baseline data highlight some of the major factors that are likely to reduce early childhood development outcomes in these contexts. These are as follows:

1. *Schools are under resourced and learning outcomes are low* ; Average class sizes in the kindergarten are extremely large at 58, and schools, particularly in Tolon, often lack books and other important resources for children. These problems extend into later ages, with only a quarter of 6-10 year olds being able to do simple addition or read even a single letter. As a result, children within the sample schools are likely to lag substantially behind what would be expected of children their age.

2. *Parental "investments" in their children are low* : The majority of primary caregivers have not conducted any form of play activity with their child in the last 3 days, and less than half have any form of play materials, even those that are home made. This could be partially due to a lack of availability of materials, only 12% of communities have a shop that stocks toys nearby, and even these are largely restricted to dolls. Investments such as books in the household have been shown to be strong predictors of child development in the Ghanaian context (Wolf and McCoy, 2017).

3. *Parental involvement in school is low* : Less than half of primary caregivers know the name of their child's Kindergarten teacher, and many teachers report the lack of parental involvement as a significant issue. This is important as parental involvement has been shown to be a strong determinant of student achievement within developing countries (Islam, 2017).

4. *Parental educational levels are low*: Only 20% of primary caregivers in our sample have ever attended formal schooling, therefore knowledge on the best educational and stimulation practices could be low. Parental education levels are highly linked to cognitive development in early childhood (Schady et al., 2015).

5. *Low socio-economic status*: Agriculture is the main source of income for the majority of households in the sample, and an average daily agricultural wage is only £1.81. Ownership of basic durables is low, with less than a half owning a bed,

or table and chair. Wealth indices constructing from these assets show that the study districts are relatively deprived compared to the national average. This is likely highly important given the large correlation between socio-economic status and early childhood development outcomes (Rubio-Codina et al., 2015).

6. *Maternal mental health issues are highly prevalent* : By any prevailing cut-off used for the SRQ-20 measure, there is a high risk of depression among primary caregivers in this sample. This is likely a major constraint given the strong correlation of maternal depression and child developmental outcomes (Parsons et al., 2012).

Successful interventions should look to tackle all of these issues; not only improving learning outcomes within school, but focusing on the home environment; addressing parental engagement, investments and mental well-being. As highlighted in the theory of change, these are all areas which we believe will be effected by the Lively Minds intervention. Through its holistic approach focusing on teachers, children and mothers, the Lively Minds intervention is well placed to tackle the main constraints to improved childhood development outcomes in this context. Therefore the baseline data provide strong support for the approach of the intervention and the underlying theory of change.

## 6.2 Lessons for scaling of the intervention

At this baseline stage, it is too early to derive strong lessons with regard to scaling of the intervention. The baseline data nevertheless reveal some contextual characteristics of the study area that are interesting to consider.

On average, the data confirm the validity of several contextual assumptions on the basis of which the scalability of the model is based. For instance, scalability is based on the assumption that the intervention makes use of relatively cheap widely available resources such as local low-educated women rather than expensive, scarce professionals. The vast majority of primary caregivers (potential volunteer mothers) in our sample have never attended any formal education. This confirms that the impacts we will measure in the context of this study will come from a scalable model that does not rely on qualified VMs. Similarly, we find that in our sample there are generally no shops nearby to purchase toys, confirming the need for a scalable model that makes use of locally available material and self-made toys/games such as that of Lively Minds. We also observe that household dwellings in our study area, particularly in Bongo, are spread out, which would make alternative parenting interventions based on home visits indeed too expensive to be scalable.

However, underlying these average statistics we highlight stark baseline differences between the two study districts, Bongo and Tolon. These are in terms of culture, socio-economic status, religion and household structure, but also interestingly in developmental outcomes for children and mental health of primary caregivers. These regional differences raise the question of whether the same intervention model is equally applicable and effective across the country or whether context-specific adaptations will need to be made when going to scale. Not only is it possible that impacts of the intervention depend on baseline levels of (intermediate and final) outcomes, intervention modalities that work in particular settings might not be sustainable in others. For example, the intervention in its current form is based on the assumption that local low-educated women are willing to volunteer their services for free. The baseline data highlight regional differences in terms of household wealth, suggesting that the opportunity cost for local women to participate in the intervention might differ across regions. This in turn might affect the extent to which the intervention in its current form can be effectively scaled up in all parts of the country. As part of our study, we will assess whether the impacts of LM's model are robust to such regional baseline differences.

# References

**Attanasio, Orazio, Raquel Bernal, Marcos Vera-Hernández, and X Peña.** 2016. "The Effects of the Transition from Home-based Childcare to Childcare Centers on Children's Health and Development in Colombia." *mimeo*.

**Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina.** 2015. "Estimating the production function for human capital: Results from a randomized control trial in Colombia." National Bureau of Economic Research.

**Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu.** 2017. "Persistence and fadeout in the impacts of child and adolescent interventions." *Journal of Research on Educational Effectiveness*, 10(1): 7–39.

**Barnett, W. Steven, and Leonard N. Masse.** 2007. "Comparative benefit–cost analysis of the Abecedarian program and its policy implications." *Economics of Education Review*, 26(1): 113–125.

**Behrman, R., J, S Wolf, and J Aber, L.** 2017. "The Impacts of Teacher Training and Parental Education on Kindergarten Quality in Ghana."

**Berlinski, Samuel, Sebastian Galiani, and Paul Gertler.** 2009. "The effect of pre-primary education on primary school performance." *Journal of public Economics*, 93(1-2): 219–234.

**Bernal, Raquel, and Camila Fernández.** 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.

**Cawley, John, James Heckman, and Edward Vytlacil.** 2001. "Three observations on wages and measured cognitive ability." *Labour Economics*, 8(4): 419–442.

**Chipimo, Peter J, and Knut Fylkesnes.** 2010. "Comparative validity of screening instruments for mental distress in Zambia." *Clinical practice and epidemiology in mental health: CP & EMH*, 6: 4.

**Frongillo, E. A., S. M. Sywulka, and P. Kariger.** 2003. "UNICEF psychosocial care indicators project." *Final report to UNICEF. Ithaca: Division of Nutritional Sciences, Cornell University*, 7.

**Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor.** 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." *Science*, 344(6187): 998–1001.

**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.

**Goodman, Robert.** 1997. "The Strengths and Difficulties Questionnaire: a research note." *Journal of child psychology and psychiatry*, 38(5): 581–586.

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

**Heckman, James J., Jora Stixrud, and Sergio Urzua.** 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411–482.

**Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz.** 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of Public Economics*, 94(1-2): 114–128.

**Islam, Asadul.** 2017. "Parental Involvement in Education: Evidence from Field Experiments in Developing Countries." Monash University, Department of Economics 02-17.

**McCoy, Dana Charles, Christopher R. Sudfeld, David C. Bellinger, Alfa Muhihi, Geofrey Ashery, Taylor E. Weary, Wafaie Fawzi, and Günther Fink.** 2017. "Development and validation of an early childhood development scale for use in low-resourced settings." *Population health metrics*, 15(1): 3.

**Parsons, Christine E., Katherine S. Young, Tamsen J. Rochat, Morten L. Kringelbach, and Alan Stein.** 2012. "Postnatal depression and its effects on child development: a review of evidence from low- and middle-income countries." *British Medical Bulletin*, 101: 57–79.

**Pisani, Lauren, Ivelina Borisova, and Amy Jo Dowd.** 2015. "International development and early learning assessment technical working paper." Save the Children.

**Raven, J. C.** 1936. "The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive Mental Tests Used in Genetic Studies." PhD diss. University of London (King's College).

**Rosenberg, Morris.** 1965. *Society and the adolescent self-image.* Princeton university press.

**Rosero, José, and Hessel Oosterbeek.** 2011. "Trade-offs between different early childhood interventions: Evidence from Ecuador."

**Rubio-Codina, Marta, Orazio Attanasio, Costas Meghir, Natalia Varela, and Sally Grantham-McGregor.** 2015. "The Socioeconomic Gradient of Child Development: Cross-Sectional Evidence from Children 6–42 Months in Bogota." *Journal of Human Resources*, 50(2): 464–483.

**Schady, Norbert, Jere Behrman, Maria Caridad Araujo, Rodrigo Azuero, Raquel Bernal, David Bravo, Florencia Lopez-Boo, Karen Macours, Daniela Marshall, Christina Paxson, and Renos Vakis.** 2015. "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries." *Journal of Human Resources*, 50(2): 446–463.

**Scholte, Willem F, Femke Verduin, Anouk van Lammeren, Theoneste Rutayisire, and Astrid M Kamperman.** 2011. "Psychometric properties and longitudinal validation of the self-reporting questionnaire (SRQ-20) in a Rwandan community setting: a validation study." *BMC medical research methodology*, 11(1): 116.

**Tuan, Tran, Trudy Harpham, and Nguyen Tuo Huong.** 2004. "Validity and reliability of the self-reporting questionnaire 20 items in Vietnam." *Hong Kong Journal of Psychiatry*, 14(3): 15.

**Van der Gaag, Jacques.** 2010. "From child development to human development." In *Development and Prevention of Behaviour Problems.* 231–242. Psychology Press.

**Wolf, Sharon, and Dana Charles McCoy.** 2017. "Household Socioeconomic Status and Parental Investments: Direct and Indirect Relations With School Readiness in Ghana." *Child Development.*

# A Appendices

## A.1 School choice and replacement

*School selection process*

District authorities were asked to identify schools with a KG. They mobilised the Head Teacher and a member of the Parent Teacher Assembly to come to a meeting where the play scheme was explained. Interested schools were asked to submit an application. District officers selected by the Director and LM met to review the applications.

Schools were eligible for the RCT if they hadn't been involved with LM or been previously rejected from the LM programme in the past. The latter condition could be relaxed if schools provided evidence about their readiness to take on the programme.

After fulfilling that condition, the following was considered: 1) the number of communities served by the school (schools serving only one community were preferred); 2) the number of children enrolled in KG (on average more than 50); 3) the number of households served by the school (on average more than 150) and 4) distance to urban areas (i.e., being not too close to markets where women have competing demands and are unlikely to have time to volunteer). Schools were also considered if the eligibility conditions were not fulfilled at the moment of the application but sometime before the implementation of the programme.

*School replacement*

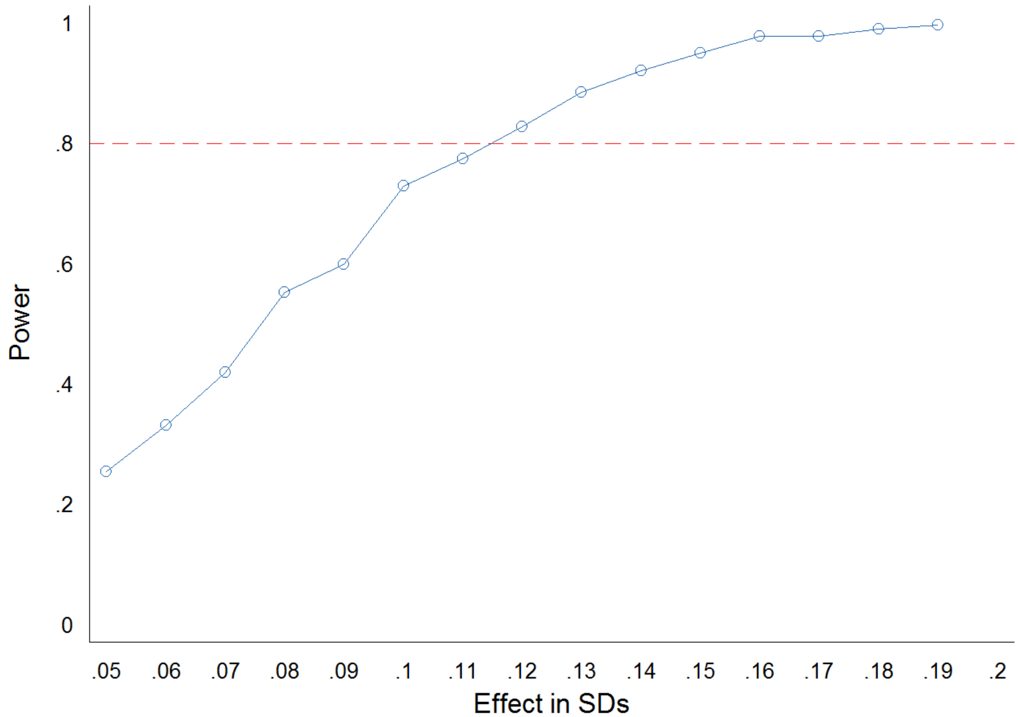
There were difficulties with two schools, one that was inaccessible due to weather conditions (e.g., flooding), and another one that became non-functional. Nevertheless, the actual number of replacements was larger due to the strategy followed. Ideally, to keep the geographical balance of the sample, a school would be replaced by another one with the same status (control or treated), within the same circuit and within the same strata (large or small school). However, the spare schools surveyed in the census did not always fulfil those requirements (i.e., no schools within the same circuit, nor with the same school size). In the latter cases, the school and its respective match (either treated or control) within the same stratum were dropped. To replace them, a new pair of schools (treated and a control) within the same stratum was added. For one school in one of the districts (Bongo), it was not possible to make the replacement as planned. After dropping the pair within the stratum, the number of schools in Bongo was reduced to 38. To keep the initial number of schools (80), a new pair of schools within the same stratum was selected in Tolon, the district that ended up with 42 schools.

## A.2 Power calculations

For the calculation of the MDEs, 40 schools (the unit of cluster) were randomly drawn, with replacement, separately from each of treatment and control groups to form a total of 1000 bootstrapped samples each containing an equivalent number of treatment and control clusters as in the original data. The midline values of the primary outcome variable (school readiness scores from IDELA) were then used to assign an age effect to both groups in each of the samples; this age effect was assumed to be normally distributed with mean 0.07 and standard deviation 0.14 (which reflected the observed differences in scores between midline and baseline in the control group). In the treatment group, a hypothesised treatment effect was also added to scores. Treatment effects were simulated with various values for the mean (all of which were a decimal fraction of the baseline standard deviation of scores), but with the final value of the outcome variable censored at 100%.

OLS regressions were then run in each of these 1000 simulated datasets, with controls for baseline value of the outcome variable, as well as strata and district fixed effects. Power displayed in Figure A1 the above graph is the proportion of these 1000 regressions which then correctly rejected the null hypothesis that treatment effects are equal to zero.

Figure A1: Power calculations




## A.3 Record of changes made to IDELA

The IDELA was extensively piloted, which led to a few minor changes being made to the original IDELA assessment :

- Item 1(f) was dropped: almost all children got this question wrong during the two pilots.
- Item 7: Apples were replaced with mangoes as children in the study area were not familiar with apples.
- Item 8: We added a four-piece puzzle to serve as an example for the six-piece puzzle.
- Item 11: Picture of a cartoon child crying was changed to an African boy crying
- Item 21: the project adapted a square shape instead of a triangle as most children in this setting were more familiar with a square shape than with a triangle.
- Item 24: The hopping question was replaced with a pencil tap game.

- Item 25: this is a health and hygiene question that was added to the usual IDELA items.

## A.4 Timeline

| EVALUATION TIMELINE | July | | | | | | | | | | | August | | | | | | | | | | | | | | | | | | | | September | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Activity** | **17** | **18** | **19** | **20** | **21** | **24** | **25** | **26** | **27** | **28** | **31** | **1** | **2** | **3** | **4** | **7** | **8** | **9** | **10** | **11** | **14** | **15** | **16** | **17** | **18** | **21** | **22-25** | **28** | **29** | **30** | **31** | **1** | **4** | **5** | **6** | **7** | **8** | **9** | **11** | **12** | **13** | **14** | **15** | **18** | **19** | **20** | **21** |
| Piloting of baseline survey instruments in CAPI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Incorporate feedback from piloting into programmed surveys | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA pilot in CAPI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Incorporate feedback from piloting IDELA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Older siblings' development and pilot | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Finalise baseline survey instruments for training | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sampling for baseline survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Recruitment of field staff (initial and adjustment) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Develop detailed field plans | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Write data check codes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Developing training materials and planning training logistics | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Training enumerators | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA training | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Finalise list of hires and sign contracts | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Prepare field logistics equipment, finances | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Teacher survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Baseline data collection -treatment communities- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Baseline data collection -control communities- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Community survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Market survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA in fulani language | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Run data checks | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Update reports/calls | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mop-up and data cleaning | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Send baseline data to IFS team | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| EVALUATION TIMELINE | September | | | | | | October | | | | | | | | | | | | | | | | | | | | | | Nov | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Activity** | **22** | **25** | **26** | **27** | **28** | **29** | **2** | **3** | **4** | **5** | **6** | **9** | **10** | **11** | **12** | **13** | **16** | **17** | **18** | **19** | **20** | **23** | **24** | **25** | **26** | **27** | **30** | **31** | **1** | **2** | **3** | **6** | **7** | **8** | **9** | **10** | **11-12** | **13** | **14** | **15** | **17** | **18** | **19** | **20** | **21-26** | **27** | **29** |
| Piloting of baseline survey instruments in CAPI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Incorporate feedback from piloting into programmed surveys | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA pilot in CAPI | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Incorporate feedback from piloting IDELA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Incorporate feedback from | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Older siblings' development and pilot | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Finalize baseline survey instruments for training | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Sampling for baseline survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Recruitment of field staff (initial and adjustment) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Develop detailed field plans | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Write data check codes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Developing training materials and planning training logistics | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Training enumerators | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA training | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Finalise list of hires and sign contracts | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Prepare field logistics - equipment, finances | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Teacher survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Baseline data collection -treatment communities- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Baseline data collection -control communities- | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Community survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Market survey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IDELA in fulani language | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Run data checks | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Update reports/calls | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mop-up and data cleaning | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Send baseline data to IFS team | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

## A.5 Baseline procedures

All training activities and baseline data collection were monitored and supported by IFS staff on multiple trips to Ghana.

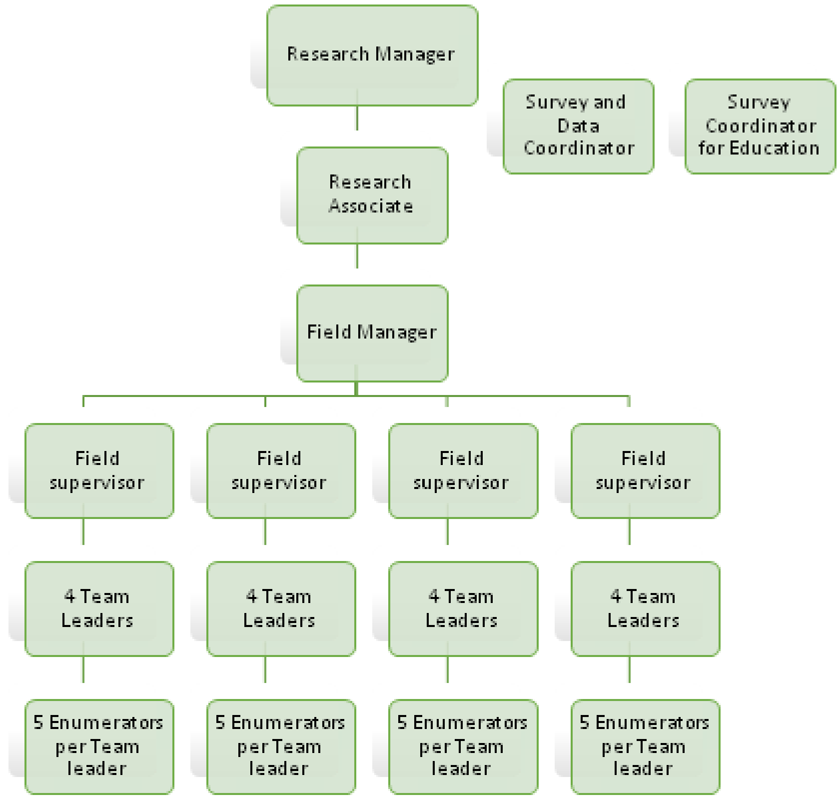
### A.5.1 Training

The training of enumerators took place between 28th August and 9th September 2017, in Tamale; 116 trainees were invited, 30 specifically for IDELA, and 86 for the remaining surveys. Of these individuals, 24 were chosen to be part of the final team for IDELA, and 68 for the other surveys. Training for the IDELA included 3 days of classroom training and two days of field practice, non-IDELA training included 8 days of in-classroom training and two days of field practice. Training on the IDELA was delivered by the survey coordinator for education, who had extensive experience of using the IDELA in another project in Ghana. Training on the other instruments was delivered by the Research Manager and Research Associate.

### A.5.2 Team structure

The data collection team consisted of 16 teams of 6 (5 enumerator and a team leader), 8 teams in Tolon and 8 in Bongo. Team leaders were always in the field with enumerators, in order to facilitate entry into a community, and solve or report any issues. Each group of 4 team leaders was managed by a field supervisor, who managed audio audits and ensured monitoring of the survey activities. The field supervisors were managed by a field manager who handled field logistics. The field manager was managed by a research associate, who oversaw recruitment, training, and project management. The whole project team was managed by the research manager, with assistance from support staff, the survey and data coordinator responsible for high frequency checks and data cleaning, and the survey coordinator for education, spearheading the the training, piloting, and monitoring of the IDELA tool. All data were collected electronically using tablets on the SurveyCTO platform.

Figure A2: Team Structure




### A.5.3 Data quality procedures

A variety of measures were undertaken in order to ensure high data quality. High frequency checks were run to maintain data quality and rectify errors, looking for duplicates, missing values, outliers and inconsistencies. Audio audits were conducted, with two individuals checking the random recordings from the previous day, listening for problems in translation, survey pace, or manner when asking questions. In addition, in at least 10% of randomly chosen submissions, auditors returned to check the data through redoing part of the survey, with incoming checks were compared to the original data on a daily basis. Any problems through these range of checks were reported to the field manager and research assistant, who acted upon them and rectified them to the best of their ability.

## A.6 Monitoring forms

Figure A3: Play Scheme monitoring form

Lively Minds Play Scheme Monitoring Form Community: _______________________________

Teacher 1 name: _______________________________

Teacher 2 name: _______________________________

| | | | | |
|---|---|---|---|---|
| | Day of week | | | |
| | Date | | | |
| | Monitor's name | | | |
| Play Scheme - Structure | Did the Play Scheme run? (1=yes, 0=no) | | | |
| | If no, was it prevented by Act of God? (1=yes, 0=no) | | | |
| | Number of Volunteers present | | | |
| | Number of children present | | | |
| | Did the children handwash? (2=with soap/ash, 1=water only, 0=no) | | | |
| | Was Scheme arranged into 5 mats with correct games on each mat? (1=yes, 0=no) | | | |
| | Were outdoor games played (1=yes, 0=no) | | | |

| | | | | |
|---|---|---|---|---|
| Play Scheme - Teaching | Are the games in a good condition? (1,2,3) | | | |
| | Is rotation punctual and orderly (1,2,3) | | | |
| | Are the rules of the games being followed? (1,2,3) | | | |
| | Are volunteers using caring & simple language (1,2,3) | | | |
| | Are the volunteers using turn-taking (1,2,3) | | | |
| | Were there 5 children per mat arranged in height order? (1=yes, 0= no) | | | |
| | Is teaching discovery-based? (1,2,3) | | | |
| | Do the children seem to be happy? (1,2,3) | | | |
| | Are the volunteers active and engaged? (1,2,3) | | | |

| | | | | |
|---|---|---|---|---|
| Comments | General comments on the Play Scheme (or if PS did not run, indicate the reason here) | | | |

Figure A4: Activity monitoring form

Lively Minds Activity Monitoring Form Community: ________________________________

Teacher 1 name: ________________________________

Teacher 2 name: ________________________________

| | | | | |
|---|---|---|---|---|
| | Date | | | |
| | Activity name | | | |
| | Monitor's name | | | |
| | Number of Volunteers in attendance | | | |
| | Did the session start on time? (1=yes, 0=no) | | | |
| Teacher 1 | Attendance (0-2 or NA) | | | |
| | If NA, reason for absence | | | |
| | How well prepared was this teacher? (0-4) | | | |
| | How well did the teacher follow the plan? (0-4) | | | |
| | Quality of delivery (including use of suitable translations) (0-4) | | | |
| | Interaction with Volunteers (including quality of supervision) (0-4) | | | |
| Teacher 2 | Attendance (0-2 or NA) | | | |
| | If NA, reason for absence | | | |
| | How well prepared was this teacher? (0-4) | | | |
| | How well did the teacher follow the plan? (0-4) | | | |
| | Quality of delivery (including use of suitable translations) (0-4) | | | |
| | Interaction with Volunteers (including quality of supervision) (0-4) | | | |
| Actions | Follow-up actions for Teachers | | | |
| | Follow-up actions for Lively Minds Staff | | | |
| Comments | General comments on the Activity (or if it did not run, indicate the reason here) | | | |

Table A1: Imbalanced variables

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| *Communities* | | | | |
| Drought shock in last 4 years (%) | 60.50 (48.91) | 41.11 (49.23) | 0.08* | 2122 |
| Shop to buy children's clothes nearby (%) | 40.00 (49.61) | 10.00 (30.38) | 0.00*** | 80 |
| *School and Teachers* | | | | |
| Desks per pupil | 0.07 (0.12) | 0.11 (0.13) | 0.06* | 2294 |
| Teacher: total years in current KG | 1.83 (1.76) | 2.46 (2.61) | 0.04** | 151 |
| Teacher: moved community for the job (%) | 57.89 (49.70) | 42.67 (49.79) | 0.05** | 151 |
| Teacher: temporary position (%) | 10.61 (30.87) | 1.41 (11.56) | 0.03** | 151 |
| Teacher: literacy knowledge (%) | 48.85 (20.63) | 56.00 (17.72) | 0.04** | 151 |
| Teacher: always praise good behaviour (%) | 64.47 (48.18) | 77.33 (42.15) | 0.03** | 151 |
| *Households* | | | | |
| Has savings (%) | 56.09 (49.65) | 46.71 (49.91) | 0.01** | 2407 |
| Owns table and chair | 37.03 (48.31) | 41.53 (49.30) | 0.09* | 2407 |
| Owns iron (Electric) | 3.11 (17.37) | 5.18 (22.17) | 0.02** | 2405 |
| Owns gas stove | 2.11 (14.36) | 3.71 (18.92) | 0.02** | 2399 |
| Expenditure on TC : Education Related items | 83.90 (376.56) | 169.18 (1108.88) | 0.06* | 2362 |
| Expenditure on YS : Food Related items | 176.53 (710.65) | 210.97 (689.44) | 0.04** | 976 |
| Expenditure on YS : Toys | 2.30 (20.04) | 3.72 (29.97) | 0.04** | 976 |
| *Primary Caregiver* | | | | |
| Caregiver since birth (%) | 83.63 (37.02) | 87.66 (32.90) | 0.06* | 2407 |
| Knows (from a list of 10 women) | 6.30 (2.58) | 5.79 (2.74) | 0.08* | 2336 |
| *Target Child* | | | | |
| Care outside household or pre school (%) | 4.62 (21.00) | 7.40 (26.19) | 0.06* | 2407 |
| Time studying at home | 0.35 (0.70) | 0.42 (0.72) | 0.06* | 2407 |
| Cough in last 30 days (%) | 52.44 (49.96) | 56.50 (49.60) | 0.04** | 2406 |
| Vomiting everything in last 30 days (%) | 29.41 (45.58) | 31.85 (46.61) | 0.04** | 2405 |
| Stomach pain in last 30 days (%) | 48.28 (49.99) | 52.67 (49.95) | 0.02** | 2406 |
| IDELA: Puzzle completion (%) | 11.49 (14.14) | 13.56 (15.52) | 0.04** | 2407 |
| IDELA: Letter identification (%) | 1.57 (6.28) | 3.08 (12.03) | 0.00*** | 2407 |
| Raw Credi Score (YS) | 8.69 (4.53) | 9.51 (4.55) | 0.01*** | 1057 |



Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## A.7 Baseline balance tests

### A.7.1 Summary : imbalanced variables

### A.7.2 Communities

Table A3: Wages (GHS)

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Agricultural (male) | 11.66 | 11.25 | 0.69 | 66 |
| | (3.85) | (4.72) | | |
| Agricultural (female) | 10.08 | 11.18 | 0.34 | 66 |
| | (3.42) | (5.55) | | |
| Non-agricultural (male) | 13.48 | 11.52 | 0.42 | 51 |
| | (7.02) | (6.92) | | |
| Non-agricultural (female) | 10.89 | 10.99 | 0.98 | 51 |
| | (6.30) | (6.34) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

*Note:* Outliers trimmed

Table A4: % of communities where:

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Main water source is borehole | 50.00 | 57.50 | 0.41 | 80 |
| | (50.64) | (50.06) | | |
| Main water source is surface water | 40.00 | 32.50 | 0.29 | 80 |
| | (49.61) | (47.43) | | |
| Any public toilet in use | 30.00 | 35.00 | 0.47 | 80 |
| | (46.41) | (48.30) | | |
| Open defacation is common | 75.00 | 80.00 | 0.73 | 80 |
| | (43.85) | (40.51) | | |
| At least one electricity connection | 65.00 | 65.00 | 0.86 | 80 |
| | (48.30) | (48.30) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: % of communities with the following within the community or nearby:

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Bank | 45.00 | 45.00 | 0.93 | 80 |
| | (50.38) | (50.38) | | |
| Bookshop | 37.50 | 27.50 | 0.48 | 80 |
| | (49.03) | (45.22) | | |
| Microfinance institution | 15.00 | 12.50 | 0.82 | 80 |
| | (36.16) | (33.49) | | |
| Shop to buy children's clothes | 40.00 | 10.00 | 0.00*** | 80 |
| | (49.61) | (30.38) | | |
| Shop to buy children's toys and games | 17.50 | 7.50 | 0.22 | 80 |
| | (38.48) | (26.67) | | |
| Weekly market | 52.50 | 55.00 | 0.77 | 80 |
| | (50.57) | (50.38) | | |
| General market | 42.50 | 42.50 | 0.77 | 80 |
| | (50.06) | (50.06) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: % of communities with the following within the community or nearby:

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Private school | 10.00 | 17.50 | 0.20 | 80 |
| | (30.38) | (38.48) | | |
| Public school | 90.00 | 87.50 | 0.55 | 80 |
| | (30.38) | (33.49) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: % of communities with:

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Familiy planning facility | 20.00 | 35.00 | 0.13 | 80 |
| | (40.51) | (48.30) | | |
| Pharmacy | 67.50 | 67.50 | 0.99 | 80 |
| | (47.43) | (47.43) | | |
| Private clinic | 12.50 | 15.00 | 0.92 | 80 |
| | (33.49) | (36.16) | | |
| Private hospital | 5.00 | 2.50 | 0.61 | 80 |
| | (22.07) | (15.81) | | |
| Public clinic | 52.50 | 60.00 | 0.42 | 80 |
| | (50.57) | (49.61) | | |
| Public hospital | 22.50 | 20.00 | 0.93 | 80 |
| | (42.29) | (40.51) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: % that experienced the following in the past 4 years

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Drought | 62.50 | 45.00 | 0.09* | 80 |
| | (49.03) | (50.38) | | |
| Erosion or landslide | 7.50 | 5.00 | 0.64 | 80 |
| | (26.67) | (22.07) | | |
| Flood | 57.50 | 50.00 | 0.70 | 80 |
| | (50.06) | (50.64) | | |
| Fire | 7.50 | 12.50 | 0.45 | 80 |
| | (26.67) | (33.49) | | |
| Pests (crops) | 50.00 | 42.50 | 0.59 | 80 |
| | (50.64) | (50.06) | | |
| Pests (humans) | 7.50 | 7.50 | 0.93 | 80 |
| | (26.67) | (26.67) | | |
| Pests (animals) | 47.50 | 40.00 | 0.59 | 80 |
| | (50.57) | (49.61) | | |
| Overflow of river | 5.00 | 0.00 | 0.12 | 80 |
| | (22.07) | (0.00) | | |
| Strong wind | 72.50 | 57.50 | 0.20 | 80 |
| | (45.22) | (50.06) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### A.7.3 Schools

Table A9: % to have ever attended pre-school by age category

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| 3 | 25.57 | 28.14 | 0.84 | 600 |
| | (43.70) | (45.04) | | |
| 4-7 | 57.43 | 54.90 | 0.67 | 1651 |
| | (49.47) | (49.79) | | |
| 7-10 | 62.83 | 57.38 | 0.27 | 2918 |
| | (48.34) | (49.47) | | |
| *Target Child* | 68.71 | 67.35 | 0.82 | 2405 |
| | (46.39) | (46.91) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: % to have ever attended school (including pre˙school) by age category

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| 3 | 31.15 | 33.22 | 0.96 | 600 |
| | (46.39) | (47.18) | | |
| 4-7 | 67.13 | 66.08 | 0.94 | 1651 |
| | (47.00) | (47.37) | | |
| 7-10 | 79.81 | 78.48 | 0.66 | 2918 |
| | (40.15) | (41.11) | | |
| 11-14 | 78.73 | 81.02 | 0.59 | 2016 |
| | (40.94) | (39.23) | | |
| 15-19 | 73.49 | 74.19 | 0.95 | 2130 |
| | (44.16) | (43.78) | | |
| *Target Child* | 77.54 | 75.90 | 0.75 | 2405 |
| | (41.75) | (42.78) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: % of those to ever attend school who have attended school in the last 12 months (including pre-school) by age category

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| 3 | 89.47 | 90.82 | 0.75 | 193 |
| | (30.85) | (29.03) | | |
| 4-7 | 96.17 | 94.87 | 0.48 | 1100 |
| | (19.22) | (22.09) | | |
| 7-10 | 96.90 | 95.88 | 0.25 | 2310 |
| | (17.34) | (19.89) | | |
| 11-14 | 93.40 | 92.57 | 0.79 | 1610 |
| | (24.84) | (26.25) | | |
| 15-19 | 81.38 | 78.63 | 0.36 | 1573 |
| | (38.95) | (41.01) | | |
| *Target Child* | 90.35 | 88.30 | 0.33 | 1845 |
| | (29.55) | (32.16) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Travel to nearest public preschool

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Travel by foot (%) | 88.58 | 88.24 | 0.85 | 2407 |
| | (31.82) | (32.23) | | |
| Travel time in minutes | 11.17 | 11.70 | 0.19 | 2299 |
| | (9.27) | (9.07) | | |

*Note:* Outliers in travel times trimmed

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: School characteristics

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Average no. Pupils per KG class | 54.05 | 51.37 | 0.38 | 76 |
| | (21.43) | (18.24) | | |
| Desks per pupil | 0.08 | 0.12 | 0.20 | 76 |
| | (0.13) | (0.13) | | |
| Has electricity | 5.00 | 15.00 | 0.18 | 80 |
| | (22.07) | (36.16) | | |
| Has a blackboard | 95.00 | 95.00 | 0.82 | 80 |
| | (22.07) | (22.07) | | |
| School has a toilet (%) | 30.00 | 47.50 | 0.10 | 80 |
| | (46.41) | (50.57) | | |
| Books available (%) | 42.50 | 52.50 | 0.49 | 80 |
| | (50.06) | (50.57) | | |
| School has a major safety hazard (%) | 42.50 | 47.50 | 0.71 | 80 |
| | (50.06) | (50.57) | | |

*Note:* Outliers in class size and desks trimmed

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### A.7.4 Teachers

Table A14: Basic characteristics

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Male (%) | 47.37 | 48.00 | 0.87 | 151 |
| | (50.26) | (50.30) | | |
| Muslim (%) | 50.00 | 52.00 | 0.88 | 151 |
| | (52.92) | (50.30) | | |
| Tertiary education (%) | 82.89 | 89.33 | 0.31 | 151 |
| | (37.91) | (31.08) | | |
| Some ECCE training (%) | 69.74 | 77.33 | 0.21 | 151 |
| | (46.24) | (42.15) | | |
| Total years of experience | 3.34 | 4.09 | 0.36 | 151 |
| | (3.94) | (4.88) | | |
| Total years in current KG | 1.83 | 2.46 | 0.04** | 151 |
| | (1.76) | (2.61) | | |
| Married | 80.26 | 89.33 | 0.10 | 151 |
| | (40.07) | (31.08) | | |
| Wealth index (standardised) | -0.01 | 0.01 | 0.67 | 151 |
| | (1.01) | (1.00) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Living situation

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Born in community where the school is located (%) | 6.58 | 12.00 | 0.11 | 151 |
| | (24.96) | (32.71) | | |
| Live in community where the school is located (%) | 15.79 | 18.67 | 0.34 | 151 |
| | (36.71) | (39.23) | | |
| Moved community for the job (%) | 57.89 | 42.67 | 0.05** | 151 |
| | (49.70) | (49.79) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Work conditions and teacher practices

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Hours worked at school (weekly) | 22.24 | 21.94 | 0.87 | 151 |
| | (9.30) | (8.81) | | |
| Hours worked preparing outside of school (weekly) | 7.62 | 6.84 | 0.47 | 151 |
| | (5.67) | (7.04) | | |
| Monthly Salary (GHS) | 948.53 | 969.03 | 0.96 | 151 |
| | (374.42) | (355.57) | | |
| Reported proportion salary paid on time (%) | 65.21 | 65.20 | 0.94 | 151 |
| | (28.37) | (25.22) | | |
| Work an additional job (%) | 5.26 | 5.33 | 0.96 | 151 |
| | (22.48) | (22.62) | | |
| Temporary position (%) | 10.61 | 1.41 | 0.03** | 151 |
| | (30.87) | (11.56) | | |
| Literacy knowledge (%) | 48.85 | 56.00 | 0.04** | 151 |
| | (20.63) | (17.72) | | |
| Respond aggressively to bad behaviour (%) | 48.68 | 50.67 | 0.55 | 151 |
| | (50.31) | (50.33) | | |
| Always praise good behaviour (%) | 64.47 | 77.33 | 0.03** | 151 |
| | (48.18) | (42.15) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Psychological well-being

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Depression score (SRQ) | 3.79 | 3.64 | 0.78 | 151 |
| | (3.80) | (3.30) | | |
| External control | 16.62 | 16.71 | 0.60 | 151 |
| | (4.16) | (3.60) | | |
| Motivation | 37.05 | 37.68 | 0.61 | 151 |
| | (4.80) | (5.13) | | |
| Job satisfaction | 89.24 | 88.76 | 0.51 | 151 |
| | (7.34) | (7.69) | | |
| Burnout | 78.42 | 76.69 | 0.35 | 151 |
| | (13.37) | (13.13) | | |

*Note:* These are all raw scores and have no direct interpretation

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### A.7.5 Households

Table A18: Household head background

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Polygamous household | 24.85 | 22.62 | 0.51 | 2407 |
| | (43.23) | (41.85) | | |
| Christian | 21.75 | 24.67 | 0.47 | 2407 |
| | (41.27) | (43.13) | | |
| Islam | 51.89 | 52.14 | 0.83 | 2407 |
| | (49.99) | (49.97) | | |
| Traditional African | 25.94 | 22.70 | 0.71 | 2407 |
| | (43.85) | (41.90) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A19: Household size

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Number of people | 9.14 | 9.25 | 1.00 | 2291 |
| | (4.62) | (4.43) | | |
| Number of adult men (17 or older) | 2.31 | 2.32 | 0.87 | 2407 |
| | (1.58) | (1.59) | | |
| Number of adult women (17 or older) | 2.50 | 2.54 | 0.94 | 2407 |
| | (2.07) | (1.98) | | |
| Number of children (16 or younger) | 4.48 | 4.20 | 0.44 | 2407 |
| | (3.75) | (3.33) | | |
| Number of young children (6 or younger) | 1.92 | 1.80 | 0.55 | 2407 |
| | (2.16) | (2.05) | | |

*Note:* Outliers in household size trimmed

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A20: % of households with following relations to TC living within

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Biological Father | 76.24 | 79.52 | 0.40 | 2407 |
| | (42.58) | (40.37) | | |
| Biological Mother | 73.80 | 78.54 | 0.22 | 2407 |
| | (43.99) | (41.07) | | |
| Grandparent | 52.73 | 56.00 | 0.12 | 2407 |
| | (49.95) | (49.66) | | |
| Uncle/Aunt | 50.29 | 50.16 | 0.89 | 2407 |
| | (50.02) | (50.02) | | |
| Cousin | 32.66 | 33.72 | 0.63 | 2407 |
| | (46.92) | (47.29) | | |
| Sibling | 82.37 | 84.21 | 0.63 | 2407 |
| | (38.13) | (36.48) | | |
| Older Sibling | 44.00 | 28.29 | 0.16 | 2407 |
| | (49.66) | (45.06) | | |
| Younger Sibling | 69.44 | 52.88 | 0.12 | 2407 |
| | (46.09) | (49.94) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A21: Household economic situation

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Monthly income (GHS) | 116.20 | 122.04 | 0.55 | 2239 |
| | (118.43) | (121.56) | | |
| Monthly expenditure (GHS) | 1679.84 | 1784.35 | 0.51 | 2337 |
| | (1230.56) | (1261.28) | | |
| Main income source : Farming own land (%) | 43.66 | 43.67 | 0.45 | 2407 |
| | (49.62) | (49.62) | | |
| Main income source: Waged work (%) | 27.62 | 28.37 | 0.40 | 2407 |
| | (44.73) | (45.10) | | |
| Main income source: Profits from small enterprise (%) | 19.40 | 18.01 | 0.96 | 2407 |
| | (39.56) | (38.44) | | |
| Main income source: Livestock (%) | 3.78 | 3.04 | 0.66 | 2407 |
| | (19.08) | (17.18) | | |
| Has a loan (%) | 35.77 | 33.80 | 0.88 | 2407 |
| | (47.95) | (47.32) | | |
| Has savings (%) | 56.09 | 46.71 | 0.01** | 2407 |
| | (49.65) | (49.91) | | |

*Note:* Outliers in income and expenditure trimmed

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Mobile Phone | 89.42 | 89.23 | 0.73 | 2407 |
| | (30.77) | (31.02) | | |
| Radio | 52.06 | 54.77 | 0.50 | 2407 |
| | (49.98) | (49.79) | | |
| Refridgerator/freezer | 3.19 | 4.77 | 0.13 | 2407 |
| | (17.58) | (21.32) | | |
| Bicycle | 76.07 | 79.61 | 0.43 | 2407 |
| | (42.68) | (40.31) | | |
| Television | 23.87 | 31.33 | 0.12 | 2406 |
| | (42.64) | (46.40) | | |
| Video/DVD/VCR | 17.21 | 21.30 | 0.32 | 2407 |
| | (37.76) | (40.96) | | |
| Motorbike/scooter | 32.49 | 32.48 | 0.83 | 2407 |
| | (46.85) | (46.85) | | |
| Car/truck/automobile | 1.18 | 0.99 | 0.66 | 2407 |
| | (10.78) | (9.89) | | |
| Tractor | 1.93 | 1.48 | 0.78 | 2407 |
| | (13.77) | (12.08) | | |
| Farm equipment (pump, plough, etc.) | 21.07 | 20.39 | 0.65 | 2407 |
| | (40.80) | (40.31) | | |
| Sewing machine | 20.49 | 17.43 | 0.21 | 2407 |
| | (40.38) | (37.96) | | |
| Bed | 37.95 | 39.64 | 0.72 | 2407 |
| | (48.55) | (48.93) | | |
| Mattress | 47.94 | 47.78 | 0.93 | 2407 |
| | (49.98) | (49.97) | | |
| Table and chair | 37.03 | 41.53 | 0.09* | 2407 |
| | (48.31) | (49.30) | | |
| Stool | 92.53 | 92.19 | 0.87 | 2407 |
| | (26.31) | (26.85) | | |
| Sofa | 14.79 | 13.82 | 0.87 | 2406 |

| | | | | |
|---|---|---|---|---|
| | (35.51) | (34.52) | | |
| Fan | 18.14 | 20.81 | 0.57 | 2407 |
| | (38.55) | (40.61) | | |
| Kerosene lamp | 18.89 | 18.67 | 0.80 | 2407 |
| | (39.16) | (38.98) | | |
| Watch | 25.52 | 22.04 | 0.33 | 2407 |
| | (43.62) | (41.47) | | |
| Woven mat | 83.63 | 80.92 | 0.70 | 2407 |
| | (37.02) | (39.31) | | |
| Torch | 89.50 | 88.16 | 0.43 | 2406 |
| | (30.67) | (32.32) | | |
| Cabinet | 16.12 | 13.91 | 0.61 | 2406 |
| | (36.79) | (34.62) | | |
| Cooking pot | 97.98 | 98.03 | 1.00 | 2407 |
| | (14.06) | (13.92) | | |
| Mosquito net | 95.47 | 93.42 | 0.16 | 2407 |
| | (20.81) | (24.80) | | |
| Iron (Electric) | 3.11 | 5.18 | 0.02** | 2405 |
| | (17.37) | (22.17) | | |
| Iron (Coal/Box) | 15.04 | 14.39 | 0.95 | 2406 |
| | (35.76) | (35.11) | | |
| Gas stove | 2.11 | 3.71 | 0.02** | 2399 |
| | (14.36) | (18.92) | | |
| Wealth Index (PCA factor score) | -0.06 | 0.06 | 0.55 | 2394 |
| | (2.05) | (2.17) | | |

Table A23: Household expenditures on different members (GHS)

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| TC : Health Related items | 451.18 | 436.41 | 0.90 | 2362 |
| | (1694.62) | (1394.49) | | |
| TC : Education Related items | 83.90 | 169.18 | 0.06* | 2362 |
| | (376.56) | (1108.88) | | |
| TC : Food Related items | 633.77 | 832.85 | 0.28 | 2362 |
| | (5044.43) | (5054.02) | | |
| TC : Toys | 7.21 | 8.98 | 0.23 | 2362 |
| | (48.82) | (66.54) | | |
| TC : Other | 681.99 | 629.49 | 0.74 | 2362 |
| | (4136.17) | (2778.77) | | |
| OS : Health Related items | 259.47 | 230.53 | 0.61 | 1447 |
| | (1136.57) | (610.61) | | |
| OS : Education Related items | 196.21 | 223.10 | 0.42 | 1447 |
| | (858.47) | (989.56) | | |
| OS : Food Related items | 545.48 | 470.09 | 0.73 | 1447 |
| | (6136.90) | (1711.32) | | |
| OS : Toys | 7.72 | 4.43 | 0.96 | 1447 |
| | (100.91) | (56.48) | | |
| OS : Other | 550.04 | 309.40 | 0.33 | 1447 |
| | (4781.07) | (1107.62) | | |
| YS : Health Related items | 318.38 | 452.92 | 0.11 | 976 |
| | (959.12) | (1832.98) | | |
| YS : Education Related items | 0.01 | 2.53 | 0.19 | 976 |
| | (0.18) | (42.39) | | |
| YS : Food Related items | 176.53 | 210.97 | 0.04** | 976 |
| | (710.65) | (689.44) | | |
| YS : Toys | 2.30 | 3.72 | 0.04** | 976 |
| | (20.04) | (29.97) | | |
| YS : Other | 692.10 | 674.50 | 0.69 | 976 |
| | (2065.93) | (2037.58) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: % experiencing the following shock since pregnancy of TC

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| A fire | 5.38 | 6.26 | 0.29 | 2403 |
| | (22.58) | (24.23) | | |
| Severe flood | 18.08 | 16.39 | 0.79 | 2403 |
| | (38.50) | (37.04) | | |
| Severe drought | 34.99 | 32.54 | 0.60 | 2403 |
| | (47.71) | (46.87) | | |
| Decrease, change in food availability | 33.56 | 31.99 | 1.00 | 2402 |
| | (47.24) | (46.66) | | |
| Livestock died | 40.54 | 39.04 | 0.82 | 2403 |
| | (49.12) | (48.81) | | |
| Crops failed | 39.08 | 37.64 | 0.62 | 2404 |
| | (48.81) | (48.47) | | |
| Livestock stolen | 12.29 | 12.18 | 0.78 | 2403 |
| | (32.85) | (32.72) | | |
| Crops stolen | 3.20 | 3.30 | 0.57 | 2401 |
| | (17.60) | (17.86) | | |
| Death/reduction in household members | 17.19 | 18.47 | 0.40 | 2400 |
| | (37.74) | (38.82) | | |
| Job loss/loss source of income/family enterprises | 4.80 | 4.77 | 0.84 | 2403 |
| | (21.38) | (21.33) | | |
| Severe illness or injury | 24.56 | 24.71 | 0.80 | 2403 |
| | (43.06) | (43.15) | | |
| Victim of crime | 1.77 | 1.73 | 0.84 | 2395 |
| | (13.20) | (13.06) | | |
| Divorced or separated | 2.45 | 2.06 | 0.56 | 2395 |
| | (15.46) | (14.22) | | |
| Birth/new household member | 18.18 | 20.03 | 0.51 | 2401 |
| | (38.59) | (40.04) | | |
| Paying for child's education | 14.09 | 15.92 | 0.19 | 2397 |
| | (34.81) | (36.61) | | |
| Moved/migrated/fled | 4.56 | 4.40 | 0.86 | 2390 |
| | (20.86) | (20.51) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A25: Family Care Indicators

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Number of different play materials | 1.07 | 1.22 | 0.15 | 2407 |
| | (1.12) | (1.23) | | |
| Have any homemade toys (%) | 46.26 | 48.11 | 0.49 | 2407 |
| | (49.88) | (49.98) | | |
| Have any bought toys (%) | 21.58 | 25.41 | 0.53 | 2407 |
| | (41.15) | (43.55) | | |
| Number of play activities (TC) | 0.17 | 0.21 | 0.14 | 2407 |
| | (0.55) | (0.62) | | |
| Number of play activities (YS) | 0.02 | 0.03 | 0.25 | 1579 |
| | (0.19) | (0.24) | | |
| Number of play activities (OS) | 0.16 | 0.18 | 0.47 | 1313 |
| | (0.52) | (0.54) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### A.7.6 Primary caregivers

Table A26: Primary caregiver characteristics

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Age | 37.00 | 35.93 | 0.07* | 2407 |
| | (12.41) | (12.09) | | |
| Christian (%) | 34.51 | 35.88 | 0.69 | 2406 |
| | (47.56) | (47.99) | | |
| Muslim (%) | 53.99 | 54.24 | 0.83 | 2406 |
| | (49.86) | (49.84) | | |
| No education (%) | 79.60 | 79.19 | 0.73 | 2407 |
| | (40.32) | (40.61) | | |
| Illiterate (%) | 92.19 | 91.61 | 0.76 | 2407 |
| | (26.84) | (27.73) | | |
| Caregiver since birth (%) | 83.63 | 87.66 | 0.06* | 2407 |
| | (37.02) | (32.90) | | |
| Born outside community (%) | 70.45 | 67.76 | 0.26 | 2407 |
| | (45.65) | (46.76) | | |
| Born outside district (%) | 15.53 | 17.19 | 0.36 | 2407 |
| | (36.24) | (37.74) | | |
| Ravens score (%) | 43.62 | 43.19 | 0.89 | 2407 |
| | (17.08) | (15.96) | | |
| PC is biological mother (%) | 75.90 | 79.28 | 0.29 | 2407 |
| | (42.79) | (40.55) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A27: Number of:

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Community groups are active in | 1.07 | 1.02 | 0.71 | 2407 |
| | (1.04) | (1.06) | | |
| Community groups with a focus on ECCE | 0.25 | 0.25 | 0.85 | 2407 |
| | (0.51) | (0.56) | | |
| People can rely on for emotional support | 2.49 | 2.35 | 0.22 | 2407 |
| | (1.86) | (1.69) | | |
| People who you would lend GHS 100 or more | 2.93 | 2.92 | 0.68 | 2407 |
| | (3.52) | (2.98) | | |
| People who you could borrow GHS 100 or more | 2.54 | 2.62 | 0.60 | 2407 |
| | (1.81) | (2.35) | | |
| Number of relatives in the community | 4.79 | 4.85 | 0.92 | 2327 |
| | (6.53) | (6.30) | | |

*Note:* Outliers in number of relatives trimmed

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A28: Of a random list of 10 women in the community, how many does the PC:

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Knows | 6.30 | 5.79 | 0.08* | 2336 |
| | (2.58) | (2.74) | | |
| Is close to | 3.11 | 2.79 | 0.33 | 2336 |
| | (2.87) | (2.75) | | |
| Talks about child with | 2.92 | 2.55 | 0.21 | 2336 |
| | (2.75) | (2.69) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A29: % who:

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Are a PTA member | 15.79 | 13.72 | 0.82 | 1674 |
| | (36.49) | (34.43) | | |
| Know teachers name | 38.75 | 44.13 | 0.12 | 1818 |
| | (48.75) | (49.68) | | |
| Doesn't know how often teacher is absent | 11.25 | 12.38 | 0.54 | 1819 |
| | (31.61) | (32.95) | | |
| Visited school in the last month | 38.20 | 40.07 | 0.48 | 1819 |
| | (48.61) | (49.03) | | |
| Ever attended a PTA meeting | 79.74 | 76.21 | 0.43 | 1367 |
| | (40.22) | (42.61) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A30: PC time use (hours)

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Sleeping | 8.02 | 8.11 | 0.51 | 2407 |
| | (1.20) | (1.25) | | |
| Farm work | 4.42 | 4.36 | 0.70 | 2407 |
| | (2.06) | (2.36) | | |
| Domestic tasks | 4.39 | 4.33 | 0.50 | 2407 |
| | (1.54) | (1.52) | | |
| Leisure | 3.28 | 3.18 | 0.23 | 2407 |
| | (1.48) | (1.40) | | |
| Caring for others | 1.96 | 1.92 | 0.81 | 2407 |
| | (1.45) | (1.40) | | |
| Paid work | 1.12 | 1.21 | 0.38 | 2407 |
| | (2.12) | (2.33) | | |
| Playing with children | 0.62 | 0.64 | 0.91 | 2407 |
| | (0.95) | (0.92) | | |
| Collecting children | 0.20 | 0.24 | 0.20 | 2407 |
| | (0.53) | (0.64) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A31: Primary caregiver health % answering yes

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Have you experienced significant weight loss in the last 12 months? | 62.89 | 60.20 | 0.30 | 2407 |
| | (48.85) | (49.47) | | |
| Have you suffered from a prolonged fever in the last 12 months? | 43.74 | 43.17 | 0.52 | 2407 |
| | (50.13) | (50.21) | | |
| Have you had chronic diarrhoea in the last 12 months? | 19.90 | 21.22 | 0.65 | 2407 |
| | (40.57) | (41.90) | | |
| Do you usually smoke cigarettes? | 0.92 | 0.90 | 0.75 | 2407 |
| | (9.57) | (9.47) | | |
| Do you usually use other form of tobacco such as chewing or snuffing tobac | 3.11 | 3.04 | 0.73 | 2407 |
| | (17.36) | (17.18) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A32: Primary caregiver well-being

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Raw Self-esteem (Rosenborg) score | 18.05 | 17.98 | 0.93 | 2407 |
| | (3.77) | (3.56) | | |
| Raw Depression (SRQ-20) Score | 8.06 | 8.32 | 0.58 | 2407 |
| | (5.29) | (5.26) | | |
| Depression (%) | 50.71 | 51.73 | 0.65 | 2407 |
| | (50.02) | (49.99) | | |
| Raw Rumination Scale Score | 10.63 | 10.88 | 0.68 | 2407 |
| | (2.91) | (2.91) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A33: Primary caregiver knowledge (KIDI)

| | Control | Treament | p-value | N |
|---|---|---|---|---|
| Parents play an important role in children's learning and development | 3.63 (0.53) | 3.65 (0.50) | 0.22 | 2407 |
| Knowing how to read and write is important for children to have a good/productive life | 3.62 (0.52) | 3.63 (0.50) | 0.45 | 2407 |
| Parents can support children's educational development at home | 3.59 (0.56) | 3.56 (0.55) | 0.97 | 2407 |
| Children can learn a lot of skills by playing games. | 3.40 (0.67) | 3.36 (0.65) | 0.64 | 2407 |
| It is possible for parents to talk with or engage children in games while doing their daily work | 3.29 (0.73) | 3.21 (0.77) | 0.34 | 2407 |
| Praising children when he/she tries to do something new is important | 3.54 (0.57) | 3.54 (0.57) | 0.66 | 2407 |

*Note:* Average score on a scale 1 (Strongly disagree) to 4(Strongly agree)

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A34: TC characteristics

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Age in months | 56.12 | 56.31 | 0.68 | 2405 |
| | (9.50) | (9.56) | | |
| Male (%) | 50.46 | 49.26 | 0.71 | 2407 |
| | (50.02) | (50.02) | | |
| Ever attended school (%) | 77.54 | 75.90 | 0.75 | 2405 |
| | (41.75) | (42.78) | | |
| Currently in school(%) | 78.25 | 77.63 | 0.91 | 2407 |
| | (41.27) | (41.69) | | |
| Care outside household or pre school (%) | 4.62 | 7.40 | 0.06* | 2407 |
| | (21.00) | (26.19) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A35: TC time use (hours)

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Sleeping | 10.07 | 10.11 | 0.72 | 2407 |
| | (1.06) | (1.02) | | |
| Playing | 4.62 | 4.69 | 0.99 | 2407 |
| | (1.97) | (2.09) | | |
| School | 4.46 | 4.41 | 0.98 | 2407 |
| | (2.53) | (2.49) | | |
| Leisure | 3.29 | 3.22 | 0.34 | 2407 |
| | (1.52) | (1.56) | | |
| Caring for others | 0.46 | 0.41 | 0.81 | 2407 |
| | (0.88) | (0.86) | | |
| Homework | 0.35 | 0.42 | 0.06* | 2407 |
| | (0.70) | (0.72) | | |
| Domestic tasks | 0.39 | 0.36 | 0.84 | 2407 |
| | (0.78) | (0.75) | | |
| Farm work | 0.35 | 0.38 | 0.48 | 2407 |
| | (0.81) | (0.96) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A36: TC basic health outcomes

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Birth weight (KG) | 3.28 | 3.24 | 0.57 | 1896 |
| | (0.73) | (0.71) | | |
| Reported health (out of 5) | 3.96 | 3.93 | 0.45 | 2407 |
| | (0.78) | (0.84) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A37: During last 30 days TC has had.. (%)

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| 3 or more loose or watery stools | 31.34 | 32.04 | 0.79 | 2395 |
| | (46.41) | (46.68) | | |
| Blood in his/her stools | 8.68 | 9.55 | 0.37 | 2391 |
| | (28.16) | (29.40) | | |
| High fever | 32.07 | 35.94 | 0.18 | 2407 |
| | (46.70) | (48.00) | | |
| Cough | 52.44 | 56.50 | 0.04** | 2406 |
| | (49.96) | (49.60) | | |
| Very fast or difficult breathing | 15.88 | 18.34 | 0.15 | 2406 |
| | (36.57) | (38.71) | | |
| Vomiting everything | 29.41 | 31.85 | 0.04** | 2405 |
| | (45.58) | (46.61) | | |
| Stomach pain | 48.28 | 52.67 | 0.02** | 2406 |
| | (49.99) | (49.95) | | |
| Serious loss of appetite | 42.77 | 42.11 | 0.98 | 2406 |
| | (49.50) | (49.39) | | |
| Skin rashes | 19.93 | 19.00 | 0.44 | 2405 |
| | (39.97) | (39.24) | | |
| Sores on feet and legs | 16.46 | 16.20 | 0.58 | 2407 |
| | (37.09) | (36.86) | | |
| Convulsions | 4.54 | 4.12 | 0.99 | 2404 |
| | (20.82) | (19.88) | | |
| Unusual tiredness | 8.16 | 10.00 | 0.57 | 2398 |
| | (27.39) | (30.01) | | |
| Unconsciousness | 1.43 | 1.65 | 0.68 | 2398 |
| | (11.89) | (12.74) | | |
| Extreme lethargy e.g. extremely weak/listless | 4.43 | 5.72 | 0.18 | 2381 |
| | (20.58) | (23.23) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A38: TC has following long term health problem.. (%)

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Physical disability | 0.67 | 0.99 | 0.79 | 2405 |
| | (8.18) | (9.89) | | |
| Mental disability | 0.34 | 0.41 | 0.84 | 2404 |
| | (5.79) | (6.41) | | |
| Fits/epilepsy/convulsions | 1.60 | 1.40 | 0.73 | 2405 |
| | (12.54) | (11.75) | | |
| Skin problems | 7.56 | 6.91 | 0.63 | 2406 |
| | (26.44) | (25.38) | | |
| Asthma/respiratory problems | 1.93 | 1.98 | 0.79 | 2403 |
| | (13.77) | (13.93) | | |
| Anaemia | 3.11 | 2.15 | 0.36 | 2399 |
| | (17.38) | (14.50) | | |
| HIV/AIDS | 0.08 | 0.25 | 0.23 | 2387 |
| | (2.91) | (4.99) | | |
| Congenital illness | 0.84 | 0.83 | 0.77 | 2394 |
| | (9.16) | (9.06) | | |
| Stomach ache/abdominal problems | 14.62 | 15.72 | 0.39 | 2405 |
| | (35.35) | (36.41) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A39: IDELA : Subdomain scores (%)

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| Socio-emotional | 27.71 | 27.28 | 28.13 | 0.78 | 2407 |
| | (17.35) | (16.75) | (17.92) | | |
| Emergent Numeracy | 27.89 | 27.61 | 28.17 | 0.91 | 2407 |
| | (14.13) | (13.53) | (14.69) | | |
| Emeregent Literacy | 19.43 | 19.53 | 19.34 | 1.00 | 2407 |
| | (13.13) | (12.69) | (13.54) | | |
| Motor Skills | 29.31 | 28.80 | 29.81 | 0.66 | 2407 |
| | (26.49) | (26.13) | (26.85) | | |
| Executive Function | 36.14 | 36.52 | 35.76 | 0.63 | 2407 |
| | (26.33) | (25.38) | (27.23) | | |
| School readiness | 27.79 | 27.74 | 27.85 | 0.96 | 2407 |
| | (14.61) | (13.91) | (15.27) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A40: IDELA : Individual item scores (%)

| | Total | Control | Treament | p-value | N |
|---|---|---|---|---|---|
| *Socio-emotional* | | | | | |
| Self-awareness | 58.16 | 57.46 | 58.83 | 0.72 | 2407 |
| | (26.80) | (26.53) | (27.05) | | |
| Number of friends | 31.17 | 31.44 | 30.91 | 0.84 | 2407 |
| | (26.29) | (26.64) | (25.95) | | |
| Emotional awareness | 6.51 | 5.98 | 7.03 | 0.54 | 2407 |
| | (20.27) | (19.27) | (21.20) | | |
| Empathy/perspective taking | 14.31 | 13.37 | 15.23 | 0.49 | 2407 |
| | (25.73) | (24.49) | (26.87) | | |
| Solving conflict | 28.40 | 28.13 | 28.66 | 0.90 | 2407 |
| | (36.39) | (35.71) | (37.06) | | |
| *Emergent Numeracy* | | | | | |
| Comparison by size and length | 78.99 | 79.93 | 78.06 | 0.27 | 2407 |
| | (26.65) | (26.00) | (27.26) | | |
| Sorting and classification | 24.10 | 23.22 | 24.96 | 0.71 | 2407 |
| | (29.91) | (29.00) | (30.77) | | |
| Shape identification | 29.12 | 28.38 | 29.85 | 0.64 | 2407 |
| | (24.83) | (24.18) | (25.44) | | |
| Number identification | 4.34 | 3.98 | 4.68 | 0.15 | 2407 |
| | (9.87) | (7.89) | (11.48) | | |
| Counting | 16.49 | 16.26 | 16.72 | 0.88 | 2407 |
| | (24.18) | (23.78) | (24.57) | | |
| Addition and subtraction | 29.68 | 30.03 | 29.33 | 0.45 | 2407 |
| | (27.75) | (27.71) | (27.79) | | |
| Puzzle completion | 12.54 | 11.49 | 13.56 | 0.04** | 2407 |
| | (14.88) | (14.14) | (15.52) | | |
| *Emergent Literacy* | | | | | |
| Expressive vocabulary | 25.69 | 25.76 | 25.61 | 0.90 | 2407 |
| | (20.98) | (21.11) | (20.87) | | |
| Print awareness | 25.90 | 26.22 | 25.58 | 0.88 | 2407 |
| | (30.49) | (30.80) | (30.20) | | |
| Letter identification | 2.33 | 1.57 | 3.08 | 0.00*** | 2407 |
| | (9.65) | (6.28) | (12.03) | | |
| First letter sounds | 9.51 | 9.74 | 9.29 | 0.98 | 2407 |
| | (18.94) | (18.94) | (18.95) | | |
| Emergent writing | 13.50 | 13.75 | 13.26 | 0.95 | 2407 |
| | (23.18) | (23.02) | (23.34) | | |
| Oral comprehension | 39.66 | 40.12 | 39.21 | 0.41 | 2407 |
| | (32.61) | (32.30) | (32.92) | | |
| *Motor Skills* | | | | | |
| Copying a shape | 35.71 | 35.81 | 35.61 | 0.89 | 2407 |
| | (41.07) | (40.62) | (41.52) | | |
| Drawing a person | 26.48 | 26.36 | 26.59 | 0.98 | 2407 |
| | (31.04) | (31.01) | (31.08) | | |
| Folding paper | 25.74 | 24.22 | 27.22 | 0.28 | 2407 |
| | (26.08) | (25.37) | (26.69) | | |
| *Executive Function* | | | | | |
| Short term memory | 53.94 | 55.21 | 52.69 | 0.27 | 2407 |
| | (30.42) | (30.02) | (30.77) | | |
| Inhibitory control | 32.13 | 32.12 | 32.14 | 0.93 | 2407 |
| | (40.34) | (39.54) | (41.12) | | |
| Pencil tapping task | 22.34 | 22.24 | 22.45 | 0.92 | 2407 |
| | (33.00) | (32.65) | (33.34) | | |



Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A41: Test scores for siblings

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| Raw Credi Score (YS) | 8.69 | 9.51 | 0.01*** | 1057 |
| | (4.53) | (4.55) | | |
| Total score (OS) | 24.96 | 26.19 | 0.28 | 1428 |
| | (9.62) | (10.50) | | |
| Maths score (OS) | 21.72 | 21.84 | 0.73 | 1428 |
| | (13.23) | (13.67) | | |
| Literacy score (OS) | 8.15 | 8.86 | 0.32 | 1428 |
| | (12.44) | (12.73) | | |
| Ravens score (OS) | 33.77 | 35.21 | 0.45 | 1428 |
| | (13.15) | (14.07) | | |
| Forward digit span score (OS) | 53.32 | 54.03 | 0.90 | 1428 |
| | (20.81) | (21.69) | | |
| Backward digit span score (OS) | 7.83 | 11.03 | 0.01** | 1428 |
| | (12.45) | (16.55) | | |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A42: Hand washing understanding

| | Control Mean | Treament Mean | p-value | N |
|---|---|---|---|---|
| *Primary caregiver* | | | | |
| Number of different times for handwashing stated | 1.96 | 1.95 | 0.60 | 2407 |
| | (1.00) | (0.99) | 0.90 | 2407 |
| Stated water is needed (%) | 87.32 | 85.88 | 0.97 | 2387 |
| | (33.29) | (34.84) | 0.72 | 1769 |
| Stated soap is needed (%) | 98.82 | 98.35 | 0.81 | 2398 |
| | (10.80) | (12.75) | 0.59 | 1793 |
| Stated one correct example of why handwashing is important (%) | 98.23 | 98.68 | 0.35 | 2401 |
| | (13.19) | (11.41) | 0.96 | 1215 |
| *Target child* | | | | |
| Number of different times for handwashing stated | 0.70 | 0.70 | 0.60 | 2407 |
| | (0.78) | (0.79) | 0.90 | 2407 |
| Stated water is needed (%) | 57.06 | 62.94 | 0.97 | 2387 |
| | (49.53) | (48.32) | 0.72 | 1769 |
| Stated soap is needed (%) | 72.73 | 75.00 | 0.81 | 2398 |
| | (44.56) | (43.33) | 0.59 | 1793 |
| Stated one correct example of why handwashing is important (%) | 20.92 | 23.38 | 0.35 | 2401 |
| | (40.70) | (42.36) | 0.96 | 1215 |

Means reported with standard errors in parenthesis. P-values from a t-test of the equivalence of means between treatment and control groups as outlined in Section 4.1, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.