Nudges to improve learning and gender parity: Supporting parent engagement and Ghana's educational response to Covid-19 using mobile phones

Pre-analysis Plan

While recent evidence from Brazil and Ivory Coast suggests that SMS messages to nudge parents' engagement in their children's education have large effects on educational outcomes, the Covid-19 pandemic raises additional concerns. In particular, learning deficits and school dropouts are likely to increase following school shutdowns, especially among vulnerable populations such as older girls who need to work to support their families or due to early marriage, childbearing and adolescent pregnancy. A further knowledge gap relates to the optimal period of exposure to the nudges, which is critical to scale-up. This study investigates whether sending nudges to parents can improve parental engagement in child education and broader development across child age groups and gender, in the low-resource setting of Ghana, by randomly assigning whether parents receive two different versions of nudges, with one version including content promoting girls' education and addressing some common stereotypes around gender, and whether the duration of these different modalities vary between three and six months.

I. Introduction

Ghana's schools have been closed since mid-March due to COVID-19. The Ministry of Education has initiated a remote-learning program aimed at reaching students through Radio, Television, and Internet to ensure continuity of learning throughout the crisis. For remote-learning to be successful, parent involvement will be essential; yet how to best communicate with parents, and whether parents with low education levels can support remote-learning, is unclear.

Ghana's Education Strategic Plan (ESP) and the Ghana Accountability Learning Outcomes Project (GALOP) recognizes parents/guardians as key stakeholders in children's education and aims to develop a communications strategy to deliver key messages to teachers, parents, and students. While parents are required by law to enroll their children in school, the level of involvement in children's education is generally low, particularly in the poorer regions in northern Ghana.

Further, the health and economic crises resulting from COVID-19 are likely to negatively affect households' economic situations, especially the most vulnerable ones. It is possible that many children, particularly older girls, will not return to school after the crisis because of the need to work to support their families or due to early marriage and childbearing, and adolescent pregnancy.

Ghana's Human Capital Index is 0.44, meaning that a child born today can only be expected to reach 44% of his/her potential. With schools closed, the situation may become more dire. The need to find low-cost, gender-sensitive solutions to minimize disruptions to learning and schooling is urgent, especially in most-disadvantaged northern regions.

Against a low learning base, the effectiveness of remote-learning, and its role in widening or mitigating inequalities, especially of gender, remain open. For many children, the ability to learn during the crisis and beyond will critically depend on their parents' engagement. However, parents often face informational barriers to support learning effectively. Further, parent engagement may vary by child gender, due to greater opportunity costs of schooling for girls (e.g. larger involvement of girls in household or care-work), lower perceived returns to girls' education, and widespread gender bias in social norms and aspirations. Providing timely, actionable information to poor and low-educated parents, including via text-messages as a low-cost intervention, can attenuate these barriers and improve parental engagement across child-age groups and gender (Bergman, 2019). If such interventions work during and after the pandemic, where stressors are greater than under non-emergency circumstances, and in a low-resource setting, is unknown.

This pre-analysis plan summarizes the design of a field experiment designed to test **the following primary hypotheses:**

1. Do behavioral nudges to parents in the form of SMS-messages increase parent engagement in government educational remote-learning activities and in general education with their children's schooling when schools reopen?

    - Hypothesis: SMS nudges increase parental engagement in their child's education and school life during and after the school shutdown.

2. Do messages change parental beliefs about returns to education and educational expectations and aspirations for each child of target age?

    Hypothesis: SMS nudges increase parents' support for and investment in their child's education and aspirations for the future, and thus children's schooling and time devoted to educational activities should increase.

3. Do messages improve children's learning outcomes in the short-term (i.e., summer) and medium-term (i.e., following school year), and schooling outcomes (i.e., enrollment, attendance) in the medium-term?

    - Hypothesis: SMS nudges improve (or at least dampen the negative effects of school shutdown on) learning and schooling outcomes both in the short-term and medium-term, increasing grades, enrollment, attendance and grade promotion rates, and decreasing school dropout rates.

        - Assuming parental educational investments and school attendance increase, it is not clear that children's learning outcomes will improve. Given low educational quality,

attending school does not necessarily translate into improved learning. This question assesses if nudges impact other dimensions of children's development that have potential prospects for longer-term well-being.

4. Are these impacts, including parental beliefs about gender norms, more equitable across girls and boys if the messages additionally focus on gender-parity in education and in behaviors/attitudes towards girls?

- Hypothesis: Focusing on gender-parity in education and in behaviors/attitudes towards girls equalizes the impact of messages across genders.

5. Do these impacts differ for younger (5-9 years) versus older (10-15) children?

- Hypothesis: Based on their greater involvement in child labor (within or outside the household), we hypothesize that the intervention may have a stronger effect on older children by increasing the time spent in educational activities vis-à-vis time spent in labor. However, given the emergency, it may be the case that older children may increase even more their work to support the family as the educational opportunity costs may have raised substantially (e.g. caring for younger siblings, or greater involvement in farm or business), so that the intervention may have a larger effect on younger children as compared to the older ones. With regards to the intervention impacts on child learning, we hypothesize that it may have larger effects on younger children based on evidence from Cote d'Ivoire.

6. Are these impacts larger and do they persist for longer if delivered for a longer duration (6 versus 3 months)?

- Hypothesis: Increasing the duration of delivery of SMS nudges increases the persistence of impacts over time. In addition, if parents continue to be nudged, impacts in the next school year will increase.

II. Intervention

The Parental Nudges Project (PNP) is a household-level intervention designed to improve school-aged children's outcomes by engaging parents in their children's learning during and after the COVID-19 pandemic. The intervention involves two text-messages (SMS) per week sent to parents / primary caregivers in simple English with behavioral "nudges" around engaging with children's learning generally and during remote learning across grades and ages. Messages include suggestions of simple activities that promote nurturing child social-emotional development and education. No curricular knowledge is required. During school closures, suggestions related to remote-learning will be included.

As school-age children (ages 5-15 years) and particularly girls in primary schools are likely to be disproportionally affected by the crisis, a randomized subset of parents will receive messages promoting gender-equitable outcomes. Nudges will encompass reminders, encouragement and activities addressing information gaps, biased beliefs, and norms behind gender inequalities in education and broader development.

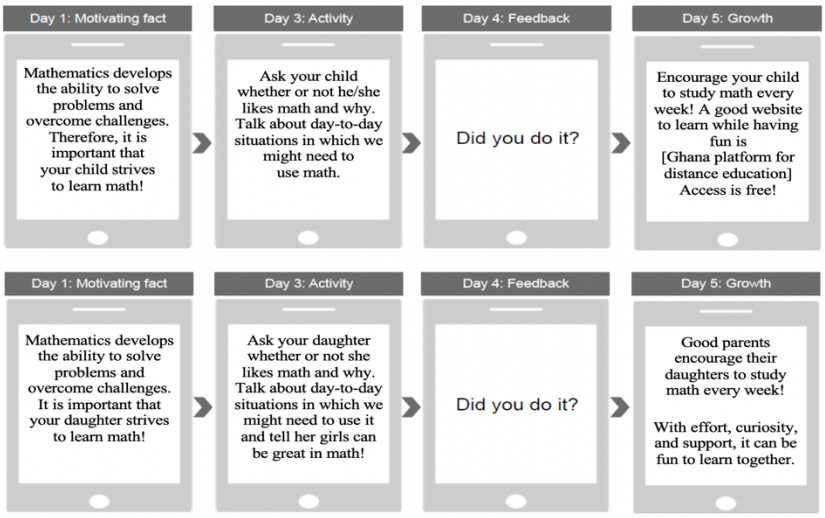
We will also vary the duration of exposure to the interventions across treatment arms. While the general parental engagement intervention has been implemented in several countries, the effectiveness of the gender component and the differing duration of exposure to messages in inducing belief and behavior change has never been tested.

The intervention is nimble, and message contents can be changed rapidly. We will adapt content as the country updates its plans to reopen schools (currently September 2020), and align them to government and World Bank remote-learning and back-to-school campaigns.

All nudges will be sent from a short code number (a 5-digit number that enables users to reply at no costs), and include the EDUQ+ tag, to clearly identify the messages under this initiative.

Eduq+ (powered by EdTech Movva) shares weekly suggestions of activities for parents to do with their children – none of them linked to curricular activities; rather, those try to bring parents closer to their children's school life by having them ask about school, discuss future plans, and share how they dealt with similar conflicts back in the day. Nudges are structured around sequences in a format inspired by READY4K!, an eight-month-long text-messaging intervention for parents of preschoolers that targets the behavioral barriers to engaged parenting (York et al., 2017). The figure below showcases two examples of the SMS sequence sent to parents assigned to the nudge program: the first sequence is not gender specific, while the second is specifically targeted at tackling gender inequalities.

Figure 1: Two examples of SMS sequence sent to parents



III. Experimental Design

The unit of randomization is households. To assign households to treatment and control, we will employ a household-level randomized controlled trial design. Households in the control group will not receive any messages during the study period. There will be no stratification as part of the randomization. Eligible and consented households identified through the Enrollment Call will be randomly assigned to receive one of the four treatment SMS text message groups or to receive no SMS text message. The randomization protocol, which will be implemented through a STATA do-file, will seek to achieve a 1:1:1:1:1 ratio across the five experimental groups. The use of a STATA do-file is to ensure that the randomization is reproducible.

The experimental groups are:

1. Treatment group 1. Behavioral nudges: Nudges to parents supporting involvement with children's learning, their child's social-emotional development, academic aspirations, and engagement in remote learning activities during the school closures and into the summer (3 months);
2. Treatment group 2. A "gender-equality boost" arm, in which some of the nudges include content promoting girls' education and addressing some common stereotypes around gender roles during the school closures and into the summer (3 months);
3. Treatment group 3. Treatment 1 implemented for 6 months into the first term of the next academic year;
4. Treatment group 4. Treatment 2 implemented for 6 months, into the first term of the next academic year;
5. Control group. No intervention / no messages.

The following surveys will be conducted (dates may change based on the evolution of the pandemic):

- Engagement Call – July/August 2020
- Parent Phone Survey I – October/Nov 2020
- Child Phone Survey I (sub-sample, N=3550) – October/Nov 2020
- Parent Phone Survey II – February 2021
- Child In-person Assessment- February 2021 (sub-sample, N=1,000; currently fund-raising to expand the sample)

The Engagement Call and Phone Surveys will be translated into the various local languages of the study regions.

The intervention will be evaluated through a household-randomized controlled trial with a goal of 2,500 households (500 in each treatment arm) in the poorest regions in Ghana (Northern, Upper East and Upper West Regions) – the exact number is still to be determined. Randomizing at the household level ensures that we take into account within-household spillovers. Within each household, beyond the parent, we will sample two children to investigate within-household spillovers and age- and gender-heterogeneity in key study outcomes: one child in the lower basic education (including early-childhood/early primary grades) (4-9 years) and one child in the upper basic education and adolescence age range (10-15 years). If there are more than two children of those target ages within the household, we will randomly select one. Ideally, we would have all children of target age in the household, but this would be challenging in a phone interview due to the shorter attention span by the interviewee when using such modality.

Each household will be assigned to one of the five treatment/control groups, as shown in table 1 below:

Table 1: Randomization strategy

	Yes (3 months)	**Yes (6 months)**	**No**
Behavioral nudges	500 households 500 caregivers 1,000 school-age children	500 households 500 caregivers 1,000 school-age children	500 households 500 caregivers 1,000 school-age children
Gender-equality boost	500 households 500 caregivers 1,000 school-age children	500 households 500 caregivers 1,000 school-age children	

Our sample is drawn from two previously completed studies. First, an impact evaluation of the Communications for Development (C4D) study[1] (2012-2016), launched by the Ghana Health Service with funding from UNICEF in 12 districts of the three poorest regions of Ghana (these were, at the time, Northern, Upper East, and Upper West, and later they have split in further regions). The sample included mothers with a child aged 0-5 years recruited in 2012. The program relied on voice messages directly delivered to female respondents through their cell phones, and the sample has high rates of mobile phone ownership (83%). In 2016, there were over 4300 families with children under age 7 years in this sample, indicating that we will have sufficient sample size for our target 2,500 families, even in a very negative scenario in which 40% of them have changed phone number and we cannot trace them in other ways[2]. Second, we will be employing households from a subsample of the Ghana Panel Survey[3]

[1] https://www.poverty-action.org/study/communication-development-ghana

[2] Note: The figure of 40% untraceability rates is based on evidence of attrition in Sierra Leone and Liberia during the Ebola crisis.

[3] https://www.src.isr.umich.edu/international/ghana-socioeconomic-panel-study/

(specifically, the Graduating Out of Poverty sub study) from the same regions in order to obtain our desired sample size.

In summary, the sample size for this study takes into account the required sample for statistical inference, possible attrition rates, and piloting of the intervention. The sample comprises of 2,500 households across the five experimental groups, 2,500 caregivers and 5,000 school-age children (2 from each household) from the selected households. Current funding permits us to conduct phone-surveys with 3,550 children, and in-person assessments for 1,000 children.

With such sample sizes, we will able to detect the following treatment effects:

OUTCOME 1: Parent's engagement in child schooling / learning **OUTCOME 2:** Parent's educational aspirations and expectations for each child **OUTCOME 3a:** Child schooling outcomes (enrollment and attendance, parent-report)		
	Sample size	
Treatment 1	500 households	
Treatment 2	500 households	
Treatment 3	500 households	
Treatment 4	500 households	
Comparison	500 households	
Assumed attrition	10%	
	Power: 0.80 For short-term follow-up (October), we will pool treatments 1 and 3, and treatments 2 and 4 to double our sample size. For the longer-term follow-up (February), where we compare differences in duration of exposure, we will analyze these arms separately. We provide estimates of MDES for both. We assume an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years.	**Power: 0.90** For short-term follow-up (October), we will pool treatments 1 and 3, and treatments 2 and 4 to double our sample size. For the longer-term follow-up (February), where we compare differences in duration of exposure, we will analyze these arms separately. We provide estimates of MDES for both. We assume an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years.
Number of clusters	Follow-up 1 and 2: 1350 (pooling treatments 1 and 3,	Follow-up 1 and 2: 1350 (pooling treatments 1 and 3, and treatments 2 and 4, relative to the control).

	and treatments 2 and 4, relative to the control). Follow up 3: 900 (for a 2-way comparison between treatment and control) (after 10% attrition accounted for)	 Follow up 3: 900 (for a 2-way comparison between treatment and control) (after 10% attrition accounted for)
Size of cluster	1 parent	1 parent
Significance level (alpha)	0.05	0.05
Sample size	Follow-up 1 and 2: 1350 Follow up 3: 900 for parents	Follow-up 1 and 2: 1350 Follow up 3: 900 for parents
Minimum detectable effect	Follow-up 1 and 2: 0.14 SD Follow up 3: 0.18 SD	Follow-up 1 and 2: 0.17 SD Follow up 3: 0.21 SD
OUTCOME 3b: Child schooling outcomes (enrollment and attendance, child-report phone survey)		
	Sample size	
Treatment 1	710	
Treatment 2	710	
Treatment 3	710	
Treatment 4	710	
Comparison	710	
Assumed attrition	10%	
	Power: 0.80 For short-term follow-up (October), we will pool treatments 1 and 3, and treatments 2 and 4 to double our sample size. We assume an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years. For children nested in households, we assume an ICC of 0.10 given the wide age range.	**Power 0.90** For short-term follow-up (October), we will pool treatments 1 and 3, and treatments 2 and 4 to double our sample size. We assume an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years. For children nested in households, we assume an ICC of 0.10 given the wide age range.
Number of clusters	Follow-up 1: 958 (assuming 10% household attrition)	Follow-up 1: 958 (assuming 10% household attrition)

Size of cluster	2 children / household	
Significance level (alpha)	**0.05**	**0.05**
Sample size	1,917 (assuming 10% household attrition)	1,917 (assuming 10% household attrition)
Minimum detectable effect	0.057 SD	0.066 SD
OUTCOME 4: Child learning and developmental outcomes in-person assessments on subset (follow up 2 only)		
	Sample size	
Treatment 1	100 households, 200 children	
Treatment 2	100 households, 200 children	
Treatment 3	100 households, 200 children	
Treatment 4	100 households, 200 children	
Comparison	100 households, 200 children	
Assumed attrition	**10%**	
	Power: 0.80 We include an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years. For children nested in households, we assume an ICC of 0.10 given the wide age range**.**	**Power 0.90** We include an R-squared of 0.1, a conservative estimate given that we have three waves of data over five years. For children nested in households, we assume an ICC of 0.10 given the wide age range.
Number of clusters	180 households	180 households
Size of cluster	2 children	2 children
Significance level (alpha)	0.05	0.05
Sample size	360 children	360 children
Minimum detectable effect	0.13 SD	0.15 SD

IV. Measures

We will collect measures at both the parent/guardian and child level.

At the **parent/guardian level** we will collect the following:

1. **Parent's engagement in child schooling and learning** (both with regards to remote-learning activities and general learning when/if schools reopen)

We will ask through three (including the enrolment call) phone interviews a battery of questions about parent's engagement on child education (e.g. reading to child, talking with the child, playing, etc.), and with government remote-learning activities. We will build a summary measure of parent engagement in child education (in general and during remote-learning) by counting the activities in which the parent report s/he is involved, similarly to Banerji et al. (2017). Further, following the same study, we will measure parent time use as we hypothesize that it will also change as part of the intervention.

2. **Parent's educational aspirations and expectations for each child**

This will be undertaken through three (including the enrolment call) rounds of phone surveys. We will use a module developed in the 'Quality Preschool for Ghana' survey (PI: Wolf) to measure parent educational aspirations and expectations. In this survey, parent's educational aspirations are defined on the basis of the following question: 'Ideally what level of formal education would you like {child's name} to complete?'. Also, parent's educational expectations ask the parent which level of education the child is expected to complete considering actual constraints to her education. Based on actual data we will gather, we will decide a meaningful cutoff point, though we are inclined to follow previous literature (e.g. Favara 2017) that identified high aspirations as completing tertiary education.

3. Gender norms

This will be assessed through three (including the enrolment call) rounds of phone surveys to parents. Gender norms will be measured through the "Gender norms and attitudes scale" (Waszak et al, 2000), which measure egalitarian beliefs about male and female gender norms. Specifically, the scale assesses agreement/disagreement with a number of statements related to the promotion of equity for girls and women and the maintaining the rights and privileges of men (14 items/2 subscales).

4. **Child schooling outcomes** (enrollment and attendance reported by parents)

This will be assessed through three (including the enrolment call) rounds of phone surveys to parents. This question aims at asking if the child was enrolled in the academic year (2019/20) in the first phone survey. Later, in the second phone survey, we will ask whether child was enrolled in the 2019/2020 (control question to check consistency or add for households that we could not interview in the first phone survey wave). Further, we will ask in the two first terms of academic year 2020/21 (i.e. mid-October 2020, mid-February 2021) about enrollment and attendance.

At the **child level** we will measure schooling, learning (literacy and numeracy), and developmental outcomes (social-emotional development, educational aspirations, time use).

We will undertake these measurements through a mix of phone and direct assessments with children.

Specifically, in October 2020, we will conduct a brief phone survey with a larger sub-sample of children in the study. ~~We are developing a short scale (max. 10-item scale for each learning competence), with which~~ We will assess very basic literacy and numeracy knowledge. There will be different scales for children in the two age groups: 5-9 years and 10-15 years, which we will devise by selecting items for

administration through phone surveys from the assessments used in the PI's Quality Preschool for Ghana Study (funded by SIEF) and an impact evaluation on school feeding that was also conducted in the north of Ghana (Aurino et al., 2018), and consider questions from Ghana's national exam. We will also draw on other item data banks such as Young Lives. In addition, we will assess children's educational aspirations (as measured by using an adaptation of the Quality Preschool for Ghana survey module focusing on children, which asks the child which level of formal education she would like to achieve if she does not have constraints to her schooling), and current school enrollment and attendance status (the latter to confirm data we collect in the household survey). Further, we will ask older children about their time use (for younger children, we will ask their parents in the phone survey) by again adapting the Young Lives survey questions about the time the child has spent in a typical days in educational activities, house- and care-work, work out of the household, leisure, etc. We hypothesize that time use is an important and potentially highly-gendered pathway for intervention impact.

Phone-based learning assessments for children are a relatively new idea and have not been widely used. We have been in discussions with a group of researchers around the world all interested in using such assessments given that it is difficult to conduct direct assessments during the pandemic. We will be sure to follow ethical principles in such assessments, including ensuring that children are protected (as outlined in Angrist et al., 2020: Principles for Phone-Based Assessments of Learning). For example, assessors can ensure that parents are aware that tests have no direct consequences for children (i.e., these are low-stakes assessments), so that adults do not discipline children if they overhear low performance. Supervisors can also monitor a sample of calls to make sure enumerators are interacting appropriately with children and youth. We are working closely also with the World Bank to develop and assess the psycho-metric properties of such scales.

In February 2021, we will conduct in-depth direct assessments on a sub-set of children (N=500 households, 1,000 children) – to be confirmed – of child learning and development. To measure child learning and developmental outcomes, will build on assessments from previous studies of the PI's in Ghana (including in Northern Ghana) to compile these measures. The target children in the study are in two age brackets: 5-9 years and 10-15 years. For the younger children, we will draw from assessments used in the Quality Preschool for Ghana Study (funded by SIEF) including items from the International Development and Early Learning Assessment (IDELA) to measure literacy, numeracy, and social-emotional development through an adaptation of the International Socio Emotional Learning Assessment (ISELA). We will also use items from the EGRA/EGMA. Importantly, the QP4G study is currently ongoing, with children now in P4 and P5. New assessments have been developed drawing on IDELA, EGRA and EGMA, and we will align these assessments as much as possible.

For the older children, we will draw on data collected from a study of Aurino's in the Northern Region to assess literacy and numeracy plus the Young Lives data. Further, we will measure child time use, educational aspirations and psycho-emotional development (see above).

V. Outcomes

We will document the effects of the treatments on the following outcomes:

Primary outcomes

1. Parent involvement in education and remote learning (depending on when the intervention actually starts)
2. Parent expectations and aspirations for their child's learning and schooling
3. Parent beliefs about gender norms
4. Children's school enrollment and attendance (as measured by parent's survey)

Secondary outcomes

1. Children's literacy and numeracy skills
2. Child Social-emotional skills
3. Educational aspirations
4. Parent and child time use

1. Parent's involvement in their child's education, as reported by the parent, in terms of time spent for school-related activities (e.g. reading to child, talking with the child, playing, etc.), and with government remote-learning activities, assessed through phone surveys;
2. Parent self-reported expectations and aspirations for their child's education, with and without actual constraints to her/his education, assessed through phone surveys;
3. Parent's self-reported beliefs about gender norms, assessed through phone surveys;
4. Children's school enrollment and attendance, as reported by the parent and children, assessed through phone surveys;
5. Children's literacy and numeracy skills, assessed through a mix of phone and direct assessments with children using a short scale for each (with different scales for children in the two age groups: 5-9 years and 10-15 years);
6. Children self-reported social-emotional skills, educational aspirations, and current school enrollment and attendance status, assessed through phone surveys;
7. Parents and children self-reported time-use which consists in questions about the time the child has spent in a typical day in educational activities, house- and care-work, work out of the household, leisure, etc (parents will answer for younger children), assessed though phone surveys.

We will be able to test for balance across treatment arms on household (e.g. age of the caregiver, number of family members, physical aspect of the house as reported by the respondent, access to drinking water

and other necessary resources, etc) and child development characteristics (e.g. physical/motor development, cognitive and language, social-emotional, etc) by conducting mean comparison tests. Since information on such outcomes was already collected in the context of the Communications for Development study in 2016 and for the Graduating Out of Poverty study in 2019, we will be able to ensure balance across groups even before the start of this intervention.

VI. Empirical analysis

The randomized design allows for the identification of causal effects of the interventions on parents and children by comparing mean outcomes between the randomized treatment arms. The analysis will follow an intention-to-treat approach, using econometric analysis for all the relevant outcomes of the intervention.

For each parent outcome described above (Section V), we will estimate the following ordinary least squares regressions indexed by parent *p* from household *h* and survey *s*:

$$Y_{p,h,s} = \beta_0 + \beta_1 NShort_h + \beta_2 GShort_h + \beta_3 NLong_h + \beta_4 GLong_h + \beta_5 Y_{p,h,0} + \beta_6 X_{h,0} + \beta_7 G_h + \theta_r + \varepsilon_{h,s}$$

Where:

- $Y_{p,h,s}$ is the outcome variable for parent *p* in household *h* and survey round s;
- NShort, GShort, NLong, and GLong are indicator variables assuming the value of 1 if the household has been randomly assigned to any of the treatment arms (Arm1: Nudges Short duration; Arm2: Gender boost short duration; Arm3: Nudges long duration; and Arm4: Gender boost long duration). We note that for the first parental assessment (the one conducted at the end of the three-month implementation of treatments 1 and 2), we will pool samples from arms 1 and 3, and 2 and 4 to estimate effects. For the final follow-up, T_p will include four treatment dummies to treatment arms 1, 2, 3 and 4 separately.
- $Y_{p,h,0}$ is the baseline outcome variable for parent *p* in household *h* (when available)
- $X_{h,0}$ is a vector of parent and household controls should there is lack of balance in the randomization

  G_h is an indicator variable assuming the value of 1 for households belonging to the GUP sample, 0 otherwise[4];
- θ_r are region fixed effects

$\varepsilon_{h,s}$ is individual error term

[4] The C4D sample was part of a randomized intervention trial. The treatment had no impacts. As a robustness check, we will also add an indicator for treatment status in the C4D sample.

We are interested in testing $\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$.

For each child outcome described above (Section V), we will estimate the following ordinary least squares regressions indexed by child *c,* living in household *h* and survey *s*:

$$Y_{c,h,s} = \beta_0 + \beta_1 NShort_h + \beta_2 GShort_h + \beta_3 NLong_h + \beta_4 GLong_h + \beta_6 X_{h,0} + \beta_7 G_h + \theta_r + \varepsilon_{c,h,s}$$

Where:

- $Y_{c,h,s}$ is the outcome variable for child *c,* living in household *h* and survey *s*
- NShort, GShort, NLong, and GLong are indicator variables assuming the value of 1 if the household has been randomly assigned to any of the treatment arms (Arm1: Nudges Short duration; Arm2: Gender boost short duration; Arm3: Nudges long duration; and Arm4: Gender boost long duration). As for the parents' outcomes, for the first child assessment (the one conducted at the end of the three-month implementation of treatments 1 and 2), we will pool samples from arms 1 and 3, and 2 and 4 to estimate effects. For the final follow-up, T_p will include four treatment dummies to treatment arms 1, 2, 3 and 4 separately.
- $X_{h,0}$ is a vector of parent and household controls should there is lack of balance in the randomization
- G_h is an indicator variable assuming the value of 1 for households belonging to the GUP sample, 0 otherwise[5];
- θ_r are region fixed effects
- $\varepsilon_{c,h,s}$ is individual error term, clustered at the household level.

We are interested in testing $\beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$.

We will consider the possibility of using the October child phone survey learning assessments as baseline values for the final round of learning outcomes.

Heterogeneity

We will assess differences in treatment effects by child gender and age group (younger children=5-9, older children=10-15 years). We will also evaluate heterogeneity in treatment effects by main parents' educational level, baseline household poverty[6], and parents' baseline gender norm attitudes.

Spillovers

[5] The C4D sample was part of a randomized intervention trial. The treatment had no impacts. As a robustness check, we will also add an indicator for treatment status in the C4D sample.

[6] The exact measurement of this indicator will depend on what is available and comparable in both the C4D and GUP samples. We anticipate it will likely be a household asset index. Alternatively, we consider using household food security as collected in the enrolment call.

We anticipate that in the case of this specific evaluation we may have "social interactions" spillovers type (as defined by Angelucci and Di Maro 2015) at different spatial levels: classroom, neighborhood and within the household. First, previous evidence from a similar intervention in Brazil highlights that there are very strong spillover effects within classrooms due to social interactions among children and between parents of children in the same class (Bettinger at al., 2020; Lalive and Cattaneo 2009). Since during the school shutdown and the summer period there are no classes, this issue becomes less relevant. For the treatment groups for which the nudges continue after classes resume, to the extent that, in our current sample, few parents overlap in terms of having children in the same classrooms, then this issue becomes minor as we randomize at the household level (of course that would be false if there is large overlap). We will track the class and the school the children are attending in order to have a magnitude of this potential spillover and, if that is large, to estimate the extent to which the treatment has spillovers on peers that were part of the comparison group or that were not part of the experiment. Based on previous evidence on peer effects of interventions that promote educational outcomes (e.g. Lalive and Cattaneo 2009), we expect these spillovers to be positive. Second, there may be neighborhood peer effects on parents' outcomes and children's schooling, such as the ones found in Bobonis and Finan (2009). Specifically, it may be likely that treatment peers (in terms of both parents and children) have influence on their control peers through their behaviors, or may share the information they received in the text. In the case of Bobonis and Finan, effects of those types of social interactions were concentrated among poorest households. Preliminary results from the Cote D'Ivoire trial show that there are positive spillovers for households within a 5-km radius, driven by younger children (early primary grades). We will try to examine spillovers at the neighborhood level as well by directly including in the questionnaire some questions about whether respondents asked for help understanding the SMS, who they asked (teacher, relative, friends, etc.), whether they asked the neighbor, if they discussed the contents of SMS with other parents in the neighborhood etc. We will also ask to the control group if anyone has shared SMS-messages related to children and education with them. Further, we can rely on household GPS coordinates[7] to more precisely estimate spillovers at the neighborhood level. Third, there may be spillovers within the household itself (which is intended as the intervention targets all children in the household). We have included this possibility directly in our research design by sampling two child per household of the target compulsory school age.

To analyze spillover effects empirically at the village level, we will use GPS data and examine the share of treated households within (i) 5 km, (ii) 5-10 km, (iii) 10+ km from each household as a predictor of child schooling and learning outcomes. This allows us to examine the share of households in proximity to each household as a predictor of outcomes.

[7] Available for GUP sample, in the case of C4D this has to be confirmed.

REFERENCES

Angrist, N., Bergman, P., Evans, D. K., Hares, S., Jukes, M. C., & Letsomo, T. (2020). Principles for Phone-Based Assessments of Learning. Working Paper.

Aurino, E., Gelli, A., Adamba, C., Osei-Akoto, I., & Alderman, H. (2018). *Food for thought? Experimental evidence on the learning impacts of a large-scale school feeding program in Ghana* (Vol. 1782). Intl Food Policy Res Inst.

Banerji, R., Berry, J., and Shotland, M. (2017). The Impact of Maternal Literacy and Participation Programs: Evidence from a Randomized Evaluation in India. *American Economic Journal: Applied Economics.*

Bergman, P. (2019). How behavioral science can empower parents to improve children's educational outcomes. *Behavioral Science & Policy*, 5(1), 52-67.

Bettinger, E., Cunha, N., Lichand, G., & Madeira, R. (2020). "*Are Effects of Informational Interventions Driven by Salience?*". Working Paper.

Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2017). Supporting Parenting through Differentiated and Personalized Text-Messaging: Testing Effects on Learning during Kindergarten. CEPA Working Paper No. 16-18. *Stanford Center for Education Policy Analysis.*

Favara, M. (2017). Do Dreams Come True? Aspirations and Educational Attainments of Ethiopian Boys and Girls. *Journal of African Economies.*

Lichand, G., & Wolf, S. (2020). "*Are Parenting Interventions Transferable Across Settings? Evaluating Key Constraints in Sub-Saharan Africa*". Working Paper.

Waszak, C., Severy, L., J., Kafafi, L., Badawi, I. (2001). Fertility Behavior and Psychological Stress: The Mediating Influence of Gender Norm Beliefs among Egyptian Women. *Psychology of Women Quarterly*, 25 (2001)

Wolf, S., Aber, J. L., Behrman, J. R., & Tsinigo, E. (2019). Experimental impacts of the "Quality Preschool for Ghana" interventions on teacher professional well-being, classroom quality, and children's school readiness. *Journal of Research on Educational Effectiveness*, 12(1), 10-37.