

Nudges to Improve Learning and Gender Parity

Preliminary findings on supporting parent–child educational engagement during Covid-19 using mobile phones

Date March 2022

Authors Elisabetta Aurino
Edward Tsinigo
Sharon Wolf

DOI 10.53832/edtechhub.0083



THE WORLD BANK



About this document

Recommended citation

Aurino, E., Tsinigo, E., & Wolf, S. (2022). *Nudges to Improve Learning and Gender Parity: Preliminary findings on supporting parent–child educational engagement during Covid-19 using mobile phones* [Technical Report]. EdTech Hub.
<https://doi.org/10.53832/edtechhub.0083>. Available at <https://docs.edtechhub.org/lib/PWU63GQS>. Available under Creative Commons Attribution 4.0 International.

Licence

Creative Commons Attribution 4.0 International
<https://creativecommons.org/licenses/by/4.0/>.
You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) for any purpose, even commercially. You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Notes

EdTech Hub is supported by UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, World Bank and UNICEF. The views expressed in this document do not necessarily reflect the views of UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, World Bank, and UNICEF.

Abstract

In this study we evaluate a digital intervention to improve low-literate caregivers' engagement with their children's education and development in rural Ghana during the Covid-19 pandemic. The programme was a text-message-based behavioural change intervention for parents / caregivers that aimed to improve caregiver engagement in children's educational activities, caregiver beliefs about returns to education, as well as children's learning, enrollment, attendance, and gender parity in education. This household-randomised trial, conducted in the North East, Northern, Savannah, Upper East, and Upper West regions of Ghana, tested four variations of the intervention, varying both duration and a gender-parity focus. Households were randomised to one of five conditions: (i) regular behavioural nudges, 12 weeks; (ii) gender-boost behavioural nudges, 12 weeks; (iii) regular behavioural nudges, 24 weeks; (iv) gender-boost behavioural nudges, 24 weeks; or (v) control. The interventions were implemented from January to April 2021 (for the 12-week groups) and January to June 2021 (for the 24-week groups). We collected data at midline (April–June 2021) and endline (August–September 2021). Our preliminary results suggest that a short, light-touch, SMS-based intervention can change caregiver behaviours and child outcomes in a rural, low-literate sample. However, the results were complex and intervention effectiveness depended on the caregiver having minimum levels of schooling. For caregivers with no education (65% of the sample), the intervention only increased caregiver expectations on reaching the desired level of education, especially among girls, but reduced educational engagement and some measures of children's school enrollment and attendance. Educational engagement among Ghanaian caregivers is low relative to peer countries ([Bornstein & Putnick, 2012](#); [Bornstein & Putnick, 2012](#)). The findings suggest that caregivers may need a base level of capital and resource (e.g., exposure to formal education) to enact the messages and increase their educational engagement with their children. Without this base level of capital, messages may increase caregivers' aspirations for their children without providing enough support to change educational investments in positive ways.

Contents

Abstract	2
Lists of tables and figures	4
Abbreviations and acronyms	4
1. Introduction	5
1.1 Background to the study	5
1.2 Purpose / aims of this evaluation	6
1.3 Context of the study	6
1.4 What this study adds to the knowledge base	7
1.5 Implications for policy and practice	9
1.7 Structure	11
2. Literature review	12
3. Methodology	14
3.1 Research questions	14
3.2 Research framework / methodology	15
3.3 Research instruments / tools	18
3.4 Research design	24
3.5 Stakeholders	25
3.6 Ethical considerations	27
3.7 Challenges	28
4. Results	32
4.1 Data	32
4.2 Limitations	32
4.3 Results	33
5. Policy Implications	50
6. Conclusion	52
Bibliography	55

List of Tables

Table 1. Background information on the primary caregiver collected at baseline.	16
Table 2. Treatment impacts on primary outcomes at midline and endline.	35
Table 3. Impact heterogeneity by caregiver education.	39
Table 4. Heterogeneity — child gender.	44
Table 5. Impact heterogeneity by child age group.	47

List of Figures

Figure 1. Study timeline	17
Figure 2. Counters used during child assessment	22
Figure 3. Sample messages	27
Figure 4. Challenges in properly setting up child assessment	31

Abbreviations and acronyms

GES	Ghana Education Service
GALOP	Ghana Accountability for Learning Outcomes Project
IDELA	International Development and Early Learning Assessment
JHS	Junior high school
MICS	Multiple Indicators Cluster Survey
MoE	Ministry of Education
SHS	Secondary high school
SMS	Short Message Service – text message

1. Introduction

In this section, we first provide background to the project, followed by details of the study and context, as well as a detailed presentation of the research questions and implications of the project for policy and practice.

1.1 Background to the study

The Covid-19 pandemic led to unprecedented extended school closures around the globe. Ghana's schools were closed from March 2020 through January 2021. In addition to evidence of growing inequalities in access to remote learning ([↑Innovations for Poverty Action, 2020](#)), there is concern that as schools have re-opened, the most vulnerable children are the ones least likely to have returned to school.

Negative macroeconomic shocks, such as recessions, health crises, or droughts increase schooling opportunity costs, inducing decreases in educational investments, especially for girls ([↑Björkman-Nyqvist, 2013](#)). Further, emerging evidence shows that the increased caregiving responsibilities stemming from the Covid-19 crisis is disproportionately assumed by women and girls ([↑Nesbitt-Ahmed & Subrahmanian, 2020](#)), risking progress towards gender parity and chances of returning to school. The need for low-cost, gender-sensitive solutions to minimise disruptions to learning is urgent.

Caregiver involvement in children's education in Ghana is low ([↑Bornstein & Putnick, 2012](#); [↑McCoy et al., 2018](#)), particularly in the more disadvantaged northern regions. Engaging parents / guardians / caregivers (henceforth defined as 'caregivers') to ensure an equitable return to school for boys and girls and learning opportunities at home is key. Yet disadvantaged caregivers face informational and social norms / expectations, which are barriers to supporting learning ([↑Bergman, 2019](#)). Many have differing perspectives about educational investment returns, especially for girls, or about what is socially expected regarding educational engagement. Providing timely, actionable information to poor and low-educated caregivers, including via text messages as a low-cost intervention, can attenuate these barriers and improve caregiver engagement across child-age groups and gender ([↑Bergman, 2019](#)). Whether such interventions work during a pandemic, where stressors are greater than under non-emergency circumstances, is unknown.

1.2 Purpose / aims of this evaluation

We evaluate an intervention that aims to improve caregiver engagement in education among low-literate parents in rural Ghana and provide the first

evidence on heterogeneity by implementation duration. First, we provided timely, actionable information to caregivers via text messages ('nudges') to attenuate caregiver behavioural and informational barriers to learning. Second, for a subset of caregivers, messages were tailored to address differing perspectives and norms about girls' education, aiming to equalise educational opportunities, caregivers' investments, and time use between learning and care-work. Finally, we tested whether there are differences in impacts based on implementation duration.

1.3 Context of the study

Our study took place in three rural regions of Ghana. Ghana's Human Capital Index is 0.44, meaning that a child born today can only be expected to reach 44% of their potential. Further, 70% and 80% of children in Grades 2 and 4 respectively, cannot read a simple word or perform basic arithmetic operations ([↑World Bank, 2018](#)). Gender inequalities in learning are wide, as shown in [↑World Bank \(2018\)](#). For instance, while national re-enrollment rates as schools re-opened are 97%, 60% of dropouts were girls ([↑World Bank et al., 2021](#)).

Our sample includes households with school-aged children (5–17 years) randomly sampled from two panel studies: The Ghana Panel Survey ([↑ECG & ISSER, no date](#)) and the Communications 4 Development (C4D) ([↑Fink et al., 2019](#)) sample in the former Northern and Upper East and West Regions. These are among the poorest regions; they are rural and educationally deprived; gender differences are also more pronounced as compared to regions in southern Ghana ([↑Osei-Assibey, 2014](#)).

Our implementing partner, Movva, sent biweekly nudges via text messages (SMS) to caregivers in simplified English, with reminders and encouragement messages targeted at habit formation. Messages included suggestions of activities that aid in social-emotional development and nurturing education and do not require curricular knowledge. Specific suggestions related to supporting children's return to school were included. The messages included reminders, encouragement and activities engaging information gaps, different understandings behind gender inequalities in education, and broader development. Each core module was delivered over two weeks, with two messages per week, and structured as a specific sequence based in behavioural economic theory to induce behaviour change — a motivating fact (Message 1), a suggested activity (Message 2), an interactive message (Message 3), and a growth message (Message 4). Movva's programme has been implemented in several countries such as Brazil (including during the Covid-19 crisis) and Cote d'Ivoire for as little as one month to an entire school year ([↑Bettinger et al., 2021](#); [↑Lichand & Wolf, 2021](#); [↑Lichand et al., 2021](#)), but

variation in the duration of exposure to messages inducing belief and behaviour change has never been tested.

1.4 What this study adds to the knowledge base

Our study makes several contributions to the knowledge base about nudge-based caregiver / parenting programmes. First, most studies testing SMS-nudge interventions to parents have been conducted in middle- or high-income countries, where caregivers likely have a base level of human capital and formal education; we conducted this study in a low-literature and low-education rural sample, growing the evidence base on whether and when these types of intervention to support child education and caregiver engagement are effective. Second, we tested whether a focus on gender-parity can make a larger difference for girls' education compared to behavioural nudges generally about parenting for all children. Third, we tested this strategy during a public health crisis — namely, the Covid-19 pandemic — and assessed the impacts of an SMS-based nudge program during a relatively stressful macro-economic and health crisis. Fourth, we tested whether the duration of messages matters in behaviour change, a question that has never been tested before.

This field experiment tested the following primary hypotheses:

1. Do behavioural nudges to caregivers in the form of SMS messages increase caregiver engagement in educational activities?

Hypothesis: SMS nudges increase caregiver engagement in their child's education and school life.

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

Hypothesis: SMS nudges increase caregivers' 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 schooling outcomes (i.e., enrollment, attendance)? Are there changes in learning outcomes?

Hypothesis: SMS nudges improve schooling outcomes both in the short-term and medium-term and decrease school dropout rates.

We focus on schooling as primary outcomes (rather than learning), as 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 caregiver beliefs about gender norms, more equitable across girls and boys if the messages additionally focus on gender parity in education and in behaviours / attitudes towards girls?

Hypothesis: Focusing on gender-parity in education and in behaviours / attitudes towards girls equalises 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 labour (within or outside the household), we hypothesise 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 labour. However, given the emergency, it may be the case that older children may increase their work, even more, to support the family as the educational opportunity costs may have risen substantially (e.g., caring for younger siblings, or greater involvement in farm or business), so that the intervention may have a greater effect on younger children as compared to the older ones. With regard to the intervention impacts on child learning, based on evidence from Cote d'Ivoire ([Lichand & Wolf, 2021](#)), we hypothesise that it may have greater effects on younger children.

6. Are these impacts greater and do they persist for longer if delivered for a longer duration (14 vs 12 weeks)?

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

Our design is a household-randomised trial. Randomisation was stratified within each of the three regions. Our final sample included 2,640 households randomly assigned to:

1. Nudges to caregivers supporting involvement with children's learning, child's social-emotional development, academic aspirations, and engagement (12 weeks).

2. A 'gender-parity boost' arm to caregivers of both boys and girls, in which some of the nudges include content promoting girls' education and addressing some common stereotypes around gender roles (12 weeks).
3. Treatment (1) implemented for 24 weeks into the first term of the next academic year.
4. Treatment (2) implemented for 24 weeks, into the first term of the next academic year.
5. Control group.

1.5 Implications for policy and practice

Our preliminary results suggest that a short, light-touch, SMS-based intervention can change caregiver behaviours. However, intervention effects vary widely by caregiver and child characteristics.

When we examined average results in the full sample, the interventions operated counter to our hypotheses, and decreased caregiver engagement, decreased self-reported school enrollment and attendance, decreased caregiver mental health, and decreased children's academic skills. These negative effects appear to be concentrated for less-advantaged caregivers and children — specifically, caregivers with no formal education, girls, and younger children. On the other hand, caregivers who had some schooling, increased their home and school engagement, consistently with our theory of change.

These findings suggest that caregivers may need a base level of resources (e.g., exposure to formal education) to enact the messages into positive behavioural changes for their children. Without this base level of human capital, the messages can backfire, as they may increase caregivers' aspirations for their children without providing enough support to change investments in positive ways. Indeed, in our study, the caregivers with no formal schooling reported a reduction in parental self-efficacy as a result of receiving these messages, while educated caregivers reported increases in this self-efficacy. This finding is in line with recent evidence that the digital version of a programme that was previously found to be effective, *reduced* caregiver mental health and increased stress, particularly among male caregivers during the pandemic in El Salvador ([↑Amaral et al., 2021](#)).

Longer-term follow-ups would be important to understand how long these changes persisted after the programme was no longer being delivered to caregivers.

These results contribute to a small but growing evidence base about SMS-based nudge interventions to parents and caregivers (see [↑Bergman, 2019](#) for a review). Importantly, the majority of studies that have found these types of programmes to improve parenting and child outcomes have been concentrated in middle- or high-income country contexts ([↑Bergman, 2019](#)). Our study is one of the first to test this type of programme in a rural, low-income, African setting, and during a public health and economic crisis.

Our findings suggest that careful consideration of the broader context of caregivers' and children's lives is needed to ensure programmes are tailored in ways that ultimately support caregiver investments in child education, caregiver-child relationships, and children's education. This attention to the context is especially pivotal in stressful times such as the current Covid-19 pandemic, which may be exacerbating existing poverty-related stressors.

We also note that our sample was predominantly Muslim, and that midline data collection occurred during Ramadan when many caregivers and many older children were fasting. In addition, our endline data collection took place during the harvesting season, and many children were working in the fields to earn money for school fees to return to school. Thus, the timing of both midline and endline data collection occurred at particularly unusual or busy times of the year. Our results suggest that the macro-context in which families receive these messages may be key to consider when they can be effective and when they may cause additional stress for caregivers. It is possible that the broader context was quite challenging for families at the time the programmes were implemented — e.g., fasting, economic hardship due to the pandemic — and that these messages related to parent and child investments caused additional stress for parents in ways that backfired.

We found that for caregivers with no formal schooling, educational expectations increased, making them more optimistic about the educational prospects for their children. In other words, the intervention may be successfully tackling an aspiration failure among the poor, which could lead to a self-sustaining poverty trap ([↑Ray, 2006](#)). However, these caregivers do not seem to have the necessary human capital resources to *act* upon the nudges, e.g., by shifting their involvement in child education at home and school. By contrast, nudges seem to be effective in improving engagement among caregivers who have a minimum level of education. There are overall very low schooling levels among caregivers: as noted, only 35% of caregivers have some schooling, and among those caregivers who attended school at some point, around half have at most completed primary. Thus, the programme seems to be effective among caregivers with a very low level of education, with the condition that they have at least *some* education.

1.7 Structure

The rest of the report is organised as follows. In [Section 2](#) we present related literature on the subject. Next, in [Section 3](#) we discuss the methodology for the study. The results of the impact evaluation are then presented in [Section 4](#). Finally, in [Section 5](#), we highlight the policy implications and draw conclusions from the findings in [Section 6](#).

2. Literature review

Ghana's Human Capital Index is 0.44, meaning that a child born today can only be expected to reach 44% of their potential. Further, 70% and 80% of children in Grades 2 and 4 respectively cannot read a simple word or perform basic arithmetic operations ([↑World Bank, 2018](#)). The need to find low-cost solutions to minimise disruptions to learning and schooling in the aftermath of school closures is urgent, especially in the most-disadvantaged northern regions. Caregiver engagement is a key input that supports children's school persistence and learning outcomes. Caregiver engagement may vary by child gender, due to different opportunity costs of schooling for girls and boys (e.g., larger involvement of girls in household or care-work, or greater time spent in work outside the household for boys), lower perceived returns to girls' education, and widespread gender bias in social norms and aspirations ([↑Alderman & King, 1998](#)). Providing timely, actionable information to poor caregivers with a low level of education, including via text-messages as a low-cost intervention can attenuate these barriers and improve caregiver engagement across child-age groups and gender ([↑Bergman, 2019](#)). If such interventions work during and after a pandemic, where stressors are greater than under non-emergency circumstances, is unknown.

Research on remote learning programmes shows that caregiver participation is key to the learning of children, although this literature is mostly from high-income settings, leaving an important gap for low-income contexts. For instance, [↑Myers et al. \(2018\)](#) note that co-viewing remote classes provides essential support for children to respond to and learn from video chat interactions. Their findings suggest that children depend primarily on live social partners to make sense of their media experiences. Similarly, [↑Troseth et al. \(2006\)](#) suggest that learning deficits from televised-learning programmes experienced by children can be overcome through caregiver–child interaction. Additional research equally confirms the benefits of using text messages (SMS) to encourage caregiver–child engagement ([↑Doss et al., 2017](#)), and a recent review of the literature found that such programmes can substantially improve children's learning outcomes at a low cost ([↑Bergman, 2019](#)). For example, results from [↑Doss et al. \(2017\)](#) show that a low-cost, personalised literacy texting intervention to parents can have a substantial effect on student academic outcomes. Further, evidence from a school-based intervention in India shows that persuasive messaging around gender can positively address gender norms among parents and children ([↑Dhar et al., 2018](#)).

Additionally, important evidence gaps exist when it comes to best practices in engaging caregivers through mobile messaging. Unanswered questions include the optimal length of exposure (i.e., how long messages need to be sent for to create long-term behaviour change), and the optimal degree of customisation, particularly to tackle caregivers' asymmetric beliefs for boys vs girls (i.e., what messages are most effective to change underlying beliefs about returns to girls' education, leading to long-term behaviour change). A further evidence gap relates to differences by child age in the effectiveness of mobile messaging interventions around caregiver engagement in child learning.

3. Methodology

Our design is a household-randomised control trial. Randomisation was stratified within each of the five regions. Households were assigned to receive either SMS nudges or no SMS nudges. Households in the comparison group did not receive any messages during the study period. Households were randomly assigned to one of five experimental groups:

1. Behavioural nudges (n = 513 households / primary caregivers): Primary caregivers received messages encouraging involvement with children's learning, their child's social-emotional development, academic aspirations, and ensuring children return to school after schools have reopened (12 weeks — 24 SMS).
2. Behavioural nudges with 'gender-parity nudges' (n = 527 households / primary caregivers): Primary caregivers received messages, in which the content built on that in the standard message treatment with nudges increasing the salience of girls and including content promoting girls' education. Some messages addressed some common stereotypes around gender roles (12 weeks — 24 SMS).
3. Behavioural nudges of longer duration (n = 540 households / primary caregivers): Primary caregivers received the same messages as Group 1 but the programme's duration was 24 weeks — into the second term of the 2021 academic year (24 weeks — 48 SMS).
4. Behavioural nudges with 'gender-parity nudges' of longer duration (n = 518 households / primary caregivers): Primary caregivers of both boys and girls received messages to parents, in which some of the nudges included content promoting girls' education and addressing some common stereotypes around gender roles during the school closures (24 weeks, into the second term of the 2021 academic year) (24 weeks — 48 SMS).
5. Comparison group (n = 530 households / primary caregivers): No messages during the study period.

3.1 Research questions

We test the following research questions:

1. Do nudges to parents in the form of SMS messages increase the rate of children returning to school and general engagement with education when schools reopen?

2. Do nudges change caregiver beliefs about returns to education and expectations and aspirations?
3. Do nudges improve children's learning and schooling outcomes (i.e., enrollment, attendance) in the medium-term?
4. Are impacts more equitable across girls and boys if messages focus on gender parity in education and in behaviours / attitudes towards girls?
5. Are impacts greater and do they persist for longer if exposure is longer (24 versus 12 weeks)?

3.2 Research framework / methodology

We conducted four main data collection activities for the study, namely, participants enrollment and caregiver survey (baseline), mini-survey, midline survey, and endline survey. Apart from the participant enrollment and caregiver survey, which was administered via phone, the midline and endline surveys were administered in person. All surveys or assessments were administered by trained enumerators. See timeline in [Figure 1](#).

We conducted the participants' enrollment and caregiver survey to screen households for eligibility (having a child of the target age group of 5–17 years), seek their consent, communicate project information, enrol them in the study, and obtain pertinent background information on the primary caregiver as detailed in Table 1 below.

Table 1. *Background information on the primary caregiver collected at baseline.*

Background information on the family of the primary giver
Family demographics
Number of school-age children in household
Engagement in remote learning
Caregiver pro-boy bias
Aspirations for education of boys and girls in household
Returns to education for boys and girls in household

The participants' enrollment and caregiver survey was conducted via phone. The sample came from two existing samples (Communication 4 Development ([↑Fink et al., 2019](#)) and Ghana Panel Survey ([↑ECG & ISSER, n.d.](#))). However, due to issues with some of the mobile phone numbers (i.e., inactive phone numbers,

wrong numbers, phone numbers switched off, and connectivity issues in the catchment areas), we conducted in-person tracking of potential households in the study areas to update or identify over 150 new mobile phone numbers. Upon calling the potential respondents, trained enumerators screened them in line with the eligibility criteria. Once the respondent met the eligibility criteria, we sought verbal consent of the primary caregiver to

- a. participate in the study
- b. allow the selected child(ren) to participate in the study
- c. inform them about the project, the intervention, and implementation modalities
- d. participate in the intervention (only those in the treatment group).

Both eligible primary caregivers and school-going children were enrolled during the participant enrollment and caregiver survey. However, no child assessment was conducted at this stage.

The midline survey was conducted to collect two main sets of information. First, to

- update the household and personal information of the primary caregiver and update the status of the eligible children selected at baseline;
- measure changes in the primary caregivers' behaviour and attitudes as well as participation in the intervention.

Second, to

- conduct child direct assessment with eligible children on literacy, numeracy, socio-economic and other useful domains of child development.

The midline survey was conducted in person at the homes of the study participants. In each study household, we first interviewed the primary caregiver to obtain child-specific information on each selected child before assessing the eligible child(ren). Interviewing the caregiver first was grounded in two reasons:

1. to confirm the availability of the eligible selected child;
2. to obtain caregiver-reported measures on each selected child in the caregiver survey.

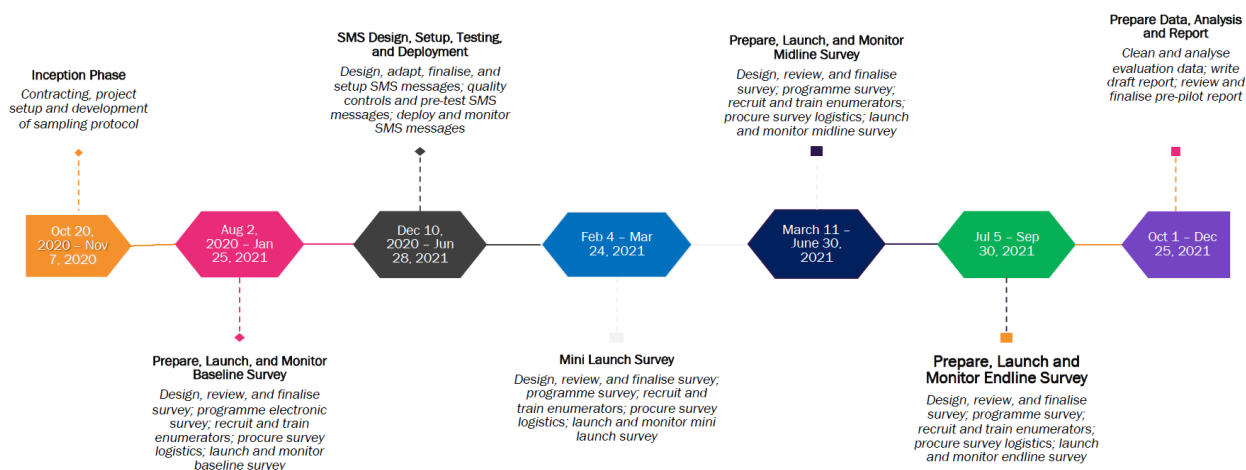
The process was automated in SurveyCTO such that once the caregiver survey was completed and data synced to the SurveyCTO server, it published the eligible child's form to enable the assessment to be conducted on the particular child.

To sample children, we used the household roster and randomly selected one child from the 5–9-year-old range, and one child from the 10–17-year-old range. This allows us to examine direct impact on children of different ages in the household.

The endline survey collected such information as household and personal information, the status of selected children, behaviour, and attitudes as well as participation in the intervention from primary caregivers. We also conducted child direct assessment with eligible children on literacy, numeracy, socio-economic, and other useful domains of child development. At endline, primary caregivers and children were interviewed in person based on their availability.

Of the 2,628 interviewed at baseline, only 41 (1.56%) were not interviewed at midline. An additional 47 households were not interviewed at endline (88 total; 3.35% of the baseline sample).

Figure 1. *Study timeline.*



When the research team arrived at each community, they first went to meet the chief or the community leader to conduct community entry protocols. The community entry is undertaken by paying a fee as a customary practice to inform the chief and his kinsmen about the presence of the team. They introduced themselves briefly and also showed their IPA ID cards and explained the activities to be carried out, the purpose, the mode of operation, and how many days they would be in the community. The chief then assigned someone to the enumerators (if possible — not all the time) to assist in

locating the households concerned. Upon reaching a household, the enumerator spoke with the household head and explained the purpose of the visit and what the entire process would entail, while the household head assisted in identifying the primary caregiver. During interviews, both enumerators and respondents sat outside where they could be visible to everyone though in a quiet space devoid of external interruption. During the rainy season, where respondents could not provide spaces like verandahs or corridors for interviews, the process was automatically rescheduled. During meetings with all those involved, Covid-19 protocols were strictly followed. All participants provided consent in order to participate in the surveys and assessments.

3.3 Research instruments / tools

Our trial and pre-analysis plan were registered in the American Economic Association's Social Science Registry.¹ Two main sets of outcome measures — primary and secondary — were used to assess the impacts of the intervention as specified in our pre-analysis plan. These outcome measures were collected using the same procedures for the intervention and the comparison groups.

3.3.1 Primary outcomes

We had five primary outcomes:

1. Caregiver engagement in education
2. Caregiver expectations and aspirations for their child's schooling
3. Caregiver expectations on returns on education
4. Caregiver beliefs about gender norms
5. Children's school enrollment and attendance (reported by both caregivers and children).

Caregiver engagement in education

At both midline and endline, caregivers reported on whether they engaged in a set of activities related to their child's education over the past three days, specific to their engagement with each of the two focal children sampled. Six activities were summed to create an index of the number of activities parents engaged in. Activities were slightly different for older and younger children to ensure developmental appropriateness. For home engagement for younger children (ages 5–9), we used the Multiple Indicators Cluster Survey (MICS)

¹ <https://www.socialscienceregistry.org/trials/6118>

home stimulation index on six activities which included reading or looking at books, telling stories, singing songs, taking the child outside the home, playing with the child, and naming / counting / drawing with the child. For older children (ages 10–17), we adapted these items based on the Young Lives surveys ([Barnett et al., 2013](#)). Activities included working on a project together, playing sports / active games / exercise, discussing time management, talking about family / community history / heritage, discussing future education and career plans, and encouraging the child to listen to or watch remote teaching. For school engagement, we used a series of seven dummy variables that were the same across the age groups. These included whether in the last month, the caregivers

1. helped the child with homework
2. asked the child what they did at school.

Whether in the last academic year they

3. attended a PTA meeting
4. attended a scheduled meeting with the child's teacher
5. attended a school or class event
6. volunteered or served on a school committee
7. participated in fundraising for the child's school.

We also measured child-reported caregiver home engagement on their education, which is measured in the same way as the caregiver-reported measure. As the two scales are moderately correlated ($r=0.40$, $p<0.01$), we analysed treatment effects on both scales.

Caregiver expectations and aspirations for their child's schooling

At midline and endline, caregivers reported on their (i) educational aspirations and (ii) expectations for two focal children in the household: 'What is the highest level of education that you WISH [child] to achieve?', and 'What is the highest level of education that you EXPECT [child] to achieve?'. Overall, most caregivers had high aspirations for their child, with close to 80% aspiring for their child to achieve a post-secondary degree. We dichotomise both variables to indicate 'low aspirations' = 1 and 'low expectations' = 1 if the caregiver aspires for their child to reach a level that is below secondary high school or expects that the child will reach a level below secondary high school).

Caregiver expectations on returns on education

At endline only, caregivers reported on their perceived returns to education using a series of four questions (adapted from [Attanasio & Kaufmann, 2014](#)). We asked six questions about perceived returns to three levels of education

(two each): Junior high school, secondary high school, and university: 1. Assume that [CHILD] completes [JHS / SHS / a university degree] and that this is his or her highest schooling degree as he or she stops going to school afterwards. From zero to one hundred, how certain are you that [CHILD] will be working at the age of 30? Assume that [CHILD] completes [JHS/SHS/a university degree] and that this is his or her highest schooling degree. Assume that [CHILD] has a job at age 30. Think about the kinds of jobs he or she might be offered and that he or she might accept. How much do you think [CHILD] will earn in a typical MONTH when he or she is about 30 years old?

Caregiver beliefs about gender norms

Gender bias was measured through two scales at baseline, midline, and endline. First, parents answered a single item: “Do you think education has a greater influence on your son’s income than your daughter’s income? (Yes/No). Second, parents answered 14 items from the “Gender norms and attitudes scale” ([Waszak et al., 2001](#)), which measures egalitarian beliefs about male and female gender norms. Specifically, the scale assesses whether parents agree or disagree with a series of statements about the promotion of equity for girls and women and maintaining the rights and privileges of men (14 items / 2 subscales). Example items include: “It is important that sons have more education than daughters”, “Daughters should be sent to school only if they are not needed to help at home”, and “Daughters should have just the same chance to work outside the homes as sons.”

Children’s school enrollment and attendance (reported by both caregivers and children)

School attendance and enrollment were assessed at midline and endline through parent and child reports, where they answered: “How many days of school has [NAME] missed in the past week of school (Monday–Friday)?” We dichotomise this variable to indicate ‘low attendance’, if the child never attends or attends infrequently (2 or 3 times per week), which corresponds to around 10% of the sample. We use child- and caregiver-reported scales for both school enrollment and attendance, as correlations between child- and caregiver responses are moderate or low, respectively (school enrollment, $r=0.65$, $p<0.01$; low attendance: 0.08 , $p<0.01$).

3.3.2 Secondary outcomes

Our secondary outcomes are divided into those we hypothesise will be mediators to the primary outcomes and those that we consider distal outcomes to the treatment.

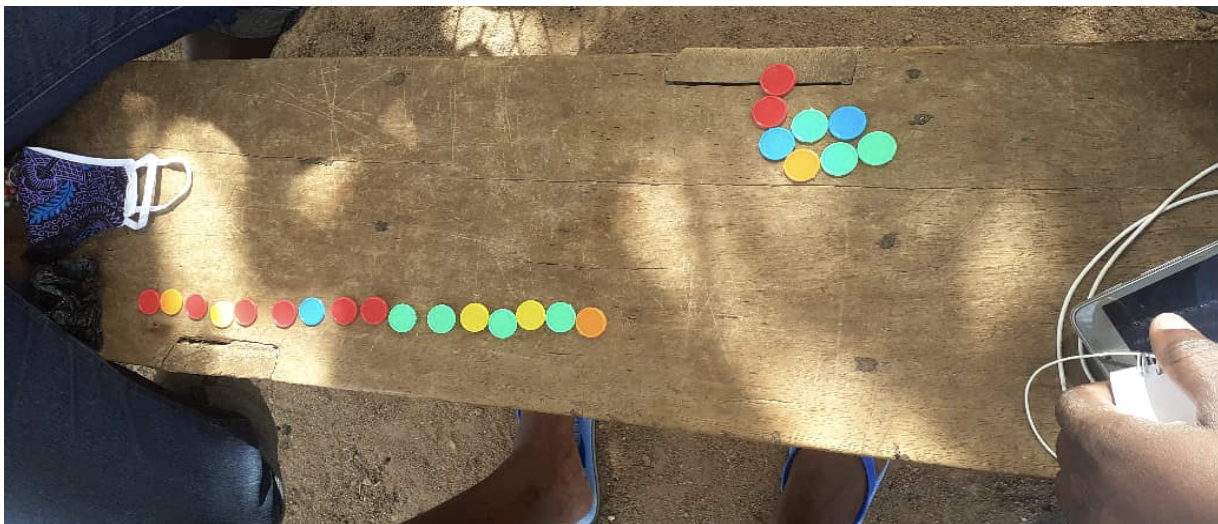
Mediators include

1. Child time use
2. Caregiver self-efficacy
3. Disciplinary practices
4. Emotional supportiveness in the home.

Child time use

Older children (10–17 years) were asked to report on the following: “On a typical weekday from Monday to Friday (not a weekend or a holiday, or a day in which a child was sick), how many hours did you spend on the following activities last week?” Children reported on the number of hours they spent on eight activities: sleep, caring for others (e.g., younger siblings or the elderly), household chores, working on the farm or other family business, working for pay, in school, studying, and engaging in leisure (e.g., playing). These items were adapted from Young Lives ([↑Barnett et al., 2013](#)). Following Young Lives, to measure these outcomes, children were given 24 counters and, using cardboard with 8 circles representing the above-listed 8 categories of activities. Fieldworkers asked the children to distribute the 24 counters according to the time spent in each of the eight tasks, as shown in Figure 2.

Figure 2. *Counters used during child assessment. Source: Fieldworker, May 2021.*



Caregiver self-efficacy

At both midline and endline, caregiver self-efficacy was measured using one scale from Bandura's Parental Self-Efficacy Scale ([↑Bandura et al., 2001](#)),

specifically the eight-item subscale related to parental self-efficacy regarding their children's schooling and learning. The scale was scored from 1–5, with 1 = nothing, and 5 = a great deal. Example items include: “How much can you do to make your children see school as valuable?” and “How much can you do to help your children get good grades in school?”

Disciplinary practices

Disciplinary practices used by caregivers were self-reported for each focal child separately using the UNICEF MICS scale ([UNICEF, 2010](#)), which asks caregivers if they have used a series of disciplinary practices in the previous month. These include three subscales that were combined in two: non-violent practices (three items) and violent practices (psychological aggression and physical violence, eight items). Caregivers were asked questions such as “did you explain why [child] behaviour was wrong in the past month?” “did you hit or slap [child] on the face, head, or ears in the past month”? “call [child] dumb, lazy, or another name in the past month?”. Questions were adapted to the different child age groups.

Emotional supportiveness in the home

Caregiver emotional supportiveness was assessed at midline and endline using a 5-item scale from the Early Childhood Longitudinal Study — Kindergarten Cohort ([Chapman, 2010](#)).² Parents answered a variety of hypothetical questions specific to supporting their children's emotional needs and answered on a 4-point scale (1 = very often true, 2 = often true, 3 = sometimes true, and 4 = never true. Example items include: “Even if I am really busy, I make time to listen to [child],” and “I encourage [child] to talk about his/her troubles.” Items were adapted to the specific child age groups.

Distal outcomes include

1. Children's academic outcomes (literacy and numeracy)
2. Children's social-emotional skills
3. Caregiver mental health.

Children's academic skills

We measured *literacy* and *numeracy* skills for children. We covered similar subskills for all age groups (e.g., reading comprehension, oral vocabulary for reading; addition and subtraction for maths). Given the wide age range in our

² This study covers children from early childhood to eighth grade.

sample, we had three assessments to cover different skill levels (i.e., an assessment for 5–9-year-olds, an assessment for 10–14-year-olds, and an assessment for 15–17-year-olds). Using tasks from International Development and Early Learning Assessment (IDELA), Early Grade Reading Assessment (EGRA), and the Young Lives surveys, we measured the following domains of skills for literacy: expressive language (oral vocabulary), non-word reading, spelling, oral reading and comprehension, and phonological awareness. Using tasks from the IDELA, Early Grade Math Assessment (EGMA) and the Young Lives surveys, we measured the following skills domains for numeracy: number identification, number / quantity discrimination, missing number patterns, number sorting, word problems, and operations (addition / subtraction, and multiplication and division). To create a summary score, the total correct number of answers on each subtask was calculated within each of the following three age groups: 5–9 years, 10–14 years, 15–17 years. This was because the number and difficulty of questions varied based on child age. Then, we summed the total number of correct answers within each of the three age groups, and finally age- and round-standardised overall scores for literacy and numeracy ($M = 0$, $SD = 1$). Thus, our scores represent children's performance relative to the peers in their age group.

Children's social-emotional skills

Social-emotional skills were only measured for older children (10–17 years). First, we measured *self-esteem* (using the Rosenberg Self-Esteem Scale; [↑Rosenberg, 1965](#)). This is measured by assessing the level of agreement of the child to ten items. Examples are: “On the whole, I am satisfied with myself”, or “At times, I think I am not good at all.” Higher levels measure higher self-esteem. Second, we measured *school motivation* using the Elementary School Motivation Scale ([↑Guay et al., 2005](#)). Children are asked how much they identify themselves with nine statements such as “I like to go to school” or “In life, it's important to go to school.”

Caregiver mental health

Caregiver mental health was measured at endline only using the Kessler Psychological Distress Scale ([↑Kessler et al., 2002](#)), a 10-item questionnaire used globally to measure general psychological distress based on questions about anxiety and depressive symptoms. Each item is scored from zero (none of the time) to four (all of the time). Items are added to create a total score, with higher scores indicating higher psychological distress and a higher likelihood of a mental health disorder.

3.4 Research design

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

For each caregiver outcome, we estimated the following ordinary least squares regressions indexed by caregiver 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 caregiver p in household h and survey rounds.
- $NShort$, $GShort$, $NLong$, 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 caregiver assessment (the one conducted at the end of the 12-week implementation of Treatments 1 and 2), we pooled samples from Arms 1 and 3, and 2 and 4 to estimate effects. For the endline, T_p included four treatment dummies to treatment arms 1, 2, 3, and 4 separately.
- $Y_{p,h,0}$ is the baseline outcome variable for caregiver p in household h (when available)
- $X_{h,0}$ is a vector of caregiver and household controls should there be a lack of balance in the randomisation.
- G_h is an indicator variable assuming the value of 1 for households belonging to the Ghana Panel Study sample, 0 otherwise;
- θ_r are region fixed effects
- $\varepsilon_{h,s}$ is an individual error term

For each child outcome, we estimated the following ordinary least squares regressions indexed by child c , living in household h and survey s :

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

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 caregivers' outcomes, for the first child assessment (the one conducted at the end of the 12-week implementation of Treatments 1 and 2), we pooled samples from Arms 1 and 3, and 2 and 4 to estimate effects. For the endline, T_p included four treatment dummies to treatment arms 1, 2, 3, and 4 separately.
- $X_{h,0}$ is a vector of caregiver and household controls should there be a lack of balance in the randomisation.
- G_h is an indicator variable assuming the value of 1 for households belonging to the GUP sample, 0 otherwise;
- θ_r are region fixed effects
- $\varepsilon_{c,h,s}$ is an individual error term, clustered at the household level.

3.5 Stakeholders

We have multiple stakeholders in Ghana, including the Ministry of Education (MoE), Ghana Education Services (GES), the World Bank, and a number of non-governmental organisations that work on education and gender in the northern regions (e.g., World Education Inc., NORSACC, and ActionAid). The SMS programme was based on Movva's EDUQ+ programme, which has been implemented and evaluated in Brazil and Cote d'Ivoire. We worked with the group of stakeholders to adapt the intervention to the Ghanaian and Covid-19 pandemic context, in collaboration with Movva. The programme shares weekly suggestions of activities for caregivers to do with their children — none of them linked to curricular activities; rather, ones that aim to bring caregivers 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 behavioural barriers to engaged parenting (York et al., 2019).

Figure 3. Sample messages.



Note: The first sequence portrays a message from the general programme; the second sequence portrays the same sequence adapted to the gender boost programme.

The idea of conducting the proposed research emerged from conversations with employees of the MoE/GES, which highlighted low levels of caregiver engagement as a key barrier to achieving equitable learning outcomes in Ghana. This issue was exacerbated by the Covid-19-induced school closures. The study therefore provides an opportunity to highlight the role of caregivers as key stakeholders in education and to better understand how best to engage caregivers. Caregiver engagement is also a key component of the Ghana Accountability for Learning Outcomes Project (GALOP), and the Ministry of Education team has expressed interest in testing different approaches for engaging caregivers as part of the GALOP research.

3.6 Ethical considerations

The study received ethical approval by the Institutional Review Board of Innovations for Poverty Action. A protocol was submitted and reviewed by the ethics review board. In-depth training for interviewing and the ethics of working with children was administered prior to starting data collection. Children were interviewed in quiet places and there was a strong emphasis on making sure that the assessments were not 'high-stakes' exams and that the information provided is confidential.

Notably, Ghana's government only requires ethical review / approval in health research studies, as per the directive from the Ghana Health Services.³ This study is not considered health research.

Before enrolling participants into the study, participants were screened, and their eligibility determined. Households had to have at least one school-aged child in the home. They were then administered the informed consent and their consent was sought to enroll them into the study and enroll their children into the study. We also sought their consent to participate in the SMS messages intervention. Only those who consented to participate in the study, allowing their children to participate in the study and participating in the SMS intervention were enrolled into the study. Consent was obtained from the participants verbally and recorded directly on the SurveyCTO form on a Samsung tablet. Caregivers who completed the surveys were given GH¢ 5 worth of airtime while their children were given Note 1 exercise books and pen / pencil.

Data collectors for both the caregiver survey and child outcome measures had prior experience working with children, were trained extensively in study protocols and research methodology, and spoke the local languages of the communities in the study. All assessments were conducted in the local languages of the participants (i.e., Dagbani, Buli, Dagaare, Gruni, Wali, Mampruli, Sisaasla). When data collectors arrived at a village, they first went to meet the chief or the community leader to conduct community entry protocols. The community entry was undertaken by paying a fee as a customary practice to inform the chief and his kinsmen about the presence of the team. The data collectors introduced themselves briefly and also showed their IPA identification cards. Then, the team explained the activities to be carried out to the chief, the purpose, the mode of operation, and how many days they would be in the community. If possible, the chief then assigned someone from the community to support the enumerators in locating the sampled households. Upon reaching a household, the enumerator identified

³ See <https://webapps.sph.harvard.edu/live/gremap/fwa.cfm?id=17> for details.

the household heads and explained the purpose of the visit and what the entire process would entail. The household head in turn assisted in identifying the child's primary caregiver. During the interviews, both enumerators and respondents sat outside where they could be visible to everyone, though in a quiet space devoid of external interruptions. During the rainy season, when respondents could not provide spaces like verandahs or corridors for the start or continuations of interviews, the process was automatically rescheduled. All Covid-19 protocols were strictly followed.

3.7 Challenges

3.7.1 Data collection challenges

We encountered the following challenges during the implementation of the data collection activities.

Contractual delays with the main funder

Our initial focus was to support caregiver engagement in remote learning as schools were closed. Unfortunately, we experienced substantial contractual delays that significantly delayed our start. We decided to start implementation in January 2021 to encourage caregivers to send their children back to schools while we were still finalising the contracts. Schools reopened on January 18, 2021. Thus, our focus shifted from supporting caregiver educational engagement during school closures to supporting caregivers in their engagement as schools re-opened.

Challenges in surveying participants by phone

In May 2020, we planned to conduct telephone surveys to collect information on caregiver and child outcome measures as in-person data collection was not feasible due to the peak of the Covid-19 pandemic in Ghana. However, the participant enrollment and caregiver baseline survey in December 2020 with the study participants, as well as lessons from conducting child phone surveys in the [Quality Preschool for Ghana](#)⁴ study (↑[Wolf, Aurino, Suntheimer et al., 2021](#)) have highlighted several challenges with telephone surveys, with implications for data quality. First, many primary caregivers had issues trusting the intentions of the research project, given that we had not been able to establish in-person contact with them before, and all contact was via the phone. Second, there were substantial network challenges that led to the need to conduct in-person tracking for hard-to-reach respondents located in many communities. Finally, using phone-based assessments is not the

⁴ See <https://www.poverty-action.org/study/improving-kindergarten-quality-ghana>
Nudges To Improve Learning and Gender Parity

optimal way to measure children's learning outcomes: in-person assessments provide far superior data. Flowing from the above, we shifted follow-up surveys in April 2021 and August 2021 to in-person surveys (instead of phone surveys) with strict adherence to Covid-19 protocols. By implementing in-person follow-up surveys, we were able to reach study participants more easily and with the assurance of higher-quality data. We conducted in-person follow-up surveys following IPA's policy on restarting face-to-face data collection and the Ghana Health Service Ethical Review Committee's protocol for conducting in-person data collection during the Covid-19 period. This included

1. Providing each field staff with sanitiser and face masks;
2. Providing each field team with a thermometer gun for checking the daily temperature of team members before and after fieldwork;
3. Practising good personal hygiene through regular hand washing, use of the sanitisers, and wearing face masks;
4. Adhering to the social distancing protocol by observing a distance of at least six feet between the study participant and the field staff.

Non-availability of resources for proper setup for conducting child direct assessments at home

Optimal administration of child assessments requires a setup with a table and two chairs — one each for the assessor and the child — to enable the assessor and the child to effectively use the assessment materials including stimulus cards in conducting the assessment. However, the lack of such resources in most households made it quite difficult to effectively conduct the assessments. In some cases, children were assessed while sitting on stones, the ground, and on trees / wood. This did not provide conducive environments for children to fully participate in the assessments and thus has implications for data quality. Given our experience with previous child direct assessment and our understanding of the study contexts, we trained our data collectors to be able to improvise in such circumstances in a way that reduced the impact of the learning space on the children and data quality, as shown in [Figure 4](#).

Figure 4. *Challenges in properly setting up child assessment. Source: Fieldworker, May 2021.*



Tracking of study participants

We designed the evaluation as a household-level intervention involving home-based data collection activities for both primary caregivers and school-going children. However, several natural and human factors affected the tracking of study participants at the endline and the planned data collection schedules. First, primary caregivers (and to some extent children) were often not available due to their participation in farming or business activities. Second, children were recruited across primary and secondary schools with varying school calendars making it difficult to identify a common period for tracking and assessing children. Third, some primary caregivers and children migrated from the study regions to other study regions or the southern part of Ghana to engage in diverse economic activities. Finally, data collection activities coincided with the rainy season, making it difficult to access communities for data collection. These factors affected our tracking efforts and data collection schedules. To reduce the impact of these challenges, reduce attrition, and ensure data quality, we implemented the following measures:

1. Interviewing study participants outside their homes including their farms, business venues, or community centres;
2. Staggering data collection activities depending on study participants — children were mostly interviewed after school hours (late afternoons) — while targeting non-available or busy study participants on weekends for data collection;

3. (Re)scheduling appointments with primary caregivers ahead of a home visit to avoid missing the study participants and ensuring that they were available for interviews;
4. Postponing data collection activities in inaccessible areas for the safety of data collectors.

Ramadan and harvesting season

The midline survey coincided with the Ramadan period, and some participants were not comfortable with participating in the data collection activity during the fasting period. Relatedly, primary caregivers engaged in harvesting activities were identified, and together with those involved in the Ramadan activity, data collection activities were adjusted to align with a more convenient time for participants, including during weekends.

3.7.2 Programme implementation challenges

The main challenges associated with the implementation of the intervention were as follows.

Difficulty in procuring a dedicated shortcode for deploying text messages

Delays in procuring a dedicated shortcode and configuring the different networks into the shortcode were one major cause of the delay in implementing the text messages intervention. To address this, the team relied on a shared shortcode from another project to kickstart the text message intervention.

Network connectivity issues

Due to the geographical location of the primary caregivers and network issues, some of the treatment households did not receive any of the text messages sent to them. SMS records from Mova show that the SMS messages were never delivered to 38 primary caregivers due to wrong numbers or non-working mobile phone numbers. One strategy used was to inform the primary caregivers of the time of day the messages would arrive so that they could be in a location that allowed them to receive the text messages.

4. Results

Our analytic approach is driven by the randomised design of the evaluation. We discuss the data and limitations with this approach, then present the impact evaluation findings for our primary and secondary outcomes.

4.1 Data

On average, caregivers were slightly above 40 years of age, and households were large — with around 10 members. On average, three household members were school-age children. The majority (60%) of the caregivers were female. Of the child sample, close to half (47%) were girls and the average age was 10 years. Based on baseline equivalency analysis (see [↑Wolf et al., 2021](#)), we conclude that the randomisation was successful in generating five equivalent groups.

The remainder of this section will outline the main limitations of this study, and later, the main results. We will start by documenting treatment effects on primary outcomes and then heterogeneity by caregiver characteristics. We will focus on caregiver schooling as the main axis of heterogeneity.

4.2 Limitations

As with any longitudinal research design, sample loss to follow-up can be a critical limitation to the internal validity of study findings. Fortunately, between baseline, midline, and endline, attrition was very low. Of the 2,628 interviewed at baseline, only 41 (1.56%) were not interviewed at midline. An additional 47 households were not interviewed at endline (88 total; 3.35% of the baseline sample). Attrition analysis for the endline sample highlights that participants withdrawing from the study were more likely to be from the Upper East and Upper West regions, more likely to be from the 24-weeks treatment groups, more likely to be a male caregiver who participated in the study, and from households with fewer children. These effects were small, ranging from 1–4 percentage points, so they should not substantially threaten the internal validity of our research design ([↑Wolf et al., 2021](#)).

Further, school enrollment and attendance data were only reported by children and caregivers. We are waiting on permission from the Ghana Education Service to collect school enrollment and attendance data from school records. This process is taking much longer than anticipated and has significantly delayed this data collection. Future analysis will investigate

correlation between survey-based data and administrative data, as well as the programme's treatment effects on administrative data.

Finally, due to the state of the Covid-19 pandemic at the start of the study, we were unable to collect baseline assessments of children and only collected a brief caregiver survey over the phone. Thus, we do not have baseline assessments of our outcomes. Notably, though, baseline equivalence was established across treatment arms, suggesting that comparisons at midline and endline between treatment and control groups are internally valid.

4.3 Results

4.3.1. Main treatment effects for primary outcomes

We use intent-to-treat estimates in our analysis. Our treatment impacts are estimated based on parent and child reports. Given small to moderate correlations between parent and child reports of the same outcomes (e.g., school enrollment), ranging from 0.10–0.55, we interpret each source separately. We present the main impacts on outcomes at both midline and endline in [Table 2](#) (Panels A and B, respectively). Results compare mean scores for treatment and control groups at each time point. We consider results as statistically significant if the p-value for the treatment coefficient is below 0.05. First, we found no effects of the intervention on changing home and school engagement, as reported by both caregivers and children.

Table 2. *Treatment impacts on primary outcomes at midline and endline.*

Variables	Caregiver: home engagement	Child: home engagement	Caregiver: school engagement	Caregiver: child enrollment	Child: enrollment	Caregiver: low attendance	Child: low attendance	Caregiver: pro-boy bias	Caregiver: aspirations- expectations gap	Caregiver: returns to JHS	Caregiver: returns to SHS	Caregiver: returns to Univ.
Panel A: Midline												
12-week, behavioural nudges	-0.145	0.048	-0.139	0.001	-0.003	0.013	-0.014	-0.019	-0.109*			
	[0.1053]	[0.0946]	[0.1145]	[0.0127]	[0.0147]	[0.0151]	[0.0182]	[0.0620]	[0.0617]			
12-week, gender boost	-0.155	0.047	-0.025	-0.033**	-0.029*	0.042***	-0.021	-0.019	-0.079			
	[0.1065]	[0.0955]	[0.1165]	[0.0137]	[0.0151]	[0.0156]	[0.0178]	[0.0627]	[0.0697]			
24-week, behavioural nudges	-0.171*	-0.031	-0.112	-0.009	-0.023	0.032**	-0.007	0.012	-0.030			
	[0.1030]	[0.0931]	[0.1152]	[0.0132]	[0.0154]	[0.0157]	[0.0182]	[0.0615]	[0.0667]			
24-week, gender boost	-0.167	-0.042	-0.057	0.008	0.006	0.017	-0.017	-0.097	0.015			
	[0.1046]	[0.0930]	[0.1139]	[0.0126]	[0.0143]	[0.0152]	[0.0179]	[0.0614]	[0.0693]			
Observations	4,671	4,673	4,673	4,673	4,670	4,673	4,178	4,673	4,166			
R-squared	0.0552	0.0808	0.0362	0.0734	0.1104	0.0916	0.0045	0.0622	0.1042			
Mean outcome var	3.037	2.202	3.204	0.926	0.897	0.108	0.123	0.0127	0.278			
Panel B: Endline												
12-week, behavioural nudges	-0.118	-0.198*	0.007	-0.001	-0.013	0.014	0.003	0.045	-0.132	0.010	-0.083	-0.123
	[0.1054]	[0.1052]	[0.1007]	[0.0131]	[0.0153]	[0.0158]	[0.0210]	[0.0632]	[0.0928]	[0.1023]	[0.0928]	[0.0818]

12-week, gender boost	-0.061	-0.035	-0.041	-0.042***	-0.020	0.048***	0.014	0.133**	-0.206**	0.117	0.064	-0.084
	[0.1055]	[0.1074]	[0.1028]	[0.0143]	[0.0152]	[0.0164]	[0.0212]	[0.0627]	[0.0949]	[0.0888]	[0.0853]	[0.0810]
24-week, behavioural nudges	-0.048	-0.036	-0.120	-0.012	-0.023	0.020	0.016	-0.023	-0.148*	0.048	-0.012	-0.081
	[0.1041]	[0.1065]	[0.0980]	[0.0136]	[0.0154]	[0.0158]	[0.0213]	[0.0624]	[0.0901]	[0.0891]	[0.0818]	[0.0770]
24-week, gender boost	-0.021	-0.040	0.084	0.001	0.007	0.015	-0.002	0.050	-0.295***	0.076	-0.022	-0.071
	[0.1050]	[0.1050]	[0.1009]	[0.0132]	[0.0147]	[0.0157]	[0.0209]	[0.0630]	[0.0882]	[0.0970]	[0.0887]	[0.0805]
Observations	4,721	4,738	4,738	4,721	4,627	4,721	4,128	4,721	4,422	2,627	3,592	4,447
R-squared	0.0553	0.0598	0.0485	0.0690	0.1057	0.0849	0.0344	0.0280	0.0867	0.0076	0.0136	0.0429
Mean outcome var	3.382	2.681	3.145	0.922	0.895	0.116	0.182	-0.0128	0.321	5.746	6.151	7.104
Notes: Caregiver = caregiver-report, child = child-report. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1												

Second, for child schooling outcomes, as reported by both caregivers and children, surprisingly, we found negative impacts of the 12-week gender boost on school enrollment and attendance at both midline and endline of similar magnitude, with a decrease of 3 percentage points (p.p.), in enrollment at midline (endline), and an increase in low attendance of 4 p.p. and 5 p.p. at midline and endline respectively. For the 24-week behavioural nudges treatment, we also found a negative effect on parent-reported child school attendance at midline, with a 4 p.p. increase in the likelihood of low school attendance, respectively.

Third, regarding gender bias, there were no impacts on pro-boy bias at midline. However, there were statistically significant increases of the 12-week gender-boost arm on pro-boy bias at endline when compared to pro-boy bias levels of the control group.

Fourth, we examined impacts on the gap between caregiver aspirations and expectations for their child's schooling outcomes. There were no statistically significant treatment impacts on the aspirations-expectations gap at midline. At endline, however, there were statistically significant impacts. The analysis of impacts on aspirations and expectations separately shows that expectations increased.

Finally, we examined caregivers' perceived returns to education at endline; that is, how much parents believe attaining a particular level of education will impact the future salaries of their children. There were no treatment impacts on perceived returns to education.

4.3.2. Subgroup treatment effects for primary outcomes

We examined heterogeneity of treatment effects for three different sets of subgroups: caregiver education (whether caregivers ever attended school (35.2% of the sample) or not); child age (5–9-year-olds versus 10–17-year-olds); and child gender. We found significant differences for each subgroup, suggesting that a full understanding of the effectiveness of the treatment needs to be understood within the context of various subgroups.

4.3.2.1. Caregiver schooling

Caregiver schooling was an important moderator of treatment impacts, with the general pattern of results suggesting negative impacts on non-educated caregivers and some positive impacts on educated caregivers ([Table 3](#)). More specifically, except for the short standard treatment, all other treatment conditions increased caregiver school

engagement for caregivers who have some schooling. These effects persisted at endline (although the level of statistical significance goes down to $p < 0.1$), and also appear on the caregiver home engagement scales for the short gender boost and the long standard treatment arms.

Table 3. *Impact heterogeneity by caregiver education.*

Variables	Caregiver: home engagement	Child: home engagement	Caregiver: school engagement	Caregiver: child enrollment	Child: enrollment	Caregiver: low attendance	Child: low attendance	Caregiver: aspirations- expectations gap	Caregiver: returns to JHS	Caregiver: returns to SHS	Caregiver: returns to Univ.
Panel A: Midline											
12-week, behavioural nudges	-0.207	-0.031	-0.244*	-0.001	-0.012	0.023	-0.016	-0.084			
	[0.1297]	[0.1132]	[0.1469]	[0.0174]	[0.0206]	[0.0209]	[0.0226]	[0.0750]			
12-week, gender boost	-0.203	-0.052	-0.209	-0.038**	-0.045**	0.051**	-0.017	-0.036			
	[0.1290]	[0.1144]	[0.1506]	[0.0187]	[0.0212]	[0.0212]	[0.0222]	[0.0882]			
24-week, behavioural nudges	-0.133	-0.051	-0.294**	-0.017	-0.040*	0.050**	-0.003	0.038			
	[0.1255]	[0.1134]	[0.1473]	[0.0182]	[0.0213]	[0.0216]	[0.0228]	[0.0815]			
24-week, gender boost	-0.264**	-0.054	-0.228	0.008	0.009	0.009	-0.009	0.031			
	[0.1286]	[0.1100]	[0.1433]	[0.0173]	[0.0194]	[0.0202]	[0.0221]	[0.0797]			
1.attended_school	0.083	0.111	-0.038	0.018	0.025	-0.031*	0.007	0.035			
	[0.1551]	[0.1419]	[0.1639]	[0.0162]	[0.0178]	[0.0181]	[0.0273]	[0.0977]			
12-week, behavioural nudges #attended_school	0.174	0.221	0.297	0.006	0.023	-0.025	0.005	-0.070			
	[0.2217]	[0.2041]	[0.2310]	[0.0235]	[0.0262]	[0.0272]	[0.0381]	[0.1318]			
12-week, gender boost #attended_school	0.132	0.276	0.525**	0.014	0.045*	-0.022	-0.013	-0.117			
	[0.2271]	[0.2057]	[0.2316]	[0.0258]	[0.0271]	[0.0293]	[0.0372]	[0.1440]			

24-week, behavioural nudges #attended_school	-0.115 [0.2182]	0.055 [0.1983]	0.534** [0.2293]	0.023 [0.0240]	0.049* [0.0274]	-0.051* [0.0281]	-0.012 [0.0379]	-0.190 [0.1410]			
24-week, gender boost #attended_school	0.278 [0.2201]	0.031 [0.2034]	0.499** [0.2330]	-0.003 [0.0235]	-0.012 [0.0267]	0.026 [0.0293]	-0.023 [0.0378]	-0.045 [0.1534]			
Observations	4,671	4,673	4,673	4,673	4,670	4,673	4,178	4,166			
R-squared	0.0580	0.0848	0.0449	0.0759	0.1169	0.0978	0.0047	0.1053			
Mean outcome var low schooling	2.914	2.054	3.062	0.909	0.868	0.135	0.122	0.261			
Panel B: Endline											
12-week, behavioural nudges	-0.245* [0.1305]	-0.146 [0.1299]	-0.044 [0.1250]	-0.010 [0.0181]	-0.025 [0.0208]	0.033 [0.0219]	0.021 [0.0261]	-0.202* [0.1131]	0.092 [0.1185]	0.023 [0.1032]	-0.053 [0.0918]
12-week, gender boost	-0.244* [0.1323]	-0.085 [0.1304]	-0.180 [0.1296]	-0.055*** [0.0194]	-0.036* [0.0212]	0.065*** [0.0221]	0.014 [0.0265]	-0.270** [0.1170]	0.076 [0.1158]	0.042 [0.1068]	-0.074 [0.0987]
24-week, behavioural nudges	-0.242* [0.1292]	-0.123 [0.1270]	-0.290** [0.1204]	-0.032* [0.0191]	-0.038* [0.0209]	0.047** [0.0218]	0.024 [0.0260]	-0.241** [0.1105]	0.067 [0.1039]	0.015 [0.0931]	0.000 [0.0826]
24-week, gender boost	-0.127 [0.1310]	-0.080 [0.1288]	-0.035 [0.1251]	-0.002 [0.0179]	0.009 [0.0197]	0.009 [0.0206]	0.018 [0.0261]	-0.362*** [0.1049]	0.079 [0.1161]	-0.014 [0.1058]	-0.012 [0.0929]
1.attended_school	-0.173 [0.1594]	0.165 [0.1639]	-0.084 [0.1464]	0.009 [0.0172]	0.011 [0.0198]	-0.021 [0.0197]	0.009 [0.0320]	-0.121 [0.1416]	-0.069 [0.1297]	-0.069 [0.1177]	0.024 [0.1135]

12-week, behavioural nudges # attended school	0.373*	-0.159	0.147	0.026	0.034	-0.055*	-0.048	0.202	-0.242	-0.304	-0.203
	[0.2207]	[0.2220]	[0.2110]	[0.0240]	[0.0286]	[0.0284]	[0.0439]	[0.1979]	[0.2250]	[0.2073]	[0.1841]
12-week, gender boost # attended school	0.527**	0.134	0.396*	0.036	0.045	-0.049	-0.002	0.186	0.114	0.059	-0.033
	[0.2180]	[0.2289]	[0.2116]	[0.0273]	[0.0279]	[0.0311]	[0.0443]	[0.1999]	[0.1765]	[0.1761]	[0.1713]
24-week, behavioural nudges # attended school	0.573***	0.253	0.500**	0.058**	0.045	-0.076***	-0.025	0.266	-0.059	-0.076	-0.237
	[0.2162]	[0.2300]	[0.2053]	[0.0240]	[0.0289]	[0.0282]	[0.0451]	[0.1918]	[0.1972]	[0.1811]	[0.1766]
24-week, gender boost # attended school	0.314	0.114	0.348*	0.009	-0.005	0.019	-0.054	0.195	-0.014	-0.022	-0.168
	[0.2189]	[0.2219]	[0.2108]	[0.0250]	[0.0280]	[0.0309]	[0.0435]	[0.1924]	[0.2113]	[0.1912]	[0.1772]
Observations	4,721	4,738	4,738	4,721	4,627	4,721	4,128	4,422	2,627	3,592	4,447
R-squared	0.0599	0.0638	0.0534	0.0738	0.1096	0.0937	0.0356	0.0877	0.0115	0.0180	0.0454
Mean outcome var	3.269	2.527	3.021	0.902	0.869	0.145	0.180	0.279	5.784	6.209	7.160
Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1											

By contrast, for caregivers who never attended school, treatment effects on school engagement were negative at both midline and endline. On the other hand, only for those caregivers that never attended school, the two gender boost arms positively shifted educational expectations at endline.

4.3.2.2. Child gender

We also found interesting differences in results based on child gender, as shown in [Table 4](#). We found impact heterogeneity by gender in how the intervention impacted both aspirations and returns to education. For aspirations, we found that caregivers' expectations for girls' education improve at both midline and endline for the 24-week gender boost arm.

The estimate of a model with a three-way interaction between caregiver's schooling and child gender (results available upon request) highlights that the positive effects of the nudges on aspirations for girls are concentrated among caregivers with no education. Thus, the programme seems to be tackling aspiration failure for the poorest girls.

Impacts on returns to education were also moderated by child gender, as they increase for boys and decrease for girls, although for the latter, coefficients are not significant.

At midline, based on caregivers' reports, the short-term gender boost treatment reduced enrollment and attendance for girls (by approximately 4 p.p.). However, for attendance, we found increases based on child-based reports. In this case, all treatment arms consistently decrease the probability of girls attending school infrequently by between 5–7 p.p. Without school record data, it is difficult to verify which source is accurate, and thus these impacts are inconclusive.

At endline, the negative effects of the short gender arm treatment on caregiver-reported enrollment and attendance persisted, while the positive effects based on the child-based attendance indicator faded out.

These opposite effects between caregiver- and child-reported attendance were not surprising, given the low correlation between these variables, but we will investigate more the reliability of these indicators when we will measure enrollment and attendance based on administrative data.

4.3.2.3. Child age

Finally, with regard to child age ([Table 5](#)), we found highly heterogeneous effects between younger (5–9-year-olds) and older (10–17-year-olds) children with regard to caregiver-reported engagement. The negative,

although not significant, effects that were observed for the overall sample for caregiver school engagement were driven by the younger children. These were only statistically significant for the short-term gender boost treatment ($p < 0.05$), but treatment effects across different arms are similar in size and direction of effects, leading to a reduction of about a third of an additional item in the caregiver home engagement scale. These effects faded out at endline.

Table 4. Heterogeneity — child gender.

Variables	Caregiver: home engagement	Child: home engagement	Caregiver: school engagement	Caregiver: child enrollment	Child: enrollment	Caregiver: low attendance	Child: low attendance	Caregiver: aspirations likely unfulfilled	Caregiver: returns to JHS	Caregiver: returns to SHS	Caregiver: returns to Univ.
Panel A: Midline											
12-week, behavioural nudges	-0.231	0.058	-0.226	-0.020	-0.012	0.018	-0.070***	-0.033			
	[0.1527]	[0.1328]	[0.1581]	[0.0169]	[0.0197]	[0.0196]	[0.0260]	[0.0248]			
12-week, gender boost	-0.265*	0.118	-0.067	-0.044***	-0.020	0.037**	-0.060**	-0.049**			
	[0.1478]	[0.1337]	[0.1515]	[0.0170]	[0.0186]	[0.0190]	[0.0268]	[0.0245]			
24-week, behavioural nudges	-0.161	-0.018	-0.142	-0.035*	-0.030	0.039*	-0.055**	-0.028			
	[0.1439]	[0.1282]	[0.1503]	[0.0178]	[0.0203]	[0.0203]	[0.0262]	[0.0252]			
24-week, gender boost	-0.202	-0.172	-0.155	-0.008	-0.002	0.024	-0.049*	0.012			
	[0.1476]	[0.1242]	[0.1485]	[0.0160]	[0.0186]	[0.0193]	[0.0268]	[0.0268]			
1.male	-0.111	-0.040	-0.000	-0.032**	-0.021	0.027	-0.043*	-0.004			
	[0.1294]	[0.1106]	[0.1333]	[0.0152]	[0.0175]	[0.0171]	[0.0244]	[0.0219]			
12-week, behavioural nudges #male	0.151	-0.018	0.156	0.038*	0.016	-0.007	0.100***	0.008			
	[0.1877]	[0.1663]	[0.1899]	[0.0218]	[0.0257]	[0.0260]	[0.0335]	[0.0298]			
12-week, gender boost #male	0.192	-0.129	0.075	0.019	-0.017	0.010	0.069**	0.012			
	[0.1829]	[0.1638]	[0.1827]	[0.0239]	[0.0258]	[0.0263]	[0.0334]	[0.0299]			
24-week, behavioural nudges #male	-0.026	-0.028	0.055	0.045**	0.011	-0.011	0.086***	0.010			
	[0.1779]	[0.1590]	[0.1810]	[0.0226]	[0.0263]	[0.0267]	[0.0327]	[0.0304]			

24-week, gender boost #male	0.055	0.241	0.185	0.027	0.012	-0.009	0.057*	-0.024			
	[0.1838]	[0.1606]	[0.1803]	[0.0215]	[0.0254]	[0.0257]	[0.0330]	[0.0314]			
Observations	4,671	4,673	4,673	4,673	4,670	4,673	4,178	4,673			
R-squared	0.0557	0.0819	0.0370	0.0744	0.1116	0.0932	0.0081	0.0484			
Mean outcome var girls	3.085	2.285	3.212	0.932	0.910	0.0930	0.114	0.129			
Panel B: Endline											
12-week, behavioural nudges	-0.120	-0.198	0.031	-0.005	-0.016	0.006	-0.027	-0.011	-0.221	-0.213	-0.243**
	[0.1465]	[0.1473]	[0.1400]	[0.0170]	[0.0214]	[0.0209]	[0.0293]	[0.0319]	[0.1493]	[0.1356]	[0.1174]
12-week, gender boost	0.053	0.084	0.005	-0.043**	-0.008	0.040*	-0.015	-0.050	-0.001	0.016	-0.139
	[0.1468]	[0.1525]	[0.1380]	[0.0178]	[0.0197]	[0.0206]	[0.0295]	[0.0313]	[0.1197]	[0.1141]	[0.1101]
24-week, behavioural nudges	0.116	0.106	-0.184	-0.024	-0.023	0.020	-0.017	-0.052*	-0.076	-0.113	-0.124
	[0.1419]	[0.1464]	[0.1314]	[0.0188]	[0.0210]	[0.0214]	[0.0291]	[0.0306]	[0.1185]	[0.1089]	[0.1016]
24-week, gender boost	0.043	-0.169	0.005	-0.008	0.013	0.025	0.002	-0.084***	0.056	-0.036	-0.073
	[0.1416]	[0.1418]	[0.1347]	[0.0174]	[0.0189]	[0.0211]	[0.0290]	[0.0296]	[0.1274]	[0.1127]	[0.1031]
1.male	0.041	-0.053	-0.118	-0.014	-0.011	0.012	-0.005	-0.019	-0.141	-0.062	-0.026
	[0.1220]	[0.1311]	[0.1093]	[0.0154]	[0.0182]	[0.0173]	[0.0252]	[0.0267]	[0.0885]	[0.0760]	[0.0735]
12-week, behavioural nudges #male	0.005	0.005	-0.011	0.007	0.005	0.015	0.052	-0.032	0.402**	0.232	0.213
	[0.1766]	[0.1871]	[0.1658]	[0.0225]	[0.0285]	[0.0274]	[0.0372]	[0.0374]	[0.1655]	[0.1489]	[0.1300]

12-week, gender boost #male	-0.205	-0.236	-0.067	0.002	-0.019	0.015	0.053	-0.002	0.207	0.083	0.099
	[0.1729]	[0.1868]	[0.1625]	[0.0244]	[0.0260]	[0.0267]	[0.0380]	[0.0363]	[0.1337]	[0.1287]	[0.1218]
24-week, behavioural nudges #male	-0.300*	-0.254	0.152	0.021	-0.001	0.002	0.060	0.036	0.218*	0.184	0.077
	[0.1750]	[0.1896]	[0.1557]	[0.0237]	[0.0272]	[0.0273]	[0.0372]	[0.0372]	[0.1287]	[0.1189]	[0.1101]
24-week, gender boost #male	-0.115	0.229	0.151	0.016	-0.012	-0.018	-0.007	0.031	0.023	0.019	0.002
	[0.1706]	[0.1808]	[0.1571]	[0.0218]	[0.0253]	[0.0262]	[0.0355]	[0.0355]	[0.1470]	[0.1336]	[0.1211]
Observations	4,721	4,721	4,721	4,721	4,610	4,721	4,113	4,721	2,627	3,592	4,447
R-squared	0.0568	0.0628	0.0534	0.0693	0.1069	0.0858	0.0364	0.0389	0.0115	0.0151	0.0441
Mean outcome var	3.457	2.771	3.222	0.927	0.909	0.104	0.171	0.182	5.727	6.120	7.057
Robust standard errors in brackets. ** p<0.01, * p<0.05, * p<0.1											

Table 5. *Impact heterogeneity by child age group.*

Variable	Caregiver: home engagement	Child: home engagement	Caregiver: school engagement	Caregiver: child enrollment	Child: enrollment	Caregiver: low attendance	Child: low attendance	Caregiver: aspirations likely unfulfilled	Caregiver: returns to JHS	Caregiver: returns to SHS	Caregiver: returns to Univ.
Panel A: Midline											
12-week, behavioural nudges	-0.258*	0.013	-0.191	0.019	0.018	-0.005	-0.026	-0.036			
	[0.1385]	[0.1321]	[0.1401]	[0.0167]	[0.0204]	[0.0202]	[0.0224]	[0.0226]			
12-week, gender boost	-0.307**	0.083	0.009	-0.019	-0.017	0.034*	-0.026	-0.061***			
	[0.1381]	[0.1301]	[0.1402]	[0.0176]	[0.0197]	[0.0199]	[0.0225]	[0.0218]			
24-week, behavioural nudges	-0.171	0.124	-0.067	0.003	0.010	0.013	-0.004	-0.043*			
	[0.1328]	[0.1289]	[0.1393]	[0.0170]	[0.0197]	[0.0197]	[0.0228]	[0.0222]			
24-week, gender boost	-0.313**	-0.163	-0.112	0.015	0.019	0.023	-0.002	-0.019			
	[0.1382]	[0.1270]	[0.1395]	[0.0171]	[0.0199]	[0.0212]	[0.0238]	[0.0230]			
1.age_group (1=10-17 yo)	-0.245**	0.008	0.066	-0.004	0.000	0.008	0.023	-0.026			
	[0.1165]	[0.0974]	[0.1037]	[0.0146]	[0.0175]	[0.0165]	[0.0218]	[0.0183]			
12-week, behavioural nudges #age_group	0.222	0.065	0.094	-0.032	-0.040	0.033	0.023	0.016			
	[0.1672]	[0.1451]	[0.1449]	[0.0206]	[0.0252]	[0.0247]	[0.0306]	[0.0264]			
12-week, gender boost #age_group	0.298*	-0.074	-0.066	-0.030	-0.025	0.017	0.010	0.036			
	[0.1711]	[0.1497]	[0.1463]	[0.0225]	[0.0243]	[0.0246]	[0.0315]	[0.0259]			
24-week, behavioural nudges #age_group	0.002	-0.302**	-0.089	-0.022	-0.065**	0.037	-0.004	0.040			
	[0.1579]	[0.1446]	[0.1441]	[0.0217]	[0.0253]	[0.0248]	[0.0304]	[0.0256]			

24-week, gender boost #age_group	0.284*	0.228	0.103	-0.014	-0.025	-0.011	-0.028	0.034			
	[0.1661]	[0.1451]	[0.1458]	[0.0198]	[0.0238]	[0.0247]	[0.0310]	[0.0261]			
Observations	4,671	4,673	4,673	4,673	4,670	4,673	4,178	4,673			
R-squared	0.0568	0.0830	0.0369	0.0759	0.1142	0.0939	0.0063	0.0485			
Mean outcome var for group=0	3.039	2.192	3.195	0.934	0.906	0.102	0.110	0.119			
Panel B: Endline											
12-week, behavioural nudges	-0.178	-0.295*	-0.018	0.029	-0.007	-0.008	-0.036	-0.058*	0.148	0.034	0.027
	[0.1660]	[0.1751]	[0.1491]	[0.0204]	[0.0255]	[0.0250]	[0.0327]	[0.0336]	[0.1332]	[0.1170]	[0.1005]
12-week, gender boost	0.025	-0.271	-0.029	-0.019	-0.027	0.039	-0.020	-0.041	0.142	0.102	-0.021
	[0.1639]	[0.1694]	[0.1467]	[0.0210]	[0.0232]	[0.0237]	[0.0319]	[0.0337]	[0.1164]	[0.0983]	[0.0977]
24-week, behavioural nudges	-0.076	-0.140	-0.067	0.011	0.009	0.003	-0.027	-0.061*	0.018	-0.060	-0.107
	[0.1670]	[0.1758]	[0.1414]	[0.0203]	[0.0228]	[0.0234]	[0.0327]	[0.0336]	[0.1229]	[0.1068]	[0.0976]
24-week, gender boost	-0.013	-0.038	0.070	0.015	0.001	0.024	-0.028	-0.101***	0.057	-0.035	-0.027
	[0.1661]	[0.1739]	[0.1462]	[0.0209]	[0.0243]	[0.0251]	[0.0315]	[0.0316]	[0.1223]	[0.1159]	[0.1023]
1.age_group (l=10-17 yo)	0.545***	-0.262*	-0.027	-0.003	-0.016	0.016	-0.042	-0.045*	0.029	-0.014	0.064
	[0.1324]	[0.1418]	[0.1126]	[0.0174]	[0.0194]	[0.0191]	[0.0274]	[0.0261]	[0.1008]	[0.0849]	[0.0796]
12-week, behavioural nudges #age_group	0.063	0.165	0.070	-0.045*	-0.008	0.032	0.060	0.047	-0.211	-0.176	-0.230*
	[0.1884]	[0.2013]	[0.1677]	[0.0247]	[0.0296]	[0.0295]	[0.0390]	[0.0376]	[0.1644]	[0.1481]	[0.1229]
12-week, gender boost	-0.113	0.362*	0.000	-0.037	0.014	0.015	0.054	-0.017	-0.041	-0.065	-0.101

#age_group	[0.1860]	[0.1934]	[0.1658]	[0.0258]	[0.0277]	[0.0290]	[0.0396]	[0.0366]	[0.1445]	[0.1278]	[0.1165]
24-week, behavioural nudges #age_group	0.028	0.179	-0.046	-0.035	-0.048*	0.026	0.067*	0.047	0.045	0.076	0.037
	[0.1909]	[0.2003]	[0.1618]	[0.0243]	[0.0270]	[0.0276]	[0.0393]	[0.0374]	[0.1448]	[0.1244]	[0.1143]
24-week, gender boost #age_group	-0.021	-0.004	0.035	-0.020	0.010	-0.014	0.041	0.055	0.029	0.019	-0.069
	[0.1897]	[0.1997]	[0.1639]	[0.0244]	[0.0282]	[0.0289]	[0.0386]	[0.0362]	[0.1498]	[0.1397]	[0.1218]
Observations	4,721	4,721	4,721	4,721	4,610	4,721	4,113	4,721	2,627	3,592	4,447
R-squared	0.0735	0.0620	0.0525	0.0726	0.1083	0.0872	0.0348	0.0395	0.0090	0.0148	0.0440
Mean outcome var for group=0	2.984	2.678	3.109	0.933	0.897	0.108	0.170	0.184	5.760	6.201	7.145
Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1											

Also, the improvement in educational expectation that was observed for the whole sample is concentrated among the younger age group, with a decrease in the caregivers' expected probability of not reaching the aspired level of education by 6 p.p. in the short gender boost sample at midline, and by 10 p.p. in the long gender boost treatment arm at endline. No other major differences are observed by child age groups.

4.3.3 Impacts on secondary outcomes

When examining impacts on secondary outcomes, we first report on what we categorise as distal outcomes (results not shown but available upon request). Overall, we found few effects on child outcomes. At the midline assessment there was a negative impact on children's academic outcomes for the 12-week gender-boost arm ($b = -0.100, p < .05$), but insignificant coefficients for all other treatment arms. These impacts faded out at endline. Further, there were no treatment effects on social-emotional outcomes at either time point.

Lastly, we found evidence that caregiver-reported psychological distress increased at endline for the 12-week behavioural nudges treatment arm, and no impacts on other treatment groups.

Regarding secondary outcomes that we categorise as mediators, overall, we did not see significant changes in child time use when looking across treatment arms at midline and endline. Regarding caregiver self-efficacy and emotional supportiveness in the home, we found no impacts at midline. At endline, however, we found reductions in caregiver self-efficacy for the 12-week behavioural nudges treatment arm ($p < .05$), and negative but non-significant coefficients for all other treatment arms. There were no impacts on emotional supportiveness in the home.

5. Policy Implications

This evaluation assessed the effectiveness of an SMS-based intervention on caregiver engagement in education, educational aspirations, and gender bias in rural, remote, and deprived communities in northern Ghana during the Covid-19 pandemic. We find that the programme contributed to improved caregiver engagement only for caregivers who have minimum levels of education. For those who do not have any schooling, the SMS intervention backfired and decreased engagement, especially with younger children. In addition, we did not find a clear difference between households that received the programme for 12 weeks and those that received it for 24 weeks. Questions remain about the optimal length of time (not too short but also not too long) during which these types of messages should be sent to participants to induce behaviour change in the short- and long term. These could be addressed by future studies. We do not find impacts on children learning outcomes.

We learnt several lessons that we hope will be of value to future researchers and policymakers aiming to increase caregiver engagement and children's schooling outcomes and use of SMS-based programmes to reach caregivers.

First, understanding the broader context in which messages are being sent is key to designing programmes in ways that ultimately remove barriers to caregiver engagement (as is the intended effect of nudge-based programmes) rather than creating additional stressors. Our evidence also suggests that tailoring messages based on caregivers' educational backgrounds may be key to supporting the most vulnerable families.

Second, through the pilot, we learnt that the programme implementation should be tailored in terms of (i) caregivers' preferred time of text message delivery; (ii) language. We adjusted the time of day that caregivers received the messages based on their stated preferences, which we believe increased their engagement with the messages based on our pilot. Further, contextualising the language of the messages to both the geographical location of the caregivers as well as their reading level was an important part of making the content more accessible to the clients. By the same token, we learnt that it is important to inform caregivers before programme implementation starts and to brand texts sent with the programme name. This is because mobile phone users often receive lots of spam in their SMS inboxes and therefore do not bother to read the texts

before deleting unsolicited messages. Branding would help them selectively delete unsolicited messages while keeping specific messages on the intervention.

Additional questions for future research include the following.

- Questions about the mechanisms through which programmes may cause additional stress to caregivers, particularly during stressful times such as the Covid-19 pandemic, and for which groups this additional stress may be greatest.
- Whether phone-based interventions reach the intended participant (particularly in households where members share a mobile phone).
- Whether programmes should be more focused around a single theme and not as broad as the EDU+ programme was, which aimed to improve relationships between caregivers and children, improve children's academic and social-emotional outcomes, improve positive disciplinary practices, and increase caregiver engagement in children's education and broader lives.

Future research may also consider how uni-directional text-based interventions compare with other interventions (e.g., face-to-face community dialogues with parents) in terms of impact and cost-effectiveness. In addition, understanding the value-added — if any — of text messages in comparison to in-person interventions.

Finally, future studies could consider involving household members beyond primary caregivers, such as older siblings, grandparents, etc. for supporting child education. This is because these other household members play critical roles in supporting children's education. Involving multiple household members with diverse skills and capabilities such as literacy and the ability to use technology when providing SMS-based nudges to caregivers may enhance the effectiveness of text message interventions. This is so for two reasons.

1. Different household members interact and support children's learning at different times of the day (e.g., older siblings, grandparents, etc.).
2. Household members who cannot read or do not use technology proficiently may rely on household members with the requisite skills for using technology.

6. Conclusion

Our preliminary results suggest that a short, light-touch SMS-based intervention can change caregiver behaviours. However, the story is complicated, and the impacts vary widely by caregiver and child characteristics. In many cases, the interventions operated counter to our hypotheses, and decreased caregiver engagement, decreased self-reported school enrollment and attendance, decreased caregiver mental health, and decreased children's numeracy skills. These negative effects appear to be concentrated for less-advantaged caregivers and children — specifically, caregivers with no formal education, girls, and younger children.

The findings suggest that caregivers may need a base level of education to enact the messages into positive changes for their children. Without this base level of capital, text-based interventions can backfire. Indeed, the caregivers with no formal schooling in our study reported a reduction in caregiver self-efficacy as a result of receiving the messages, while educated caregivers reported increases in this self-efficacy.

These results contribute to a small but growing evidence base about SMS-based nudge interventions to parents and caregivers. Importantly, the majority of studies that have found these types of programmes to improve parenting and child outcomes have been concentrated in middle- or high-income country contexts (see [Bergman, 2019](#) for a review). Our study is one of the first to test this type of programme in a rural, low-income, African setting, and during a public health and economic crisis. Longer-term follow-ups would be important to understand how long the observed changes persisted after the programme was no longer being delivered to caregivers. Our findings suggest that careful consideration of the broader context of caregivers' and children's lives is needed to ensure programmes are tailored in ways that ultimately support caregiver investments, caregiver-child relationships, and children's education. For example, an SMS-based programme implemented and evaluated during the pandemic in El Salvador targeting caregiver mental health also found that the programme reduced caregiver mental health and increased stress, particularly among male caregivers ([Amaral et al., 2021](#)).

We also note that our sample was majority Muslim,⁵ and that midline data collection occurred during Ramadan when many caregivers and many older children were fasting. In addition, our endline data collection took place during the harvesting season, and many children were working in the fields to earn money for school fees to return to school. Thus, the timing of both midline and endline data collection occurred at particularly unusual times of the year. Our results suggest that the macro-context in which families receive these messages may be key to consider when they can be effective and when they may cause additional stress for caregivers. It is possible that the broader context was quite challenging for families at the time the programmes were implemented — e.g., owing to fasting and economic hardship due to the Covid-19 pandemic — and that the messages related to parent and child investments caused additional stress for parents in ways that backfired.

By contrast, nudges seem to be effective in improving caregiver engagement with their children when caregivers have a minimum level of education. We recall that ours is a sample with overall very low education among caregivers. As noted, only 35% of caregivers in our sample have some schooling, and among those caregivers who -attended school at some point, around half of them have at most completed primary.

We have three planned next steps to continue to investigate caregivers' experiences with this programme. First, as soon as we have approval from the Ghanaian Ministry of Education and with funds from another grant, we will collect school administrative records to examine a third source of data on children's school enrollment and attendance, as well as assess impacts on these outcomes several months after endline.

Second, in January 2022, we conducted a qualitative study with 30 randomly selected treatment group participants. Using semi-structured interviews, we asked caregivers about their experiences with the programme, their perceptions of their role in supporting their child's education generally and in the context of the Covid-19 pandemic. We also examined how non-literate parents engaged with the messages. We will receive the transcripts soon and we hope this data will shed additional light on some of our counterintuitive findings.

Third, through a new grant from the LEGO Foundation, and with two Ghanaian colleagues (Richard Appiah at the University of Ghana and

⁵ Although we did not explicitly ask about religion, data show that 60% of caregivers and 35% of children were fasting during Ramadan.

Esinam Avornyo at the University of Cape Coast), we will conduct a community-based participatory research project to inform a deeper understanding of caregiver attitudes towards engagement in their children's education generally, and girls specifically, as well as to inform the adaptation of the Parental Nudges Project programme (PNP) intervention to increase caregiver educational engagement and gender parity. We will work in eight diverse communities in Ghana to gain a deeper understanding of caregiver perceptions about investments in children's learning, their engagement in child education and the challenges they currently face, and to validate and improve the context-specificity of an existing behavioural caregiver / parenting intervention to support caregiver educational engagement. The communities were selected to understand perceptions of education and learning from a diverse group of stakeholders, 2 rural communities in the northern region, 1 peri-urban community in the North; 1 cocoa-growing community; 1 peri-urban community in the Greater Accra Region; 1 southern rural community).

Bibliography

This bibliography is available digitally in our evidence library at <https://docs.edtechhub.org/lib/PWU63GQS>

- Alderman, H., & King, E. M. (1998). Gender differences in parental investment in education. *Structural Change and Economic Dynamics*, 9(4), 453–468. [https://doi.org/10.1016/S0954-349X\(98\)00040-X](https://doi.org/10.1016/S0954-349X(98)00040-X). Available from <https://www.sciencedirect.com/science/article/pii/S0954349X9800040X>. (details)
- Amaral, S., Dinarte Diaz, L., Dominguez, P., & Perez-Vincent, S. M. (2021). *Helping families help themselves? Heterogeneous effects of a digital parenting program*. The World Bank. <https://doi.org/10.1596/1813-9450-9850>. Available from <http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-9850>. (details)
- Attanasio, O. P., & Kaufmann, K. M. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109, 203–216. <https://doi.org/10.1016/j.jdeveco.2014.04.003>. Available from <https://linkinghub.elsevier.com/retrieve/pii/S0304387814000455>. (details)
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. <https://doi.org/10.1111/1467-8624.00273>. Available from <https://onlinelibrary.wiley.com/doi/10.1111/1467-8624.00273>. (details)
- Barnett, I., Ariana, P., Petrou, S., Penny, M. E., Duc, L. T., Galab, S., Woldehanna, T., Escobal, J. A., Plugge, E., & Boyden, J. (2013). Cohort profile: the Young Lives study. *International Journal of Epidemiology*, 42(3), 701–708. <https://doi.org/10.1093/ije/dys082>. Available from <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dys082>. (details)
- Bergman, P. (2019). How behavioral science can empower parents to improve children's educational outcomes. *Behavioral Science & Policy*, 5(1), 52–67. <https://doi.org/10.1353/bsp.2019.0004>. Available from <https://muse.jhu.edu/article/748230>. (details)

- Bettinger, E., Cunha, N., Lichand, G., & Madeira, R. (2021). *Are the effects of informational interventions driven by salience?* (No. 350; Working Paper Series). University of Zurich Department of Economics. <https://www.econ.uzh.ch/static/wp/econwp350.pdf>. (details)
- Björkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics*, 105, 237–253. <https://doi.org/10.1016/j.jdeveco.2013.07.013>. Available from <https://www.sciencedirect.com/science/article/pii/S0304387813001120>. (details)
- Bornstein, M. H., & Putnick, D. L. (2012). Cognitive and socioemotional caregiving in developing countries. *Child Development*, 83(1), 46–61. <https://doi.org/10.1111/j.1467-8624.2011.01673.x>. Available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3270892/>. (details)
- Chapman, C. (2010). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS) correct theta scores for the kindergarten through eighth grade data collections errata*. National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010052>. Available from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010052>. (details)
- Dhar, D., Jain, T., & Jayachandran, S. (2018). *Reshaping adolescents' gender attitudes: evidence from a school-based experiment in India* (Working Paper No. 25331). National Bureau of Economic Research. <https://doi.org/10.3386/w25331>. Available from <https://www.nber.org/papers/w25331>. (details)
- Doss, C., Fahle, E., Loeb, S., & York, B. (2017). *Supporting Parenting through Differentiated and Personalized Text-Messaging: Testing Effects on Learning During Kindergarten*. <https://cepa.stanford.edu/content/supporting-parenting-through-differentiated-and-personalized-text-messaging-testing-effects-learning-during-kindergarten>. (details)
- ECG, & ISSER. (n.d.). *EGC-ISSER-Northwestern Ghana Panel Survey*. Yale Economic Growth Center. Retrieved March 18, 2022, from <https://egc.yale.edu/data/egc-isser-northwestern-ghana-panel-survey>. (details)
- Fink, G., Karlan, D., Udry, C., & Osei, R. (2019). *Communications for Development*. Harvard Dataverse. <https://doi.org/10.7910/DVN/CQRNC7>. Available from

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CQRNC7>. (details)

Guay, F., Marsh, H. W., Dowson, M., & Larose, S. (2005). *Assessing academic motivation among elementary school children: the Elementary School Motivation Scale (ESMS)*. Australian Association for Research in Education Annual Conference, Parramatta.
<https://www.aare.edu.au/data/publications/2005/gua05378.pdf>. (details)

Innovations for Poverty Action. (2020). *RECOVR Ghana: Tracking the Effects of the COVID-19 Pandemic*. Innovations for Poverty Action.
<https://www.poverty-action.org/recovr-study/recovr-ghana-tracking-effects-covid-19-pandemic>. (details)

Kessler, R., Andrews, G., Colpe, L., Hiripi, E., Mroczek, D., Normand, S.-L., Walters, E., & Zaslavsky, A. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976.
<https://doi.org/10.1017/S0033291702006074>. (details)

Lichand, G., & Wolf, S. (2021). *Arm-wrestling in the classroom: the non-monotonic effects of monitoring teachers* (No. 357; Working Paper Series). University of Zurich Department of Economics.
<https://www.econ.uzh.ch/en/research/workingpapers.html?paper-id=1039>. Available from
<https://www.econ.uzh.ch/en/research/workingpapers.html?paper-id=1039>. (details)

Lichand, G., Christen, J., & van Egeraat, E. (2021). *Behavioral nudges to prevent learning losses and dropout risk during the pandemic: experimental evidence from Brazil* (Working Paper No. 363). University of Zurich Department of Economics.
<https://doi.org/10.2139/ssrn.3724386>. Available from
<https://papers.ssrn.com/abstract=3724386>. (details)

McCoy, D., Salhi, C., Yoshikawa, H., Black, M., Britto, P., & Fink, G. (2018). Home- and center-based learning opportunities for preschoolers in low- and middle-income countries. *Children and Youth Services Review*, 88, 44–56. <https://doi.org/10.1016/j.childyouth.2018.02.021>. (details)

Myers, L., Crawford, E., Murphy, C., Aka-Ezoua, E., & Felix, C. (2018). Eyes in the room trump eyes on the screen: effects of a responsive co-viewer

- on toddlers' responses to and learning from video chat. *Journal of Children and Media*, 12(3), 275–294.
<https://doi.org/10.1080/17482798.2018.1425889>. (details)
- Nesbitt-Ahmed, Z., & Subrahmanian, R. (2020, April 23). Caring in the time of COVID-19: Gender, unpaid care work and social protection. *UNICEF Evidence for Action*.
<https://blogs.unicef.org/evidence-for-action/caring-in-the-time-of-covid-19-gender-unpaid-care-work-and-social-protection/>. (details)
- Osei-Assibey, E. (2014). Nature and dynamics of inequalities in Ghana. *Development*, 57(3–4), 521–530. <https://doi.org/10.1057/dev.2015.25>. Available from
<http://ugspace.ug.edu.gh/bitstream/handle/123456789/24257/Nature%20and%20Dynamics%20of%20Inequalities%20in%20Ghana.pdf?sequence=2&isAllowed=y>. (details)
- Ray, D. (2006). Aspirations, poverty, and economic change. In *Understanding Poverty* (pp. 409–422). Oxford University Press.
<https://doi.org/10.1093/0195305191.003.0028>. Available from
https://www.google.co.uk/books/edition/Understanding_Poverty/qVNLt3Ce1EC?hl=en&gbpv=0. (details)
- Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and Commitment Therapy*, 61. (details)
- Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, 77(3), 786–799. <https://doi.org/10.1111/j.1467-8624.2006.00903.x>. (details)
- UNICEF. (2010). *Child disciplinary practices at home: evidence from a range of low- and middle-income countries*.
<https://data.unicef.org/resources/child-disciplinary-practices-at-home-evidence-from-a-range-of-low-and-middle-income-countries/>. (details)
- 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(3), 197–208. <https://doi.org/10.1111/1471-6402.00021>. Available from
<http://journals.sagepub.com/doi/10.1111/1471-6402.00021>. (details)
- Wolf, S., Aurino, E., Osei-Akoto, I., & Ackwerh, E. Y. B. (2021). *Nudges to improve learning and gender parity: Supporting parent engagement and Ghana's educational response to COVID-19 using mobile phones*.

- World Bank Strategic Impact Evaluation Fund (SIEF).
<https://www.worldbank.org/en/programs/sief-trust-fund/brief/ghana-nudges-to-improve-learning-and-gender-parity-supporting-parent-engagement-and-ghana-s-educational-response-to-covi>. (details)
- Wolf, S., Aurino, E., Suntheimer, N., Avornyo, E., Tsinigo, E., Jordan, J., Samanhyia, S., Aber, J. L., & Behrman, J. R. (2021). *Learning in the time of a pandemic and implications for returning to school: effects of COVID-19 in Ghana* (CPRE Working Papers). Consortium for Policy Research in Education (CPRE).
https://repository.upenn.edu/cgi/viewcontent.cgi?article=1027&context=cpre_workingpapers. Available from
https://repository.upenn.edu/cgi/viewcontent.cgi?article=1027&context=cpre_workingpapers. (details)
- World Bank, UNESCO, & UNICEF. (2021). *The state of the global education crisis: a path to recovery*.
<https://documents1.worldbank.org/curated/en/416991638768297704/pdf/The-State-of-the-Global-Education-Crisis-A-Path-to-Recovery.pdf>. (details)
- World Bank. (2018). *Poverty and shared prosperity 2018: piecing together the poverty puzzle*. World Bank.
<https://doi.org/10.1596/978-1-4648-1330-6>. Available from
<http://hdl.handle.net/10986/30418>. (details)
- York, B. N., Loeb, S., & Doss, C. (2019). One step at a time: the effects of an early literacy text-messaging program for parents of preschoolers. *Journal of Human Resources*, 54(3), 537–566.
<https://doi.org/10.3368/jhr.54.3.0517-8756R>. Available from
<http://jhr.uwpress.org/content/54/3/537>. (details)