Arm-Wrestling in the Classroom: The Non-Monotonic Effects of Monitoring Teachers

By Guilherme Lichand and Sharon Wolf*

Teacher absenteeism and shirking are common problems in developing countries. While monitoring teachers should ameliorate those problems, mobilizing parents to do so often leads to small or even negative effects on learning outcomes. This paper provides causal evidence that this might result from non-monotonic effects of monitoring teachers. Cross-randomizing nudges to teachers and parents in Ivory Coast – to motivate and monitor teachers directly, and to mobilize parents –, we find that in schools where either parents or teachers are nudged, student dropouts decrease by nearly 50%. In contrast, in schools where both are nudged, there is no effect on dropouts.

JEL: C93, D23, D91, I25
Keywords: Moral Hazard; Monitoring; Education; Teachers

A famous cartoon portrays parents’ reactions when their child comes home with an ‘F’ grade, in the 1960’s and in the 2000’s: parents from the past confront their child about her poor grade; in contrast, parents from the present side with the child to confront the teacher instead. While teacher accountability might have indeed changed over time, the cartoon hints at a more fundamental question: is monitoring teachers good or bad for children’s education? This paper investigates the effects of monitoring teachers through different strategies, providing first-hand causal evidence that the answer to that question might depend on monitoring intensity.

Teacher effectiveness has been identified as a key hurdle to human capital formation in developing countries. Across seven Sub-Saharan African countries, Bold et al. [2017] document that not only do many teachers lack minimum knowledge in language and mathematics, they are also very often absent from the classroom (on average, 44% of the time). That outcome is unsurprising in face of moral
hazard: parents delegate the provision of costly inputs to their children’s education but cannot perfectly observe the extent to which teachers put effort into it. Since teachers are typically not paid by performance, this tends to induce under-provision of effort relative to what parents would have optimally set. As such, having parents monitor teachers more closely should increase effort and lead to better educational outcomes. Having said that, monitoring might actually crowd out intrinsic motivation and effort (Fehr and Gächter 2000), e.g. because teachers perceive it as an indication of distrust (Frey 1993). Which effect dominates? Does too much monitoring backfire, with parents and teachers “arm-wrestling” over children’s education?

We study this question in the context of Ivory Coast, a West-African country where about half of 6th-graders are considered not competent in reading and nearly 2/3 are not competent in Math (World Bank 2018). We cross-randomize nudges to teachers and caregivers across 296 classrooms (2nd, 4th and 6th grades) in 100 public schools. Nudges to teachers aim at increasing their effort directly, encouraging attendance and time-on-task, and monitoring them about compliance with suggested activities. Because the program was run in partnership with the Ministry of Education, it carried an official weight. Nudges to parents aim at increasing their engagement in their children’s education, including showing up in school to monitor teachers. In treated schools, teachers report that parents show up to inquire about their children to a much greater extent: in those schools, 24% of teachers report that the typical parent shows up at least once a week, compared to 14% in the control group. Both teachers and caregivers receive nudges twice a week through text messages (SMS) over the course of the whole school year. Our experimental design has several advantages. Nudging teachers and parents independently allows us to estimate the causal effects of monitoring teachers through different strategies – directly, or through community monitoring – on teachers’ effort and educational outcomes. In turn, nudging teachers and parents concurrently allows us to estimate the causal effects of higher-intensity monitoring.

Results are as follows. We find that nudges to teachers decrease student dropouts across all grades by 2.23 p.p., a nearly 50% reduction relative to control schools. Similarly, nudges to parents decrease dropouts by 2.47 p.p. In contrast, in schools where both parents and teachers are nuded, there is no impact on dropouts relative to the control group (a -0.32 p.p. effect size). While there is no one-to-one relationship between student dropouts and teachers’ attendance, the way the latter responds to the different interventions is revealing about the

---

3 For many public school students in developing countries, primary caregivers are not the parents but, rather, close relatives (e.g. grandparents). In this paper, we use parents and caregivers interchangeably for ease of exposition.

4 Nudges to parents have been shown to significantly improve educational outcomes in Brazil; see Bettinger et al. (2020).

5 We also randomized whether parents received Eduq+ over SMS or automated phone calls (Interactive Voice Response units, IVR). Results are shown in Appendix E.
underlying mechanism. While teachers’ attendance is not affected when they are nudged independently, it actually decreases relative to the control group when teachers and parents are nudged concurrently. This is not because nudges to parents are disruptive on their own: in schools where only parents are nudged, teachers’ attendance actually increases relative to the control group (although not statistically significant). The evidence is consistent with the hypothesis that monitoring mitigates moral hazard at first, but eventually backfires as its intensity becomes too high.

Since the only dimension of teachers’ effort that we capture is attendance, it could be that high-intensity monitoring leads to worse educational outcomes not because teachers put in less effort as a result, but rather because monitoring (inefficiently) displaces teachers’ effort to other margins (e.g. from teaching in the classroom to grading homework). We rule out that alternative mechanism in our setting by taking advantage of heterogeneous treatment effects by teachers’ attendance at baseline. Among below-median attendance teachers, absenteeism and dropouts both decrease relative to the control group to the same extent across all treatment arms – including when parents and teachers are nudged concurrently. As such, there is no mechanical displacement induced by high-intensity monitoring. In reality, backfiring is driven by above-median attendance teachers, whose attendance decreases in response to all treatment arms relative to the control group. In schools where both parents and teachers are nudged concurrently, student dropouts actually increase, by over 20% relative to the control group. Monitoring teachers who were already putting in more effort before the onset of the intervention seems to crowd out their intrinsic motivation. While their students’ likelihood of dropping out decreases at first despite lower teacher attendance, as monitoring intensity increases, educational outcomes eventually deteriorate.

Consistent with non-monotonic effects of monitoring, teachers, parents become more optimistic about their children’s numeracy skills when either parents or teachers are nudged independently, but, strikingly, become more pessimistic when both are nudged concurrently. In tune, corporal punishment decreases significantly when teachers are nudged independently, but does not change when both teachers and parents are nudged. While there are no differences in students’ self-reported effort or in their standardized test scores across treatment conditions, the patterns of treatment effects on parental beliefs and corporal punishment certainly impact students’ motivation to stay in school.

Is non-monotonicity a general property of how teachers respond to monitoring intensity, or is this result a special feature of intervening on parents and teachers concurrently? In particular, if teachers dislike having parents ‘step on their toes’, sending text messages to both could trigger backlash because it violates social

---

norms about the division of labor (Görges and Nosenzo 2020). To answer that question, we revisit all 296 teachers and 1 caregiver per classroom (randomly drawn) in a follow-up survey, eliciting how they would change their inputs to children’s education in response to different levels of effort by the other party. Given the particular nature of team production in education – it is a many-to-1 problem, whereby many parents combine inputs with one teacher – we ask teachers about their planned effort in response to different effort levels by caregivers varying also the share of caregivers who undertake that effort. We find that teachers’ planned effort indeed decreases when too many parents show up too often in school, consistent with the “arm-wrestling” mechanism underscored by our experimental findings. To that point, estimating heterogeneous treatment effects of nudges to parents with respect to the share of caregivers who show up in school regularly (reported by students), we find that the treatment effect on student dropouts is $U$-shaped with respect to parental engagement: effect sizes eventually become positive at the high-end of the distribution. This not only confirms that non-monotonic responses of teachers’ effort to monitoring intensity are likely to emerge even when they are not directly targeted by monitoring interventions, but also rules out that our results are an artifact of how additional elements covered by nudges to parents (unrelated to monitoring) interact with nudges to teachers.

Our non-monotonicity finding helps reconcile seemingly inconsistent results in the literature on parent-teacher moral hazard. On the one hand, [Kremer and Holla 2009] enumerates successful monitoring interventions that mitigate the effects of informational asymmetries between parents and teachers, particularly when able to provide conditional incentives (e.g. Duflo, Hanna and Ryan 2012). On the other hand, [Banerjee et al. 2010] finds that participatory programs to monitor teachers have no effect on teacher’s effort or students’ test scores, and [Wolf et al. 2019] finds that while a teacher training intervention in Ghana has large positive effects on literacy among preschool children, the positive learning effects disappear when parents are assigned to parent-teacher association meetings to inform them of the new teaching practices and encourage them to engage with the teacher. Our results suggest that non-monotonic effects of monitoring likely explain why some programs work, but not others – especially when community monitoring happens in combination with programs targeting teachers, which might also be perceived as monitoring.

Our results also contribute to a large literature on the economics of education that discusses how parents’ and teachers’ inputs matter for children’s human capital production (Glewwe and Kremer 2006). Observational data has been extensively used to back up claims about how those inputs interact, sometimes in favor of complementarity (e.g. Fryer Jr, Devi and Holden 2016; World Bank 2018; Agostinelli, Saharkhiz and Wiswall 2019), and sometimes in favor of substitutability (Greaves et al. 2019; Chang, Cobb-Clark and Salamanca 2020).\footnote{In our sample, both parents and teachers see each other’s inputs as complements in the education production function; see Appendix A.}
Going beyond the question on how different inputs get combined through the education production function, we show that parents’ and teachers’ effort levels respond endogenously to changes in each others’ inputs. This is the first paper to show that teachers’ best-response function is non-monotonic with respect to parents’ effort, making educational inputs behave like complements or substitutes over different ranges of the support distribution.

Beyond education, our empirical findings showcase that monitoring can crowd out intrinsic motivation even in the absence of financial incentives, and decrease effort above and beyond merely displacing it to other margins that are not monitored to the same extent. While laboratory experiments document that monitoring above a certain threshold can decrease work effort (e.g. [Falk and Kosfeld 2006], [Dickinson and Vileval 2008]), to our knowledge, this is the first paper to provide rigorous evidence for that phenomenon in the field.

I. Background

This section describes the Ivorian educational setting in subsection [A] followed by details about our interventions in subsection [B].

A. Education in Ivory Coast

Education in Ivory Coast is pervasively low-quality: Ivorian students’ literacy and numeracy skills are rated among the poorest in West Africa ([World Bank 2018]). Out of the 7 million Ivorian school children, most have to rely on public schools for their primary education, especially outside of Abidjan (its main economic center). In those schools, children are taught by professionals who often have very limited training and precarious knowledge about the subjects they teach ([Bold et al. 2017]). What is worse, it is often the case that teachers say there is little they can do if parents do not have the necessary education ([World Bank 2018]). Since educational achievement in the country is very low, this hints at a potential vicious cycle, whereby neither parents nor teachers contribute significantly to producing children’s education. While in our sample less than 20% of teachers concur with that statement, over 25% of parents do. What is more, 30-35% of both teachers and parents agree that there is little teachers can do if parents have too many personal or financial problems – which are of course widespread in the poor setting where the study takes place.

Primary education is organized in three cycles: CP (grades 1 and 2), CE (grades 3 and 4) and CM (grades 5 and 6). Children are assessed at the end of each cycle (CP2, CE2 and CM2). Because we are interested in outcomes such as grade repetition, our study focuses on those final years of each cycle, comprising 2nd, 4th and 6th graders. In our sample, the typical number of students is 49 per

8Related (quasi-)experimental studies are [Das et al. 2013], which shows substitution between spending by parents and schools, and [Fryer Jr, Devi and Holden 2016], which shows substitution across different academic subjects.
classroom at CP2, 46 at CE2 and 40 at CM2. Since each teacher is responsible for the two grades within a cycle, that means they oversee 80-100 students in a typical classroom. Primary completion rates are low (73.3%), with the majority of children dropping out before completing the three primary cycles [9]. In our sample, average grade repetition rates across all primary grades are over 15%, and yearly dropout rates average 4.7%. Children learn very little during the first two years of primary education: we find no difference in second-graders' average scores in literacy and numeracy standardized tests at the beginning and at the end of the school year [Wolf, Lichand and Deambrosi, 2020].

Our study takes place in the cocoa producing regions of Aboisso and Bouafle. Besides the usual challenges of developing country settings, along with Ghana, Ivory Coast hosts approximately 2/3 of the world production of cocoa. This has been linked to one of the highest incidences of child labor worldwide, with almost 2 million children employed in cocoa fields [10]. In our sample, nearly 45% of children report to actively help with farm labor. What is more, corporal punishment is quite widespread in this setting as a disciplining strategy, raising additional challenges to children’s human capital formation in face of the correlational evidence about its negative effects on socio-emotional development [Wolf and Suntheimer, 2020b].

B. Nudges to parents and teachers

The educational program we offer parents in the context of our experiment (Eduq+, powered by the Brazilian EdTech Movva) delivers nudges via text messages (SMS) directly to parents’ and teachers’ mobile phones [11]. Nudges are organized in thematic sequences – comprised of four text messages –, with two messages delivered each week. Inspired by READY4K! (York, Loeb and Doss, 2018), sequences start with a motivating fact, followed by a suggested activity; users then receive an interactive message in the following week, posing them a question linked to the activity suggested the week before; last, a “growth” message meant to highlight that engaging in activities such as those should be routine concludes the sequence. Content is catered to students’ age group. The intervention was implemented over the entire 2018-19 school year.

In the context of our study, nudges to teachers aim at increasing their attendance and time-on-task while teaching. Messages suggest pedagogical strategies, encourage teachers to show up to school everyday and discourage corporal punishment. Every other week, an interactive message checks on whether teachers are complying with suggested activities. Because the program is run in partnership with the Ministry of Education, it carries an official weight.

Nudges to parents aim at boosting motivation and beliefs about returns to

[11] Movva (http://movva.tech) delivers nudges to engage parents in their children’s education across Brazil and Ivory Coast. One of the authors (Guilherme) is Movva’s co-founder and chairman.
investments in children’s education. Messages suggest non-curricular activities to do with children, and encourage parents to take an interest in their children’s school life. Some messages explicitly encourage caregivers to show up to parent-teacher meetings and, in treated schools, teachers report that parents show up to inquire about their children to a much greater extent (in the treatment group, 24% of teachers report that the typical parent shows up in school at least once a week, compared to 14% in the control group).\textsuperscript{12}

II. Empirical strategy

This section discusses our empirical strategy. Subsection II.A introduces our experiment design, followed by a summary of the outcomes we draw upon in subsection II.B. Last, subsection II.C describes the equations we estimate, our treatment of standard errors, and how we deal with multiple comparisons.

A. Experiment design

All details of the experimental design and a pre-analysis plan were pre-registered at the AEA RCT Registry on October 31, 2018 (AEARCTR-0003385). Appendix B presents the pre-analysis plan in full, and Section B.B2 highlights the elements of the analyses that deviate from what had been pre-specified.

We cross-randomize nudges to teachers and parents across 296 classrooms (2nd, 4th and 6th grades) in 100 Ivorian public schools.\textsuperscript{13} Randomization was undertaken at the school-level to prevent spillovers.\textsuperscript{14} Out of the 100 schools, 25 were randomly assigned to receive no intervention (the control group); 25 were assigned to have only teachers nudged; 24 were assigned to have only parents nudged; and 26 were assigned to have both teachers and parents receive nudges.\textsuperscript{15}

At each school, enrollment was voluntary: teachers and caregivers merely had to provide their phone numbers to receive nudges at no cost. Because the interventions are undertaken in partnership with the Ministry of Education, all CP2, CE2 and CM2 teachers consented to participate. All primary caregivers with phones were also enrolled, either directly in school meetings to publicize the program, or indirectly via the school principal; in both cases, they could opt-out of the program at any point via text message.

We also randomized whether caregivers received nudges over text or audio messages (Interactive Voice Response units, IVR). We find no significant differences between the two delivery modes; results are shown in Appendix E.

\textsuperscript{12}Nudges to parents have been shown to significantly improve educational outcomes in Brazil; see Bettinger et al. (2020).

\textsuperscript{13}Four of those schools only covered the first two cycles of primary education.

\textsuperscript{14}We estimate geographic spillovers of nudges to parents based on random variation in treatment saturation within different radius of students’ households; see Appendix E.

\textsuperscript{15}The reason for the asymmetry is that local partners were particularly interested in the arm at which both parties were nudged.
Appendix A shows that the different treatment arms are balanced: there are no significant differences in baseline characteristics of students, teachers or schools across cells.

**B. Data and outcomes**

The first set of outcomes we have access to is administrative data on grade repetition and dropouts, at the classroom level, shared by school districts upon request of the Ministry of Education.

The second set of outcomes we have access to is data on parents’ and teachers’ inputs (reported by students), teachers’ motivation and burnout, parental beliefs about their child’s school performance (captured by their best guess of their child’s math/language grade), the extent of corporal punishment (reported by parents) and child labor (reported by parents and students), students’ effort (self-reported) and standardized numeracy and literacy test scores, from baseline and end line surveys. The school year ran from October 2018 (when we collected our baseline survey) to June 2019 (when we collected end line survey). Within each school, we randomly drew 13 CP2 students and 12 CE2 students to be surveyed at each wave (we tried to track the same students at both waves, with a 91% success rate). As such, our sample comprises 1,159 CP2 students and 1,114 CE2 students surveyed both at baseline and end line, along with their primary caregivers and teachers. For test scores, we adapted PASEC\(^{17}\) and IDELA (Pisani, Borisova and Dowd \(^{18}\) items after extensive piloting to prevent floor effects.

The third set of outcomes we have access to is data on teachers’ career plans, parents’ and teachers’ beliefs about returns to inputs by each party, and their best-response functions in response to the other party’s inputs, collected in an extra follow-up wave in October 2019, at the very beginning of the following school year. For that wave, we surveyed the 96 CM2 teachers for the first time, and revisited all 200 CP2 and CE2 teachers and 1 caregiver per classroom (randomly drawn, given budget constraints) across all 296 classrooms in our sample. When it comes to beliefs about returns, we elicit parents’ and teachers’ estimates of what the typical child’s math and language grades would have been under different combinations of effort levels by each party (how frequently – never, once a year, once a month, once a week or everyday – parents show up in school, and how frequently teachers help students who are lagging behind). Last, when it comes to best-response functions, we elicit parents’ and teachers’ planned effort levels in response to different effort levels by the other party (same scale). For teachers, we ask how they would change their effort in response to how frequently parents show

\(^{16}\)In a companion paper (Wolf and Lichand \(^{2020}\), we characterize the main predictors of learning outcomes in Ivory Coast, including additional outcomes not shown in this paper such as children’s socio-emotional skills.


\(^{18}\)A complete account of the effects of the intervention on learning outcomes is presented in a companion paper (Wolf and Lichand \(^{2020}\)).
up in school unannounced to talk about what their children are learning, under two scenarios: (a) for 1 out of 10 parents showing up at the specified frequency, and (b) for 10 out of 10 parents.

C. Estimation

For grade repetition and dropouts, we estimate intention-to-treat effects of each treatment cell using OLS regressions, as follows:

\[ Y_{cs} = \alpha + \beta_1 \text{Teachers}_s + \beta_2 \text{Parents}_s + \beta_3 \text{Both}_s + \varepsilon_{cs}, \]

where \( Y_{cs} \) stands for the fail/dropout rate within classroom \( c \) at school \( s \); \( \text{Teachers}_s = 1 \) if school \( s \) was assigned to have only teachers nudged, and 0 otherwise; \( \text{Parents}_s = 1 \) if school \( s \) was assigned to have only parents nudged, and 0 otherwise; \( \text{Both}_s = 1 \) if school \( s \) was assigned to have both teachers and parents nudged, and 0 otherwise; and \( \varepsilon_{cs} \) is an error term. We are interested in testing \( \beta_1 \leq 0, \beta_2 \leq 0 \) and \( \beta_3 \leq \beta_1, \beta_2 \).

For survey outcomes that were measured at both baseline and end line, we include student fixed-effects, as follows:

\[ Y_{icst} = \alpha + \beta_1 \text{Teachers}_{st} + \beta_2 \text{Parents}_{st} + \beta_3 \text{Both}_{st} + \theta_i + \varepsilon_{icst}, \]

where \( Y_{icst} \) stands for the outcome (e.g., standardized test score) for child \( i \) in classroom \( c \) at school \( s \) measured at wave \( t \); all treatment indicators equal 0 at \( t = 0 \); and \( \theta_i \) stands for student fixed-effects. For outcomes measured at the extra follow-up wave, we cannot include student fixed-effects (since those outcomes had not been elicited in previous waves).

We cluster standard errors at the classroom level across all specifications, allowing outcomes to be arbitrarily correlated among students (and their caregivers) under the same teacher. For outcome variables based on multiple questions (such as scales for parental engagement), we combine different variables using summary measures to deal with family-wise error rates, following Kling, Liebman and Katz (2007).

As specified in our pre-analysis plan, we build summary measures for outcomes with multiple components. Parent’s effort comprises the extent to which parents help with or ask about homework, school notes or about how school was; help organizing school materials; incentivize attendance and studying; attends parent-teacher meetings; and talks to the teacher. Corporal punishment comprises different questions to caregivers about whether they find appropriate to punish children across different instances (or had recently done so). Students’ effort comprises self-reported time use in academic activities, including the extent to which children study at home; read books or magazines; and listen to the radio.\(^{19}\)

\(^{19}\)We include listening to radio as part of the engagement summary measure in face of the evidence
Since our treatments affected parental engagement, assignment influenced our ability to track students at end line relative to the control group (see Appendix A). For this reason, we estimate Lee bounds (Lee, 2009) for our main survey outcomes to account for the effects of selective attrition on sample composition (see Appendix G).

Last, because of issues such as phone sharing and network availability, not all subjects might have received nudges as assigned. Appendix D handles imperfect compliance with the treatment assignment by estimating an instrumental variable model, whereby treatment assignment is used as an instrument for the share of caregivers who, at the end line survey, acknowledge that they received text messages from the school over the course of the academic year.

III. Results

We start by estimating the effects of the interventions on school dropouts. Table 1 estimates equation 1 using administrative data on dropouts. Column (1) presents estimates pooling observations for classrooms of all grades; columns (2) to (4) present treatment effects on CP2, CE2 and CM2, respectively.

Results are as follows. Across all grades (column 1), in schools where only parents are nudged, dropouts decrease by 2.47 p.p. – a huge effect size, over 50% the average dropout rate in the control group, statistically significant at the 1% level. In schools where only teachers are nudged, dropouts decrease by 2.23 p.p., significant at the 5% level. In contrast, nudging both teachers and parents concurrently has a tiny and statistically insignificant effect size (-0.32 p.p., less than 7% the control group average). The coefficient of the joint intervention is statistically different from those of nudging either parents (p = 0.02) or teachers (p = 0.04) independently. Such patterns are identical within each grade: the effect size of the joint intervention is always much smaller than those of intervening independently on each party. Effects are especially large among older children: strikingly, dropout rates fall by nearly 75% at the final primary cycle in response to nudging either parents or teachers. Still, the treatment effect of nudges to both within that grade is again not only statistically insignificant, but less than 10% the dropout rate in the control group.

Next, Figure 1 showcases heterogeneous treatment effects with respect to teachers' baseline attendance (over the previous 14 days, reported by students), splitting the sample at the median. Teachers' median attendance at baseline is pretty high (90.4%) but high-variance, with some teachers missing over 50% of classes over the two previous weeks even at the very beginning of the school year. Panel A in Figure 1 shows that nudging parents or teachers independently has sizeable negative effects on dropouts across all teachers. Among teachers with that this develops language skills, particularly in the earlier grades (Abimpaye et al., 2019). In fact, only 35% of children in our sample report listening to radio at baseline. Appendix C shows that the same pattern holds for grade repetition, although not statistically significant within any grade.
### Table 1—Treatment effects on student dropout rates

<table>
<thead>
<tr>
<th></th>
<th>All grades</th>
<th>CP2</th>
<th>CE2</th>
<th>CM2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parents</strong></td>
<td>-0.0247***</td>
<td>-0.0147</td>
<td>-0.0286</td>
<td>-0.0316*</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.0153)</td>
<td>(0.0177)</td>
<td>(0.0166)</td>
</tr>
<tr>
<td><strong>Teachers</strong></td>
<td>-0.0223**</td>
<td>-0.0086</td>
<td>-0.0270</td>
<td>-0.0318*</td>
</tr>
<tr>
<td></td>
<td>(0.0094)</td>
<td>(0.0151)</td>
<td>(0.0175)</td>
<td>(0.0163)</td>
</tr>
<tr>
<td><strong>Both</strong></td>
<td>-0.0032</td>
<td>-0.0017</td>
<td>-0.0038</td>
<td>-0.0041</td>
</tr>
<tr>
<td></td>
<td>(0.0092)</td>
<td>(0.0150)</td>
<td>(0.0173)</td>
<td>(0.0159)</td>
</tr>
<tr>
<td><strong>Control group mean</strong></td>
<td>0.0468</td>
<td>0.0458</td>
<td>0.0515</td>
<td>0.0428</td>
</tr>
</tbody>
</table>

|                  |            |          |          |          |
| **Parents = Both [p-value]** | 0.0228 | 0.3928  | 0.1598  | 0.0948  |
| **Teachers = Both [p-value]** | 0.0398 | 0.6453  | 0.1830  | 0.0859  |

| **Observations** | 296 | 100 | 100 | 96 |
| **R-squared**    | 0.0364 | 0.0121 | 0.0441 | 0.0676 |

*Note:* Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Panel A: Student dropouts

Panel B: Teacher attendance

Figure 1. Heterogeneous treatment effects of monitoring on teacher attendance and student dropouts, by teachers’ median baseline attendance

Note: In Panel A, student dropouts available only at the classroom-level. In Panel B, teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Median of teachers’ baseline attendance is 90.37%. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. P-values shown in dark brackets.
above- (below-) median attendance at baseline, nudges to parents decrease student dropouts by 66.6% (22.6%), and nudges to teachers decrease those by 33.8% (39.7%). In contrast, nudging parents and teachers concurrently still decreases dropouts among below-median teachers (by 28.8%), but increases them by over 20% among above-median ones.

Figure 2 showcases treatment effects on parents’, teachers’ and children’s inputs, displaying effect sizes for the different treatment arms.

![Figure 2. Treatment effects on parents’, teachers’ and children’s inputs](image)

Note: Effect sizes are reported for intention-to-treat estimates, with student fixed-effects for all outcomes except parent monitoring. For this outcome, we use survey responses from the extra follow-up (CP2, CE2, CM2). Parent monitoring = 1 when teachers report that the caregiver of the typical child in their classroom last year showed up in school “Sometimes” or “Always”, and 0 otherwise. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. Standard errors clustered at the classroom level.

Parents show up in school to monitor teachers to a greater extent in response to nudges; if anything, the effect size is even larger (and significant at the 10% level) when teachers are nudged concurrently. In contrast, teacher attendance is not significantly affected on average by nudges to either parents or teachers inde-
pendently, and even decreases when both are nudged (although not statistically significant). Having said that, as Panel B in Figure 1 shows, those results mask important heterogeneity. On the one hand, nudges always increase attendance by teachers with below-median baseline attendance (significantly at the 10% level) when either parents or teachers are nudged independently. On the other hand, nudges always decrease it among above-median teachers, significantly so when parents and teachers are nudged concurrently. Since the increase in attendance among below-median teachers is the lowest in schools where both parents and teachers are nudged, and since the decrease in attendance among above-median teachers is the highest in those schools, the net treatment effect of monitoring on teacher attendance in the aggregate becomes negative as monitoring becomes too high.

Next, even though treatment effects on parental beliefs are noisily estimated in Figure 2, their ranking of effect sizes is consistent with non-monotonicity. Parents become more optimistic about their children’s math performance (0.14 s.d.) when they are nudged independently; when only teachers are nudged, beliefs still improve, but to a lesser extent (0.04 s.d.); when both are nudged, strikingly, parents become more pessimistic about their children’s math grades (-0.09 s.d.). Treatment effects on corporal punishment also conform to those patterns: while children are physically disciplined to a significantly lesser extent when only teachers are nudged (by 0.28 s.d., significant at the 1% level), this effect size falls by nearly 2/3 when teachers and parents are nudged concurrently.

In turn, students’ self-reported effort increases in response to all treatments relative to the control group by 0.32-0.36 s.d. (all significant at the 1% level), and the intervention does not systematically improve learning outcomes. All in all, evidence suggests that lower dropouts must be mediated by students’ motivation, driven by a combination of parents being more optimistic about their children’s school performance, and children being exposed to corporal punishment to a lesser extent. Symmetrically, higher dropout rates among students of high-effort teachers nudged concurrently with parents are likely driven by lower teacher attendance and parents being more pessimistic about their children’s numeracy skills.

When it comes to the issue of whether non-monotonicity is a general property of how teachers respond to monitoring intensity or an artifact of the specific features of our setup, we tackle that question in two ways. First, we analyze best response functions with respect to the other party’s effort. Figure 3 displays the results.

Parents’ best-response function is monotonic: the extent to which parents state

---

21 Nudges to parents actually improve math standardized test scores significantly within CP2 (see Appendix C). For second-graders, it is indeed the case that learning improves when parents are nudged, but not when teachers are (independently or jointly). Having said that, the overall effect on dropout rates is driven by CE2 and CM2.

22 One reason for why it is hard to detect effects on learning outcomes are geographic spillovers. Appendix D documents that numeracy standardized test scores of students in the control group significantly improve with the share of students whose parents receive nudges within a 5km radius of their homes.
Figure 3. Parents’ and teachers’ best-response functions

Note: Panel A: “During the last school year, how would your involvement in the education of [CHILD] have changed if [CHILD]’s teacher had asked you to come to school [FREQUENCY] to talk about what [CHILD] was learning?”; “During the last school year, how would your involvement in the education of [CHILD] have changed if [CHILD]’s teacher had been present in school teaching [FREQUENCY]?” Parents could answer on a scale from 1 to 5, where 1 means "decrease a lot" and 5 means "increase a lot". The effort measure standardizes their answers to each question.

Panel B: “How would your effort in teaching your classroom last year have changed in case 1 (10) out of 10 PARENTS had showed up in school unannounced [FREQUENCY] to talk about what their child is learning?” Teachers could answer on a scale from 1 to 5, where 1 means "decrease a lot" and 5 means "increase a lot". The effort measure standardizes their answers to each question.

that they would engage in their children’s education is increasing in the frequency at which teachers request them to come to school. Incidentally, from the way parents specify they would change effort in response to teachers’ requests, we can infer that the average teacher never demands that parents show up in school. We can also infer that parents expect teachers to be in class teaching more often than once a week, although not everyday; if that were the case, parents would increase their effort substantially, as much as if teachers would request them to come to school everyday.

When it comes to teachers’ best-response function, while it has a similar shape to that of parents when 1 out of 10 parents show up in school more frequently, it bears an inverted U-shaped relationship with the frequency of monitoring when it comes to 10 out of 10 parents showing up in school: teachers’ planned effort level dramatically decreases when all parents show up every day – down to roughly the same level as if parents only showed up once a year.

The second way in which we tackle that question is by estimating heterogeneous treatment effects of nudging parents independently with respect to the share of parents who show up in schools regularly (according to students).

Figure 4 shows that nudging parents decreases student dropouts only within classrooms where 85% of parents or less show up in school regularly. Beyond that share, at the high-end of the distribution, the intervention actually increases
IV. Concluding remarks

This paper documents first-hand that monitoring teachers’ effort has non-monotonic effects. Our interventions to monitor teachers either directly or through community engagement massively decreased school dropouts, but combining both interventions failed to improve educational outcomes. Too high monitoring intensity backfires because it deteriorates high-effort teachers’ motivation, leading to backlash. Lower attendance by high-effort teachers triggers a chain reaction: parents become more pessimistic, corporal punishment no longer decreases, and students drop out of school to the same extent as if no intervention were in place.
Evidence from teachers’ best-response function and heterogeneous treatment effects of nudges to parents suggests that results are not a special feature of our setup, but rather a general property of team production in children’s education.

Our findings match qualitative evidence about frustrated interactions between teachers and parents, documented in the education literature (e.g., [Wolf and Suntheimer 2020a; Chikutuma 2017]). They are also consistent with Frey’s 1993 conjecture, and with laboratory findings (e.g., Falk and Kosfeld 2006; Dickinson and Villeval 2008). In Dickinson and Villeval (2008), whether monitoring backfires above a certain threshold depends on whether reciprocity can be inferred by the agent from the monitoring level set by the principal, and on the nature of the relationship between principals and agents. They find that repeated interactions increase the likelihood of backlash, and particularly so when employers’ payoffs depend primarily on workers’ output – which is precisely the nature of the interactions between parents and teachers: children’s education is regarded as primarily a function of teachers’ inputs, and interactions are non-anonymous and repeated over a long horizon.

Educational policies could benefit from incorporating insights based on our findings, particularly in developing countries – where education is a critical component of investments in children’s human capital under stark informational asymmetries, with far-ranging implications for growth and inequality (Glewwe and Kremer 2006). First, while monitoring interventions in education typically focus on teachers (as in Duflo, Hanna and Ryan 2012), we find that those might trigger backlash if teachers consider they are already putting in high effort. As such, effectively intervening on teachers might require carefully eliciting their beliefs in order to inform targeting. Second, when it comes to participatory programs, while they often encourage caregivers to monitor teachers quite intensely (e.g., Banerjee et al. 2010), adapting that approach in face of non-monotonicity – either by targeting a smaller share of parents in the community, or by moderating the extent to which parents are encouraged to approach teachers – would likely increase their effectiveness.

While our results shed light on puzzles in the literature, they also raise a number of additional questions. For instance, could complementary interventions raise the monitoring levels that high-effort teachers find acceptable? Do financial incentives (in particular, pay for performance) prevent or magnify burnout and backlash in response to intense monitoring? Those are promising avenues for future research.
REFERENCES


Online Appendix A - Balance and Selective Attrition Tests

This Appendix presents balance and selective attrition tests. Tables A1 and A2 display sample means for school characteristics (panel A), caregiver characteristics (panel B), child characteristics (panel C) and teacher characteristics (panel D) across each of the four treatment arms (Control, Parents only, Teachers only, and Both), along with p-values for ANOVA tests of joint significance for differences in means across treatments arms. We find few differences, small in magnitude and likely to emerge by chance given the large number of covariates; most importantly, F-tests for joint significance rule out systematically imbalances between each treatment and the control group at the 10% significance level.

Next, table A3 tests for selective attrition across treatments arms. As both interventions led to higher parental engagement, assignment to either treatment significantly increased the odds that we successfully tracked students at end line. We find that students whose parents received messages either independently or jointly with teachers were over 4 p.p. more likely to be successfully tracked by surveyors at end line, relative to an 88% success rate in the control group. For this reason, in Appendix G we compute bounds for treatment effects on outcomes based on our end-line survey with students.
Table A1—Descriptive statistics and balance tests

<table>
<thead>
<tr>
<th>Panel A: School characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Regions</td>
<td>2.466</td>
<td>2.436</td>
<td>2.703</td>
<td>2.036</td>
<td>F-test for nudges to teachers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students enrolled</td>
<td>258.44</td>
<td>281.83</td>
<td>283.48</td>
<td>273.35</td>
<td>F-test for nudges to parents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of teachers</td>
<td>6.20</td>
<td>5.57</td>
<td>5.72</td>
<td>6.19</td>
<td>Table A1—Descriptive statistics and balance tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability of learning material</td>
<td>2.56</td>
<td>2.63</td>
<td>2.78</td>
<td>2.75</td>
<td>ANOVA test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly teacher absenteeism</td>
<td>1.64</td>
<td>1.35</td>
<td>1.32</td>
<td>1.35</td>
<td>Both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of students absent per day</td>
<td>7.12</td>
<td>11.22</td>
<td>11.16</td>
<td>8.38</td>
<td>Both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test for nudges to parents</td>
<td>0.889</td>
<td></td>
<td></td>
<td></td>
<td>Both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test for nudges to teachers</td>
<td>0.194</td>
<td></td>
<td></td>
<td></td>
<td>Both</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Caregiver characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female share</td>
<td>0.42</td>
<td>0.37</td>
<td>0.45</td>
<td>0.40</td>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>40.61</td>
<td>40.15</td>
<td>39.75</td>
<td>40.91</td>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>1.14</td>
<td>1.14</td>
<td>1.09</td>
<td>1.15</td>
<td>P-values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. parental engagement</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>-0.06</td>
<td>both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindset with respect to children</td>
<td>8.17</td>
<td>8.24</td>
<td>8.44</td>
<td>8.38</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Failure mindset</td>
<td>7.20</td>
<td>7.42</td>
<td>7.79</td>
<td>7.39</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. parental beliefs (numeracy)</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.03</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. parental beliefs (literacy)</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.06</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations about future school performance</td>
<td>7.96</td>
<td>8.00</td>
<td>7.96</td>
<td>8.00</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child attendance</td>
<td>0.61</td>
<td>0.61</td>
<td>0.57</td>
<td>0.65</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness-to-pay for intervention</td>
<td>23.27</td>
<td>22.09</td>
<td>20.66</td>
<td>19.68</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental health</td>
<td>47.04</td>
<td>45.42</td>
<td>45.88</td>
<td>45.39</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. physical punishment</td>
<td>0.00</td>
<td>0.05</td>
<td>0.07</td>
<td>-0.09</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believe necessary to physically punish a child</td>
<td>0.22</td>
<td>0.24</td>
<td>0.26</td>
<td>0.25</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor activities: Work in cacao plantation</td>
<td>0.20</td>
<td>0.12</td>
<td>0.18</td>
<td>0.12</td>
<td>Only</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: P-values computed with standard errors clustered at the classroom level.
Table A2—Descriptive statistics and balance tests

<table>
<thead>
<tr>
<th>Panel C: Child characteristics</th>
<th>Sub-sample means</th>
<th>ANOVA test</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Parents Only</td>
<td>Teachers Only</td>
</tr>
<tr>
<td>Literacy</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Numeracy</td>
<td>0.49</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Social-emotional</td>
<td>0.65</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>Working memory</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>Attention</td>
<td>0.73</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td>Female share</td>
<td>0.47</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>Impulsivity scale</td>
<td>15.58</td>
<td>14.92</td>
<td>15.73</td>
</tr>
<tr>
<td>Labor activities: Work in cacao plantation</td>
<td>0.42</td>
<td>0.37</td>
<td>0.39</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>7.90</td>
<td>7.85</td>
<td>7.91</td>
</tr>
<tr>
<td>Mindset</td>
<td>9.09</td>
<td>8.92</td>
<td>9.21</td>
</tr>
<tr>
<td>Teacher attendance</td>
<td>0.89</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>Std. parental engagement</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td>Std. student effort</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>F-test for nudges to parents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test for nudges to teachers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel D: Teacher characteristics

<table>
<thead>
<tr>
<th></th>
<th>Sub-sample means</th>
<th>ANOVA test</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Parents Only</td>
<td>Teachers Only</td>
</tr>
<tr>
<td>Age</td>
<td>41.13</td>
<td>38.54</td>
<td>38.42</td>
</tr>
<tr>
<td>Education</td>
<td>3.88</td>
<td>4.26</td>
<td>4.18</td>
</tr>
<tr>
<td>Language skills</td>
<td>1.35</td>
<td>1.13</td>
<td>1.08</td>
</tr>
<tr>
<td>Teacher self-reported attendance</td>
<td>0.33</td>
<td>0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Believe other teachers attendance</td>
<td>1.18</td>
<td>1.17</td>
<td>1.12</td>
</tr>
<tr>
<td>Arrive late</td>
<td>1.08</td>
<td>0.82</td>
<td>0.73</td>
</tr>
<tr>
<td>Leave early</td>
<td>0.14</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>Std. motivation</td>
<td>0.00</td>
<td>0.17</td>
<td>-0.09</td>
</tr>
<tr>
<td>Std. failure mindset</td>
<td>-0.06</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Std. job satisfaction</td>
<td>0.08</td>
<td>0.10</td>
<td>-0.09</td>
</tr>
<tr>
<td>F-test for nudges to parents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test for nudges to teachers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: P-values computed with standard errors clustered at the classroom level.
Table A3—Selective attrition test

<table>
<thead>
<tr>
<th></th>
<th>Tracked at end line</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Parents</td>
<td>0.0417*</td>
</tr>
<tr>
<td></td>
<td>(0.0230)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>(0.0225)</td>
</tr>
<tr>
<td>Both</td>
<td>0.0415*</td>
</tr>
<tr>
<td></td>
<td>(0.0227)</td>
</tr>
<tr>
<td>Control group mean</td>
<td>0.880</td>
</tr>
<tr>
<td>P-value Parents=Teachers=Both</td>
<td>0.658</td>
</tr>
</tbody>
</table>

Observations 2,475
R-squared 0.0035

Note: OLS regression with dependent variable equal to 1 if the participant surveyed at baseline was also present at end line, and 0 otherwise. Standard error clustered at the classroom level. Significance levels are denoted by *** p<0.01, ** p<0.05, * p<0.1.
This Appendix presents the pre-analysis plan in full in subsection B.B1. At the time of pre-registration, we concentrated our hypotheses on issues linked to whether the effects of the intervention evaluated in Bettinger et al. (2020) in Brazil would be transferable to the Sub-Saharan African context. Subsection B.B2 discusses the instances in which our analyses deviate from what had been pre-specified, followed by subsection B.B3 which presents the pre-analysis plan for the extra follow-up wave motivated by our analysis based on the end line data.

B1. Pre-analysis plan
Are Parenting Interventions Transferable Across Settings?
Evaluating Key Constraints in Sub-Saharan Africa

Pre-analysis Plan

While evidence from Brazil suggests that SMS messages to nudge parents’ engagement in their children’s education have large effects on educational outcomes, such an intervention might not work as intended in poorer settings, for at least two reasons: parents have a much higher likelihood of being illiterate—such that text-based interventions may fail to induce behavior change—and teachers have less formal training and a much higher probability of being absent from schools—such that learning outcomes may not improve even if student’s attendance increases. This paper investigates whether those reasons are critical constraints for transferring the intervention across settings, in the context of Ivory Coast, by randomly assigning whether parents receive nudges over text or voice messages, and by cross-randomizing nudges across parents and teachers.

I. Introduction

Both youth and adult literacy rates in Côte d’Ivoire are low, estimated at 53% and 44%, respectively (UNESCO Institute for Statistics, n.d.), with large disparities for marginalized groups, including females and poor children. With the goal of improving learning and literacy through universal education, Ivorian authorities are employing various strategies as part of reforms to meet this goal. One approach to promoting education and learning is to increase parental awareness and engagement in their child’s education, known to be key predictors of children’s academic achievement (see Castro et al., 2015 for a meta-analysis). To do that, Ivorian authorities plan to test Eduq+, a platform powered by MGov, a Brazilian social impact startup, which allows schools to send messages to parents with information about children’s attendance and grades, and which nudges them with motivating facts and suggested activities to engage them in their children’s school life. The SMS-based version of Eduq+ has been shown to be effective in Brazil, where communication with parents had large impacts on attendance, test scores and grade promotion rates (Cunha et al., 2017). Nevertheless, in contexts with low adult literacy rates, audio-based interventions for caregivers may be more effective than text-based.

Moreover, in the context of Côte d’Ivoire, it is not clear that an increase in students’ attendance rate would automatically translate in better learning outcomes. Indeed, a number of previous evaluations of children’s literacy and numeracy skills, show that pupils in Côte d’Ivoire perform relatively worse than in other developing countries, suggesting that they learn relatively little in school (see for example PASEC, 2016). In this setting, a key element behind poor learning outcomes may be teachers, who are often absent
from the classroom in Sub-Saharan Africa (Kremer and Holla, 2009), and poor instructional time use inside the class. Indeed, a study conducted in Ghana and in the Brazilian state of Pernambuco, found that the percentage of time that students were engaged in learning activities was approximately 39% in Ghana, compared to 63% in Brazil (Abadzi, H., 2007).

The program to be tested in Côte d’Ivoire, has two important features that differ from the one previously evaluated in Brazil. First, Eduq+ will send messages not only to parents, but also to teachers, with nudges to encourage higher attendance and suggested activities on how to implement best practices and customize their classes to fit the needs of their students, in order to impact the quality of teaching. Moreover, for parents, we will consider the differential impacts of audio-messaging vs. text-messaging, as well as which characteristics of children and parents can boost the educational gains from the program and why.

Rather than just comparing average treatment effects of text vs. audio messages, we are interested in heterogeneity. Parent’s characteristics – from literacy to language spoken at home to patterns of attention – may be critical for determining which type of intervention is most effective for each family. Moreover, children’s characteristics may also support or constrain to the success of the intervention, such as that parents may be more willing to engage and invest in children with higher baseline academic skills.

Applying this framework will allow us to understand more about how to foster successful home-school relationships, as well as how to tailor behavioral nudge interventions to different children and parents to be most effective, with implications for national educational interventions.

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

1. Does nudging parents via SMS messages improve academic and behavioral learning outcomes?

   - Hypothesis: Eduq+ (audio or text delivery) improves attendance, grades and promotion rates, and decreases school dropout rates.

     ➢ Assuming parental educational investments and school attendance increase, it is not clear that children’s learning outcomes will improve. Given low educational quality, attending school does not necessarily translate improved learning. This question assesses if Eduq+ impacts other dimensions of children’s learning that have potential prospects for longer-term well-being.

2. What delivery mode for SMS nudges to parents has the largest impacts on academic and behavioral learning outcomes: text or audio?

   - Hypothesis: Parent’s characteristics are key moderators for the impacts of the program and hence determine which delivery mode has higher impacts. We posit that text delivery (SMS) will have
larger impacts amongst (a) literate parents who primarily communicate with their children in French; (b) parents with relatively higher visual attention vis-à-vis audio attention; and (c) parents with relatively higher trust in text vis-à-vis audio messages.

3. Do text messages to teachers increase the impacts of audio/text messages to parents?

- Hypothesis: The effects of Eduq+ (audio or text delivery) on attendance, grades and promotion rates, and decreases school dropout rates should be higher for children whose teachers also receive nudges to boost their motivation and support them with information about best practices.

Besides assessing the impact of the intervention on children’s learning outcomes, and determining the most effective delivery mode for nudging parents, we are also interested in testing two additional exploratory hypotheses:

4. How do child characteristics (baseline literacy and numeracy skills, working memory, attention, motivation, social-emotional and self-regulatory skills, time use, self-esteem and mindset) moderate impacts on (i) changes in parents’ education investments, and (ii) children’s learning outcomes?

- Hypothesis: Larger impacts on parents’ investments will occur for children who have higher levels of academic skills at baseline. Children’s non-cognitive skills will moderate impacts on learning outcomes, with children who have higher levels of non-cognitive skills benefiting more from the intervention.

5. Do text messages to teachers improve teachers’ attendance and students’ learning outcomes?

- Hypothesis: text messages to teachers will increase attendance and time spent on learning activities, resulting in an improvement in children’s learning outcomes.
II. Intervention

The intervention, which has been designed and will be implemented by MGov and the Ministry of Education of Côte d’Ivoire, will randomly assign students’ parents and teachers to receive messages by the schools and by MGov. In particular, parents will be assigned to one of three treatment groups:

1. An “audio treatment”, in which parents will receive in the form of audio messages: (i) up to one message per week sent by school to parents about students’ attendance or performance, and (ii) two messages per week sent by MGov, as nudges, where parents will receive suggestions of simple activities that aid in social-emotional development, and which do not demand any curricular knowledge.

2. A text treatment in which the same information will be provided in the form of a message over SMS,

3. A control group, in which parents will not receive any message.

Additionally, teachers will also be assigned to one of two treatment groups:

1. A text treatment in which they will receive weekly messages over SMS with tips on activities to do with students and way to customize their classes to increase children’s learning,

2. A control condition, in which they will not receive any message from the platform.
III. Experimental Design

The intervention will be evaluated through a school-randomized control trial with 100 schools in 2 regions in Côte d’Ivoire. In order to minimize spillovers, randomization will be done at the school level, which will be assigned to one of six treatment groups, as shown in table 1.

Table 1: Randomization strategy

<table>
<thead>
<tr>
<th>Eduq+ Teachers</th>
<th>Eduq+ Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (arm A)</td>
</tr>
<tr>
<td>No (arm D)</td>
<td>25 schools</td>
</tr>
<tr>
<td></td>
<td>Students: 5,000 (survey data for 625)</td>
</tr>
<tr>
<td></td>
<td>Teachers: 50</td>
</tr>
<tr>
<td>Yes (arm E)</td>
<td>25 schools</td>
</tr>
<tr>
<td></td>
<td>Students: 5,000 (survey data for 625)</td>
</tr>
<tr>
<td></td>
<td>Teachers: 50</td>
</tr>
</tbody>
</table>

We will use administrative data on test scores (if made available), attendance, and drop-out rates, as well as primary data collected through surveys with parents, teachers and head teachers, and direct assessments with students. Specifically, data collection will occur through: (i) direct child assessments administered in schools, with 25 randomly chosen students per school, (ii) surveys with teachers and head teachers administered in school, and (iii) directly administered surveys with caregivers in their home. This will occur at two time-points: baseline (October 2018 for parents, when assessors will already be going to communities for program enrolment, and October 2018 for children and teachers, during the first month of school), and follow up (June-July 2019, at the end of the school year). Finally, we will also administer a head-teacher survey at baseline, to assess school characteristics that may serve as barriers or supports for the successful implementation of the program.

IV. Measures

The questionnaire administered to parents will collect the following data:

- socio-demographic characteristics such as gender, relationship with the child, age, marital status, years as primary caretaker of the child, language spoken at home and education,
• Language of instruction skills to assess parents’ ability to comprehend the content of the text and audio messages,
• Trust in audio vs. text messages,
• Parent’s involvement in their child’s education,
• Parent self-reported mindset with respect to children (growth or fixed) and failure mindset,
• Parent’s expectations, aspirations for their child’s education, and beliefs on the child’s school performance and attendance,
• Audio and visual working memory, measured through a simple picture-span task,
• Audio and visual attention, measured through a simple stroop task,
• Self-reported hypothetical willingness to pay to receive weekly messages about their child’s school life,
• Parent’s self-reported mental health,
• Parent’s self-reported discipline practices,
• Child's engagement in labor activities (as reported by the parent).

The questionnaire administered to children will collect the following information:

• Socio-demographic characteristics, such as gender and age,
• Numeracy skills, measured through a test developed to assess children’s mathematics skills, which includes exercises from the PASEC (programme d'analyse des systèmes éducatifs de la Confemen), the EGMA (early grade mathematics assessment), and the IDELA (International Development & Early Learning Assessment),
• Literacy skills, measured through a test developed to assess children’s French skills, based on EGRA (early grade reading assessment) and Tulane University Survey of Child Cocoa Labour,
• Visual working memory, measured through a simple picture-span task,
• Visual attention, measured through a simple stroop task,
• Child self-reported social-emotional skills, directly assessed through items used in IDELA,
• Child self-reported impulsivity scale,
• Child self-reported labor activities, which will be measured through the questionnaire developed by Dr. Kaja Jasinska and has already been used in Côte d'Ivoire,
• Child self-reported motivation,
• Child self-reported self-esteem,
• Child self-reported mindset (growth or fixed),
• Parents’ involvement in child’s education as reported by the child,
• Child self-reported time use,
• Teacher’s attendance as reported by the child.

The questionnaire administered to teachers will collect the following data:

• Socio-demographic characteristics, such as gender, age and closeness to the school,
• Teacher language of instruction skills to assess their French reading comprehension skills,
• Teacher trust in text messages,
• Teacher self-reported education and training,
• Teacher self-reported motivation,
• Teacher self-reported attendance,
• Teacher self-reported mindset (growth or fixed) and failure mindset,
• Teacher self-reported job satisfaction.

Finally, the questionnaire administered to head-teachers will collect the following data:

• School characteristics such as meals provided at school and availability of classrooms,
• Availability of teaching and learning materials,
• Students’ and teachers’ attendance.
V. Outcomes

We will document the effects of the treatments on the following outcomes categories for students enrolled in the second and fourth year of primary school (aged ~7 and ~9 years old respectively), the final grades of the first two primary school cycles:

1. Students’ school attendance, grade retention and drop-out rates measured by administrative records and/or by schools’ inputs at MGov’s platform;
2. Students’ literacy and numeracy skills, measure through direct assessments;
3. Students’ cognitive performance in tasks aimed at assessing working memory and attention:
   - Child visual attention, measured through a stroop-task;
   - Child visual working memory, measured through a picture span task.
4. Students’ socio-emotional and self-regulatory skills:
   - Child self-reported social-emotional skills, directly assessed through items used in IDELA;
   - Child self-reported impulsivity scale.
5. Additional students’ outcomes:
   - Child engagement in labor activities, as reported by the child and by the parent, measured through the questionnaire developed by Dr. Kaja Jasinska and has already been used in Côte d’Ivoire,
   - Child self-reported motivation,
   - Child self-reported self-esteem,
   - Child self-reported mindset (growth or fixed),
   - Child self-reported time use.
6. Parents’ outcomes:
   - Parent self-reported hypothetical willingness to pay to receive weekly messages about their child’s school life;
   - Parent’s involvement in their child’s education, as reported by the parent and the child, in terms of time spent for school-related activities
   - Parent self-reported mindset with respect to children (growth or fixed) and failure mindset,
   - Parent self-reported expectations, aspirations for their child’s education, and beliefs on the child’s school performance and attendance,

1 We focus on those grades such that we can estimate impacts on retention grades, since students cannot be failed on the first and third grades, the beginning of each of those cycles.
- Parent self-reported mental health,
- Parent self-reported discipline practices,

7. Teachers’ outcomes:

- Teacher self-reported mindset and failure mindset;
- Teacher self-reported motivation;
- Teacher self-reported job satisfaction;
- Teacher’s attendance (measure in three ways: (i) self-reported, (ii) reported by the students, and (iii) as reported by administrative records and/or by schools’ inputs at MGov’s platforms).

Additionally, we will assess how students’, their parents and teachers’ baseline characteristics moderate children’s response to the different versions of the program (audio and text), both in terms of outcomes measured by administrative records (grade retention and school attendance), but also their learning outcomes in both academic and behavioral domains.

Since we have several outcome variables for each outcome categories, we will conduct a multiplicity of tests within each category. Estimating separate regressions for each outcome would substantially inflate the probability of false positives above stated significance levels. For this reason, we will build summary measures for each outcome category above, 1 through 7, Following Kling, Liebman and Katz (2007), we will normalize all outcomes to z-scores, and run seemingly unrelated regressions (SUR) to compute effect sizes for each summary measure, within outcome category.
VI. Empirical analysis

1. Does Eduq+ Parents improve academic and behavioral learning outcomes?

Comparing arm A to arms B+C will allow us to measure whether sending messages to parents is effective in improving students’ behavioral and learning outcomes. For each outcome $j$ described above, we will estimate the following regressions indexed by school $m$, individual $i$ and survey $t$:

$$Y_{m,i,t}^j = \beta_0 + \beta_1 Parents_m + u_{m,i,t}$$

Where:

- $Y_{m,i,t}^j$: Outcome variable;
- $Parents_m$: Dummy variable equal to 1 if parents in school $m$ are assigned to receive the intervention (either text or audio), 0 otherwise.

We are interested in testing $\beta_1 = 0$.

2. What delivery mode for Eduq+ has the largest impacts on academic and behavioral learning outcomes: audio or text?

Comparing arm B to arm C will allow us to measure which delivery mode for Eduq+, audio or text, is more effective in improving students’ behavioral and learning outcomes. For each outcome $j$ described above, we will estimate the following regressions:

$$Y_{m,i,t}^j = \beta_0 + \beta_1 Audio_m + \beta_2 Text_m + u_{m,i,t}$$

Where:

- $Y_{m,i,t}^j$: Outcome variable;
- $Audio_m$: Dummy variable equal to 1 if parents in school $m$ are assigned to receive the messages in the form of audio, 0 if the messages are received over SMS.
- $Text_m$: Dummy variable equal to 1 if parents in school $m$ are assigned to receive the messages in the form of audio, 0 if the messages are received over SMS.

We are interested in testing $\beta_1 = \beta_2$. 
3. Do text messages to teachers increase the impacts of audio/text messages to parents?

In order to assess whether sending messages to teachers increases the impact of audio/text message to parents, we will estimate variants of the following regression for each outcome \( j \) described above in the groups ((1), (2), (3) and (4)):

\[
y_{m,tt} = \beta_0 + \beta_1 Parents_{m} + \beta_2 Teacher_{m} + \beta_3 Parents_{m} \times Teacher_{m} + u_{m,tt}
\]

Where:

- \( y_{m,tt} \): Outcome variable;
- \( Teacher_{m} \): Dummy variable equal to 1 if the teacher is assigned to receive the intervention; 0 otherwise.
- \( Parents_{m} \): Dummy variable equal to 1 if parents are assigned to receive the intervention (either text or audio), 0 otherwise.

We are interested in testing \( \beta_3 = 0 \).

4. How do child and parents characteristics moderate impacts on (i) changes in parents’ education investments, and (ii) children’s learning outcomes?

We will explore heterogeneity of treatment according to the following variables:

- Baseline children’s literacy and numeracy skills;
- Baseline children’s attention and working memory skills;
- Baseline motivation (measured for children and teachers);
- Baseline children’s social-emotional and self-regulatory skills;
- Baseline children’s time use;
- Baseline children’s self-esteem;
- Baseline children’s, parents’ and teachers’ mindset (growth or fixed);
- Gender of the child;
- Baseline parents’ involvement in their child’s education;
- Baseline parent’s aspiration and beliefs about their children’s school performance;
- Baseline parent’s trust in audio vs visual messages;
- Baseline parents’ attention and working memory
- Baseline parents’ language skills;
- Baseline teachers’ job satisfaction;
• Baseline teacher self-reported mindset and failure mindset;
• Baseline teacher self-reported motivation;
• Baseline school characteristics.

For each outcome j described above, we will estimate the following regressions:

\[ Y_{m,tt}^j = \beta_0 + \beta_1 Parents_m + \sum \beta_2^j (Parents_m \times Y_{m,tt-1}^j) + \sum \beta_3^j (Parents_m \times X_{m,tt-1}^k) + \beta_4 Y_{m,tt-1}^j + \beta_5 X_{m,tt-1}^k + u_{m,tt} \]

Where:

• \( Y_{m,tt}^j \): Outcome variable;
• \( Parents_m \): Dummy variable equal to 1 if parents are assigned to receive the intervention (either text or audio), 0 otherwise;
• \( Y_{m,tt-1}^j \): Outcome variable at baseline;
• \( X_{m,tt-1}^k \): Participants’ characteristic at baseline (as specified above)

We are interested in applying machine learning techniques (such as random forests) to test \( \beta_2^j = 0 \) and \( \beta_3^k = 0 \) trading-off goodness of fit against over-fitting.

5. Do text messages to teachers improve teachers’ attendance and students’ learning outcomes?

Comparing arm D to arm E will allow us to measure whether sending messages to teachers is more effective in increasing teachers’ attendance, and in improving students’ behavioral and learning outcomes.

For each outcome j described above, we will estimate variants of the following regression:

\[ Y_{m,tt}^j = \beta_0 + \beta_1 Teacher_m + u_{m,tt} \]

Where:

• \( Y_{m,tt}^j \): Outcome variable;
• \( Teacher_m \): Dummy variable equal to 1 if the teacher is assigned to receive the intervention; 0 otherwise.

We are interested in testing \( \beta_1 = 0 \).
B2. Deviations from pre-analysis plan

We had anticipated at the time of pre-registration that SMS nudges to parents might not have improved educational outcomes as in Brazil, for two reasons. First, because teacher absenteeism might be much larger in Ivory Coast than in Brazil (there was no data available on it, and our measure of teachers’ attendance come from surveying students at baseline and end line). For this reason, we cross-randomized nudges to parents and teachers, hypothesizing that nudging teachers might actually improve outcomes to a greater extent, particularly when parents were nudged concurrently.

The second reason for why SMS nudges might not be as effective in Ivory Coast was because illiteracy rates might be much larger than in Brazil. For this reason, we randomized whether nudges to parents were send via text or audio messages (see Appendix E).

While we did not anticipate non-monotonic effects of monitoring teachers, our design was well-suited to study that question. We did not pre-register that we would estimate heterogeneous treatment effects, splitting the sample by median teacher attendance at baseline or interacting treatment assignment with the share of parents who show up in school regularly, but those are used only to shed further light on the mechanism behind the main effects – which are all estimated using analyses registered in the pre-analyis plan.

Next, we pre-specified that we would analyse treatment effects separately by grade. While in the main paper we concentrate our analysis in the whole sample to maximize statistical power, in Appendix C we present results broken down by grade, showing that while effects on dropouts and grade repetition are strongly driven by CM2 students, learning outcomes improve significantly only for CP2. Treatment effects on learning outcomes are analyzed in much greater detail in Wolf and Lichand (2020).

Last, while we had not originally planned to analyze parents’ and teachers’ best-response functions, we extended the pre-analysis plan to register that analyses before the extra follow-up survey data was collected. This extension is presented in the next subsection. Outcomes not analyzed in this paper as presented in Wolf, Lichand and Deambrosi (2020) and Wolf and Lichand (2020).

B3. Pre-analysis plan extension: extra follow-up wave
Are Parenting Interventions Transferable Across Settings? 
Evaluating Key Constraints in Sub-Saharan Africa 

Pre-analysis Plan – Extension for additional follow-up 

I. Introduction 

In this additional follow-up wave to the original study, we intend to survey once more all teachers and a subset of parents from participating schools. The survey instruments intend to help us understand the experimental findings. Mainly, we are interested in shedding light on the negative interaction effects: why is it that sending nudges to both parents and teachers ends up having no positive effects on educational outcomes, when sending nudges to parents only has very positive and significant effects? To answer that question, this follow-up focuses on additional dimensions of parents’ behavior – in particular, whether they show up in school unannounced –, parents’ and teachers’ best response functions within the strategic problem of jointly producing children’s human capital, and their beliefs about the returns of this human capital production function to a variety of inputs. 

Beyond that main goal, we are also interested in collecting student-level data on grade repetition, dropouts and re-enrollment, since this data was only provided by the Government aggregated for each classroom. Last, we are also interested in better understanding the effects of the interventions on child labor more deeply, by eliciting more detailed measures of physical work in agriculture. 

The first step is to characterize planned inputs of each party (i.e. parents and teachers) in response to changes in inputs by the other party, as well as beliefs of each party, in terms of the distribution of expected returns (marginal and cross-derivatives) to different inputs by parents and teachers. The second step is to estimate treatment effects of the original interventions on additional measures of parents’ behaviors, on the distributions of those planned inputs and beliefs, and on those additional school outcomes and additional measures of child labor. As part of that effort, we also elicit incentivized measures of willingness to pay for the intervention under different treatment assignment of the other party, to gain a better understanding of interaction effects above and beyond self-reported non-incentivized beliefs and actions. 

Those goals are summarized in the following research questions: 

1. What is parents’ and teachers’ effort conditional on different input levels of the other party?
1. **Hypothesis:** Teacher effort changes non-monotonically with parents’ inputs: teachers plan to increase effort as a response to higher parental inputs starting from very low levels; however, for at least some inputs, planned effort decreases after inputs are high enough (e.g.: when too many parents show up in school too often).

2. **What are parents’ and teachers’ beliefs about the marginal and cross derivatives of students' learning to marginal increases in inputs by parents and teachers?**

   - **Hypothesis:** Teachers believe that marginal returns to their inputs are low, and that at least some of parents’ and teachers’ inputs exhibit substitutability.

3. **Does the intervention decrease grade repetition and dropouts, and increase re-enrollment for the next school year?**

   - **Hypothesis:** Nudges to parents decrease grade repetition and dropouts, and increase re-enrollment. Nudges to teachers do not affect those outcomes, even when sent in combination with nudges to parents.

4. **Does the intervention decrease child labor in agriculture?**

   - **Hypothesis:** Nudges to parents decrease the intensity of child labor in agriculture, measured by the number of hours worked, the physicality of labor, the extent to which children handle dangerous tools, and the extent to which they sustain injuries, even though they increase the number of occasions in which children help parents in agriculture.

II. **Intervention and experimental design**

There is no additional intervention at this stage. This follow-up wave of data collection, which we have designed and which will be implemented jointly with Innovations for Poverty Action (IPA), intends at capturing the additional research questions mentioned in the previous section.

The survey will cover the same 100 schools that were surveyed in the first phase of the project in 2 regions in Côte d’Ivoire. The sample of subjects for this follow-up consists of 296 randomly selected students’ parents (1 per classroom) and all 296 teachers – 100 from CP2 teachers, 100 from CE2 and 96 from CM2 – that took part in the study. The 200 CP2 and CE2 teachers have already been surveyed in the first phase, while the 96 CM2 teachers will be interviewed for the first time. Similarly, the parents consist of 100 parents whose child are CP2 students, 100 parents whose child are CE2 students and 96 parents whose child are CM2 students. Similar to teachers, the 200 CP2 and CE2 students’ parents have
already been surveyed in the first phase, while the 96 CM2 students' parents will be interviewed for the first time. To ensure maximum efficiency during field work, two replacements for each parent drawn to be surveyed will be randomly selected in case s/he is absent.

A particular sample selection strategy will be applied to parents of CM2 students in control schools for which we have no information. In this case, the procedure that applied to select children whose parents are going to be interviewed is the following: once enumerators get to a school, they get the class list of CM2 students from last year. Using this class list and a random seed that corresponds to the size of the class and the day of the interview, the enumerator will randomly select one child whose parent will be contacted in order to organize the interview. Two substitutes will be selected in case the principal respondent is unreachable.

For this follow-up wave, we will use primary data collected through surveys with parents and teachers. Specifically, data collection will occur through: (i) surveys with teachers administered in school, and (ii) directly administered surveys with parents in their home. This will occur during the month of November.

III. Measures

In this follow-up survey, we will collect information in four main categories of measures and outcomes:

1. We want to understand and assess teachers' and parents' best responses to the other party's inputs, and their beliefs about returns of investments in children's education by each of them, in the case where both are targeted simultaneously by the program as well as in the case where each is targeted independently.

In this category, the questionnaire administered to parents will collect the following data:

- Parents' effort best response function to teachers' inputs (taken as a given parameter). Parents are asked to report what they believe would have happened in terms of their effort in parenting for different frequencies at which teachers engage in various activities. Responses correspond to the magnitude of changes in baseline effort, and are reported on a scale from 1 to 5;
- Parents' views on binding constraints with respect to their child's education, such as teachers' limitation due to lack of parents' education, teachers' limitation due to parents' personal or financial problems, parents' limitation due to teachers' absenteeism and parents' limitation due to teachers' lack of training. These measures draw from and build up on the 2018 World
Development Report\(^1\): Parents need to indicate how much they agree/disagree with each statement, on a scale from 1 to 5;

- Parents’ best guess of different dimensions of their child's school life such as grades, how often they show up in school, how often they help with homework, how often they apply corporal punishment, how often teachers are in the classroom teaching, how often teachers are able to answer their child's questions, how often teachers focus most of their time in the classroom on their child and how often teachers apply corporal punishment. Grades need to be reported on a 0-10 scale (used in Côte d'Ivoire, rounded to the nearest 0.5); frequency is reported on a scale from 1 to 7;

- Parents’ best guess of the impact of marginal changes in the aforementioned dimensions of their child's school life on learning for math and French. Parents are asked to estimate what grades would be (0-10 scale) in each case;

- Parents’ best guess of the impact of joint changes in their own and teachers' inputs on learning for math and French. Parents are asked to estimate what grades would be (0-10 scale) in each case;

- Parents’ best guess of the average impact of the program on their child's grades under different scenarios, such as whether messages are sent to parents only, to teachers only, or to both. Elicitation is incentivized with the possibility of winning airtime credit by selecting the correct bracket for average treatment effects (scale from 1 to 7).

The questionnaire administered to teachers will collect the following data:

- Teachers' effort best response function to parents' inputs (taken as a given parameter). Teachers are asked to report what they believe would have happened in terms of their effort in teaching for different frequencies at which parents engage in various activities. Responses correspond to the magnitude of changes in baseline effort, and are reported on a scale from 1 to 5;

- Teachers' views on binding constraints with respect to children's education, such as teachers' limitation due to parents’ lack of education, teachers' limitation due to parents’ personal or financial problems, parents' limitation due to teachers’ absenteeism and parents' limitation due to teachers' lack of training. Teachers are asked to indicate how much they agree/disagree with each statement, on a scale from 1 to 5;

- Teachers’ best guess of different dimensions of the typical student’s school life, such as grades, how often parents show up in school, how often parents help with homework, how often parents apply corporal punishment, how often they are in the classroom teaching, how often they are

able to answer all questions asked in the classroom, how often they focus most of their time in the classroom students lagging behind and how often they apply corporal punishment. Grades need to be reported on a 0-10 scale (used in Côte d'Ivoire, rounded to the nearest 0.5); frequency is reported on a scale from 1 to 7;

- Teachers’ best guess of the impact of marginal changes in the aforementioned dimensions of their child’s school life on learning for math and French. Teachers are asked to estimate what grades would be (0-10 scale) in each case;
- Teachers’ best guess of the impact of joint changes in their own and parents’ inputs on learning for math and French. Teachers are asked to estimate what grades would be (0-10 scale) in each case;
- Teachers’ best guess of the average impact of the program on grades of the typical student under different scenarios, such as whether messages are sent to parents only, to teachers only, or to both. Elicitation is incentivized with the possibility of winning airtime credit by selecting the correct answer. The scale of responses goes from 1 to 7;
- Teachers’ views on optimal aggregate parental inputs. Teachers are asked to report the optimal frequency (from their perspective) at which different shares of parents should engage in different activities, such as showing up unannounced at school to ask for help interpreting an SMS or to inquire about what their child is learning. Teachers report how often each activity should take place (within a school year), for different shares of parents, on a scale from 1 to 5.

2. We want to understand and assess the demand for the nudge program by parents and teachers.

The questionnaire administered to parents will collect the following data:

- Parents’ willingness to pay for receiving weekly messages about their child's school life through the use of Becker-DeGroot-Marshak (BDM) method, incentivized by the possibility of actually receiving messages or earning airtime credit. Amounts proposed range from 0 to 5000 CFA;
- Parents’ hypothetical willingness to pay such that the teachers or other parents receive weekly messages about their child's school life under different scenarios, also through BDM, except that their decision has no probability of being actually implemented. Amounts proposed range from 0 to 5000 CFA,

The questionnaire administered to teachers will collect the following data:

- Teachers’ willingness to pay for receiving weekly messages about their child's school life through BDM, incentivized by the possibility of actually receiving messages or earning airtime credit. Amounts proposed range from 0 to 5000 CFA;
• Teachers’ hypothetical willingness to pay such that they receive a different version of the program (based on a smartphone app) or such that parents receive weekly messages about their child's school life under different scenarios, also through BDM, except that their decision has no probability of being actually implemented. Amounts proposed range from 0 to 5000 CFA,

3. We want to understand how the intervention impacts the extent to which parents show up in school unannounced, grade repetition, school dropouts, re-enrollment and child labor.

The questionnaire administered to parents will collect the following data:

• Parents’ self-report of the frequency at which they showed up to school unannounced to request help reading messages received in the context of the program, or to ask about any dimension of their child’s school life;
• Parents’ self-report of whether their child dropped out of school, repeated the grade, and re-enrolled;
• Parents’ best guess of their child’s engagement in different labor activities. For some of the questions, parents are asked to estimate the time spent working by their child, on a scale from 1 to 6. For one question, parents are asked when the child works, on a scale from 1 to 8. For the remaining questions, they are asked whether the child is involved or not in different labor activities (yes or no).

The questionnaire administered to teachers will collect the following data:

• Teachers’ best guess of which students dropped out of school, repeated the grade, and re-enrolled. This tries to approximate child-level administrative data measured, as teachers are asked to tick the names corresponding to each question in their class list;
• Teachers’ best guess of which students had their parents request help reading messages received in the context of the program. This tries to approximate child-level administrative data measured, as teachers are asked to tick the names corresponding to each question in their class list;
• Teachers’ best guess of students’ engagement in different labor activities. Some of those questions try to approximate child-level administrative data measured, as teachers are asked to tick the names corresponding to each question in their class list (only for the subset of students who have already been surveyed in previous waves). Teachers are also asked to estimate the time spent working by the typical child, on a scale from 1 to 6. For one question, teachers will be asked when the child works, on a scale from 1 to 8.
4. We want to understand teachers' willingness to invest in their own professional development and in children's education.

The questionnaire administered to teachers will collect the following data:

- Teachers' self-reported professional aspirations, measured on a scale from 1 to 5 corresponding to different career options they can select;
- Teachers' self-reported willingness to spend time on their own professional development, measured on a scale from 1 to 7.

IV. Empirical analysis

Our empirical analysis will consist in two main steps, as discussed previously: a descriptive step, and an impact evaluation step. Both steps are explained in detail below.

A. Descriptive

1. Estimating each party’s best response function to the other party’s inputs, and their beliefs about how returns to investment in children's education vary with inputs of both teachers and parents, has the potential to shed light on whether substitutability is behind negative interaction effects of the intervention. We will also characterize how well parents and teachers accurately anticipate average treatment effects, to gain further understanding about whether their beliefs and best response functions have the potential to provide a rational explanation to observed treatment effects. To complement that descriptive evidence on beliefs and strategic responses, we are interested on parents' and teachers' more general perceptions about complementarity between their inputs.

In the equation below, we are interested in $\beta_1 > 0$ (complementarity) or $\beta_1 < 0$ (substitutability). We also include a quadratic term to allow for potential non-monotonicity ($\beta_2$ and $\beta_1$ with different signs):

$$Y_{ck} = \alpha + \beta_1 X_{cj} + \beta_2 X_{cj}^2 + \epsilon_{ck}$$ (I)

Where:

- $Y_{ck}$: outcome variable for the student (parent k’s child or the typical student within classroom c, in the case of teachers) reported by subject k (parent or teacher) in classroom c;
• $X_{cj}$: Vector of inputs (potentially singleton) by subjects $j$ (either $k$, $-k$ or both) in classroom $c$.

Importantly, because we only elicit perceived returns relative to local perturbations – i.e. given each party’s best guess of the outcome for the (typical) student –, different parts of the distribution are estimated using different sub-samples.

2. Estimating the relationship between teachers' satisfaction and the share of parents engaged in different monitoring activities will give us insights into potential non-monotonicity of monitoring that does not necessarily emerge out of rational strategic responses to changes in inputs.

In the equation below, we are interested in $\beta_1 > 0$ (complementarity) or $\beta_1 < 0$ (substitutability). We also include a quadratic term to allow for potential non-monotonicity ($\beta_2$ and $\beta_1$ with different signs):

$$Y_c = \alpha + \beta_1 S_c + \beta_2 S_c^2 + \epsilon_c$$  \hspace{1cm} (2)

Where:

• $Y_c$: teacher in classroom $c$’s self-reported satisfaction;
• $S_c$: share of parents of children in classroom $c$ engaged in a particular activity.

3. Estimating the accuracy of parents’ and teachers’ prediction of average effects of the program under different scenarios will help potentially rationalize treatment effects with their beliefs and best response functions. We will mainly focus on descriptive statistics, among teachers and parents, of both their beliefs and the distance between those beliefs and the correct answer.

4. Estimating teachers' and parents' willingness to pay will help us understand how much they value the program. Once again, we will focus on summary statistics, among teachers and parents. We are also interested on how WTP changes with the share of parents engaging on different activities, following equation 2.

5. Estimating teachers’ career aspirations will help us understand how much effort they are willing to invest in improving themselves as teachers, which will ultimately impact their quality in the classroom. We will therefore investigate summary statistics such as mean and median amounts, distribution and dispersion.
B. Impact evaluation

We will document the effects of treatment assignment on the following outcomes categories for students enrolled in the second, fourth and sixth year of primary school (CP2, CE2 and CM2), the final grades of the first three primary school cycles:

1. Students dropout rate, grade repetition rate and re-enrollment rate, according to both teachers and parents;
2. Different measures of child labor, according to both parent and the teacher. To deal with family-wise error rates for related hypotheses testing, we will create a summary measure (Kling, Liebman and Katz, 2007) of child labor for each respondent.
3. Parents' outcomes:
   - Self-reported frequency at which parents show up in school unannounced;
   - Parents' willingness to pay for receiving weekly messages about their child's school life;
   - Parents' hypothetical willingness to pay for having teachers receiving the program and for having other parents receiving weekly messages about their child's school life;
   - Parents' prediction accuracy of the average impact of the program on their child's grades.
4. Teachers' outcomes:
   - Teachers' willingness to pay for receiving weekly messages designed to them;
   - Teachers' hypothetical willingness to pay for receiving the smartphone-based program and for having parents receiving weekly messages about their child's school life;
   - Teachers' prediction accuracy of the average impact of the program on the typical student;
   - Teachers' self-reported professional aspirations. To deal with family-wise error rates for related hypotheses testing, we will create a summary measure (Kling, Liebman and Katz, 2007)

Following the main pre-analysis plan, we will study those questions through the following regressions, separately for different grades:

\[ Y_{sci} = \alpha + \beta_1 Parents_s + \beta_2 Teacher_s + \beta_3 Parents_s \times Teacher_s + \varepsilon_{sci} \]  

(3)

Where:

- \( Y_{sci} \): Outcome variable for child i in classroom c at school s;
- \( Parents_s \): Indicator variable equal to 1 if parents in school s are assigned to receive the intervention, and 0 otherwise;
• Teacher: Indicator variable equal to 1 if teachers in school s are assigned to receive the intervention, and 0 otherwise;

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

Additionally, we will assess how parents’ and teachers’ estimates of different dimensions of children can moderate children's response to the program. These dimensions are how often parents show up in school, how often parents help with homework, how often parents apply corporal punishment, how often teachers are in the classroom teaching, how often teachers are able to answer questions asked in the classroom, how often teachers focus most of their time in the classroom on specific students and how often teachers apply corporal punishment. We will also assess how parents' and teachers' views on binding constraints can moderate children's response to the program.

For the moderation analysis, we will estimate:

$$Y_{sci} = \alpha + \beta_1 Treatment_s + \beta_2 X_{sci} + \beta_3 Treatment_s \times X_{sci} + \epsilon_{sci}$$

Where:

• $Y_{sci}$: Outcome variable for child i in classroom c at school s;
• $Treatment_s$: Indicator variable equal to 1 if teachers/parents/both parties in school s are assigned to receive the intervention, and 0 otherwise;
• $X_{sci}$: Vector of covariate (candidates for moderation) for children i in classroom c at school s.
Online Appendix C - Additional Results

This Appendix compiles additional tables and figures to complement the analyses presented in the main text of the paper. Figure C1 displays histograms for the distribution of teacher attendance at baseline and end line (the latter, only for the control group), based on students’ survey responses at each point in time and averaged at the classroom level. While median teacher attendance is quite high in our sample (above 90%), likely because our surveys took place at the very beginning and at the very end of the school year, at baseline the distribution is more spread out, with some teachers absent as much as 50% of the time.

Next, Table C1 presents treatment effects on grade repetition, also based on administrative data shared by the Ministry, available only at the classroom level. While none of the coefficients are precisely estimated, aggregate patterns are similar to those of dropouts: while the effect size of nudging parents independently across all primary cycles is -0.01 (an almost 7% decrease in grade repetition), that of nudging parents and teachers concurrently is less than half that magnitude. Moreover, focusing on the final primary cycle, where treatment effects on grade repetition seem to be concentrated in terms of effect sizes (while still only imprecisely estimated), we find that while nudging parents independently decreases grade repetition by 1/3 and nudging teacher independently decreases grade repetition by about 10% relative to the control group, the effect size of nudging both on grade repetition is actually positive.

Last, we present heterogeneous treatment effects on survey outcomes by primary cycle. Table C2 and Figure C2 restrict attention to CP2 students, while Table C3 and Figure C3 turn to CE2 students. While corporal punishment and student effort respond similarly to the interventions across the two primary cycles, treatment effects on parent monitoring and beliefs, and on students’ numeracy test scores, are concentrated on CP2 students.

23 We did not survey CM2 students, for whom we only have access to administrative data shared by the Ministry of Education.
Figure C1. Histogram of teacher attendance

Note: Histogram of teacher attendance at the classroom level. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students.
Table C1—Treatment effects on student grade repetition rates

<table>
<thead>
<tr>
<th></th>
<th>All grades (1)</th>
<th>CP2 (2)</th>
<th>CE2 (3)</th>
<th>CM2 (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>-0.0104</td>
<td>0.0068</td>
<td>0.0081</td>
<td>-0.0498</td>
</tr>
<tr>
<td></td>
<td>(0.0207)</td>
<td>(0.0267)</td>
<td>(0.0346)</td>
<td>(0.0442)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0148</td>
<td>0.0405</td>
<td>0.0166</td>
<td>-0.0138</td>
</tr>
<tr>
<td></td>
<td>(0.0204)</td>
<td>(0.0264)</td>
<td>(0.0343)</td>
<td>(0.0433)</td>
</tr>
<tr>
<td>Both</td>
<td>-0.0049</td>
<td>0.0255</td>
<td>-0.0535</td>
<td>0.0136</td>
</tr>
<tr>
<td></td>
<td>(0.0201)</td>
<td>(0.0261)</td>
<td>(0.0339)</td>
<td>(0.0424)</td>
</tr>
<tr>
<td>Control group mean</td>
<td>0.1527</td>
<td>0.1358</td>
<td>0.1739</td>
<td>0.1481</td>
</tr>
<tr>
<td>Parents = Both [p-value]</td>
<td>0.7851</td>
<td>0.4819</td>
<td>0.0754</td>
<td>0.1478</td>
</tr>
<tr>
<td>Teachers = Both [p-value]</td>
<td>0.3285</td>
<td>0.5659</td>
<td>0.0416</td>
<td>0.5194</td>
</tr>
<tr>
<td>Observations</td>
<td>296</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0057</td>
<td>0.0293</td>
<td>0.0517</td>
<td>0.0245</td>
</tr>
</tbody>
</table>

Note: Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table C2—Treatment effects on parents’, teachers’ and children’s inputs for CP2

<table>
<thead>
<tr>
<th></th>
<th>Par. monit.</th>
<th>Teach. att.</th>
<th>Par. beliefs</th>
<th>Corp. punish.</th>
<th>Stud. effort</th>
<th>Num. std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>0.2609**</td>
<td>0.0345</td>
<td>0.2103</td>
<td>-0.1096</td>
<td>0.3006***</td>
<td>0.1645*</td>
</tr>
<tr>
<td></td>
<td>(0.1172)</td>
<td>(0.0268)</td>
<td>(0.1576)</td>
<td>(0.1430)</td>
<td>(0.0970)</td>
<td>(0.0949)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.3043**</td>
<td>0.0066</td>
<td>0.0909</td>
<td>-0.3385**</td>
<td>0.3075**</td>
<td>0.0367</td>
</tr>
<tr>
<td></td>
<td>(0.1172)</td>
<td>(0.0206)</td>
<td>(0.1439)</td>
<td>(0.1393)</td>
<td>(0.1482)</td>
<td>(0.0588)</td>
</tr>
<tr>
<td>Both</td>
<td>0.3077***</td>
<td>-0.0128</td>
<td>-0.0979</td>
<td>-0.0276</td>
<td>0.2512**</td>
<td>0.0689</td>
</tr>
<tr>
<td></td>
<td>(0.1137)</td>
<td>(0.0394)</td>
<td>(0.1516)</td>
<td>(0.1304)</td>
<td>(0.1068)</td>
<td>(0.0899)</td>
</tr>
<tr>
<td>Indiv. FE</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>96</td>
<td>2,124</td>
<td>2,087</td>
<td>2,522</td>
<td>2,559</td>
<td>2,559</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0953</td>
<td>0.6451</td>
<td>0.7546</td>
<td>0.6653</td>
<td>0.6456</td>
<td>0.8559</td>
</tr>
</tbody>
</table>

Note: Outcomes are as follows: (1) Parent monitoring; (2) Teacher attendance; (3) Parental beliefs; (4) Corporal punishment; (5) Student effort; (6) Numeracy std. test score. Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Effect sizes are reported for intention-to-treat estimates. For parent monitoring, we use survey responses from the extra follow-up (only CP2). Parent monitoring = 1 when teachers report that the caregiver of the typical child in their classroom last year showed up in school “Sometimes” or “Always”, and 0 otherwise. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Standard errors clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.
Figure C2. Treatment effects on parents’, teachers’ and children’s inputs for CP2

Note: Effect sizes are reported for intention-to-treat estimates, with student fixed-effects for all outcomes, for the sub-sample of CP2 students. Parent monitoring = 1 when teachers report that the caregiver of the typical child in their classroom last year showed up in school “Sometimes” or “Always”, and 0 otherwise. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. Standard errors clustered at the classroom level.
Table C3—Treatment effects on parents’, teachers’ and children’s inputs for CE2

<table>
<thead>
<tr>
<th></th>
<th>Par. monit.</th>
<th>Teach. att.</th>
<th>Par. beliefs</th>
<th>Corp. punish.</th>
<th>Stud. effort</th>
<th>Num. std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>0.0644</td>
<td>0.0115</td>
<td>0.0611</td>
<td>-0.2161</td>
<td>0.4253***</td>
<td>-0.0487</td>
</tr>
<tr>
<td></td>
<td>(0.1193)</td>
<td>(0.0257)</td>
<td>(0.1690)</td>
<td>(0.2090)</td>
<td>(0.1293)</td>
<td>(0.0715)</td>
</tr>
<tr>
<td>Teachers</td>
<td>-0.1131</td>
<td>0.0301</td>
<td>-0.0311</td>
<td>-0.2451*</td>
<td>0.4015***</td>
<td>-0.0852</td>
</tr>
<tr>
<td></td>
<td>(0.1208)</td>
<td>(0.0336)</td>
<td>(0.1360)</td>
<td>(0.1252)</td>
<td>(0.1495)</td>
<td>(0.0771)</td>
</tr>
<tr>
<td>Both</td>
<td>0.0091</td>
<td>-0.0051</td>
<td>-0.0835</td>
<td>-0.2101*</td>
<td>0.4069***</td>
<td>-0.0821</td>
</tr>
<tr>
<td></td>
<td>(0.1179)</td>
<td>(0.0240)</td>
<td>(0.1674)</td>
<td>(0.1172)</td>
<td>(0.1072)</td>
<td>(0.0790)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indiv. FE</td>
<td>90</td>
<td>2,136</td>
<td>1,971</td>
<td>2,371</td>
<td>2,416</td>
<td>2,416</td>
</tr>
<tr>
<td>Obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0247</td>
<td>0.6307</td>
<td>0.7139</td>
<td>0.6627</td>
<td>0.6353</td>
<td>0.8580</td>
</tr>
</tbody>
</table>

Note: Outcomes are as follows: (1) Parent monitoring; (2) Teacher attendance; (3) Parental beliefs; (4) Corporal punishment; (5) Student effort; (6) Numeracy std. test score. Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Effect sizes are reported for intention-to-treat estimates. For parent monitoring, we use survey responses from the extra follow-up (only CE2). Parent monitoring = 1 when teachers report that the caregiver of the typical child in their classroom last year showed up in school “Sometimes” or “Always”, and 0 otherwise. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Standard errors clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.
Figure C3. Treatment effects on parents’, teachers’ and children’s inputs for CE2

Note: Effect sizes are reported for intention-to-treat estimates, with student fixed-effects for all outcomes, for the sub-sample of CE2 students. Parent monitoring = 1 when teachers report that the caregiver of the typical child in their classroom last year showed up in school “Sometimes” or “Always”. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. Standard errors clustered at the classroom level.
Online Appendix D - Imperfect Compliance

This Appendix accounts for imperfect compliance with the interventions by estimating an instrumental variables’ model. We restrict attention to survey outcomes since we do not have individual-level data on student dropouts. In our end-line survey, parents and teachers were asked whether they received text messages from the school over the course of the school year. We use that binary indicator as a measure of compliance within the treatment group, and rely on treatment assignment as an instrument to estimate the ‘clinical’ effects of the interventions – the expected differences between the treatment and control groups if all subjects assigned to the interventions had been reached by the nudges. In computing the measure of compliance, we disregard survey answers by control group participants as schools often reach out to parents for reasons unrelated to our interventions. For that reason, our IV estimates likely provide lower bounds for the true treatment effects on the treated (as we over-estimate the share of compliers in the treatment group).

Due to a variety of reasons – from phone sharing across different household members to network availability issues –, parents’ compliance with our treatment according to that measure was rather low. Only 49% of caregivers assigned to be nudged independently recall receiving text messages from the school, a similar share when it comes to audio messages (46.8%). Our measure of compliance incidentally provides further insight about the non-monotonicity mechanism. While 53% of parents recall receiving audio messages when teachers are concurrently nudged, only 38.9% of parents recall text messages from the school in that situation, suggesting that teachers likely helped parents read and interpret text messages – but to a lesser extent when they were concurrently nudged. In contrast, teachers’ compliance with the intervention is much higher: 82.7% of them recall receiving text messages when nudged independently. That share is even higher when parents are nudged concurrently (reaching 92.3% when parents are nudged via SMS), reinforcing the argument that community monitoring makes direct monitoring even more salient, contributing to non-monotonicity.

Table D1 and Figure D1 report IV estimates. Results yield very similar patterns to those in the main text.\(^{24}\)

\(^{24}\)Parent monitoring is not included in the estimates since it was only collected in the extra follow-up wave by asking teachers about the frequency at which the caregiver of the typical child showed up in school, hence, unmatched to the end-line data on compliance with the interventions.
Table D1—Treatment effects on parents’, teachers’ and children’s inputs (IV)

<table>
<thead>
<tr>
<th></th>
<th>Teach. att.</th>
<th>Par. beliefs</th>
<th>Corp. punish.</th>
<th>Stud. effort</th>
<th>Num. std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>0.0416</td>
<td>0.2004</td>
<td>-0.2898</td>
<td>0.7014***</td>
<td>0.1468</td>
</tr>
<tr>
<td></td>
<td>(0.0307)</td>
<td>(0.1783)</td>
<td>(0.2353)</td>
<td>(0.1480)</td>
<td>(0.1220)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0188</td>
<td>0.0542</td>
<td>-0.3402***</td>
<td>0.4216***</td>
<td>-0.0311</td>
</tr>
<tr>
<td></td>
<td>(0.0211)</td>
<td>(0.1036)</td>
<td>(0.1031)</td>
<td>(0.1214)</td>
<td>(0.0573)</td>
</tr>
<tr>
<td>Both</td>
<td>-0.0395</td>
<td>-0.2281</td>
<td>0.1382</td>
<td>0.2814</td>
<td>0.0310</td>
</tr>
<tr>
<td></td>
<td>(0.0476)</td>
<td>(0.2180)</td>
<td>(0.2248)</td>
<td>(0.2127)</td>
<td>(0.1440)</td>
</tr>
</tbody>
</table>

Indiv. FE | Yes | Yes | Yes | Yes | Yes |
Obs. | 3,128 | 3,058 | 4,266 | 4,274 | 4,274 |
R-squared | -0.0073 | -0.0076 | 0.0039 | 0.0325 | -0.0001 |

Note: Outcomes are the following: (1) Teacher attendance; (2) Parental beliefs; (3) Corporal punishment; (4) Student effort; (5) Numeracy std. test score.

Parents = 1 if parents answered "yes" to the question "In the last months, did you receive regular messages by phone concerning parenting and your child’s education?" at end line, and 0 otherwise; Teachers = 1 if teachers answered "yes" to the question "In the last months, did you receive regular messages by phone by your school or by Movva?" at end line, and 0 otherwise; and Both = 1 in schools where both parents and teachers answered "yes" to these questions, and 0 otherwise. Effect sizes are reported for treatment-on-treated estimates by instrumenting for individuals who actually received the treatment using assignment to treatment as instrument. Standard errors clustered at the classroom level. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this table since we could not construct instruments for this measure collected in the extra follow up. *** p<0.01, ** p<0.05, * p<0.1.
Figure D1. Treatment effects on parents’, teachers’ and children’s inputs (IV)

Note: Effect sizes for treatment effects on the treated, with student fixed-effects across all outcomes. Treatment assignment used as instrumental variable for a binary indicator equal to 1 if treated parents/teachers recall receiving text messages from the school, and 0 otherwise (including all parents and teachers in the control group). Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this figure since we could not construct instruments for this measure collected in the extra follow up. Dark grey bars stand for treatment effect sizes of nudges to parents; light grey bars, those of nudges to teachers; and white bars, those of nudges to both parents and teachers. Standard errors clustered at the classroom level.
This Appendix unbundles nudges to parents, contrasting the impacts of audio messages to those of text messages. Table E1 presents the results. As mentioned in the main text, we find no significant differences across most outcomes. While audio messages have slightly larger effect sizes when it comes to parental beliefs (when parents are nudged independently) and corporal punishment (when teachers are nudged concurrently), we fail to reject the hypothesis that the effects of audio and text nudges to parents are statistically identical across all outcomes.
Table E1—Treatment effects on parents’, teachers’ and children’s inputs (audio vs text messages)

<table>
<thead>
<tr>
<th></th>
<th>Par. Teach.</th>
<th>Par. Corp.</th>
<th>Stud. Num. std.</th>
<th>monit. att. beliefs punish. effort test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0523</td>
<td>0.0162</td>
<td>-0.2826***</td>
<td>0.3550***</td>
</tr>
<tr>
<td></td>
<td>(0.0688)</td>
<td>(0.0199)</td>
<td>(0.0915)</td>
<td>(0.1023)</td>
</tr>
<tr>
<td>Parents (text)</td>
<td>0.0964</td>
<td>0.0228</td>
<td>-0.0100</td>
<td>0.3942***</td>
</tr>
<tr>
<td></td>
<td>(0.0847)</td>
<td>(0.0265)</td>
<td>(0.1294)</td>
<td>(0.1233)</td>
</tr>
<tr>
<td>Parents (audio)</td>
<td>0.1035</td>
<td>0.0234</td>
<td>-0.3212</td>
<td>0.3208***</td>
</tr>
<tr>
<td></td>
<td>(0.0855)</td>
<td>(0.0258)</td>
<td>(0.2008)</td>
<td>(0.0971)</td>
</tr>
<tr>
<td>Both (text)</td>
<td>0.1314</td>
<td>-0.0138</td>
<td>-0.0736</td>
<td>-0.0285</td>
</tr>
<tr>
<td></td>
<td>(0.0823)</td>
<td>(0.0316)</td>
<td>(0.1473)</td>
<td>(0.0980)</td>
</tr>
<tr>
<td>Both (audio)</td>
<td>0.1111</td>
<td>-0.0070</td>
<td>-0.0995</td>
<td>-0.1852**</td>
</tr>
<tr>
<td></td>
<td>(0.0830)</td>
<td>(0.0279)</td>
<td>(0.0812)</td>
<td>(0.1167)</td>
</tr>
<tr>
<td>Both (text) = Both (audio) [p-value]</td>
<td>0.943</td>
<td>0.985</td>
<td>0.268</td>
<td>0.194</td>
</tr>
<tr>
<td>Both (text) = Both (audio) [p-value]</td>
<td>0.832</td>
<td>0.873</td>
<td>0.907</td>
<td>0.353</td>
</tr>
</tbody>
</table>
| **p < 0.01, **p < 0.05, *p < 0.1.**

Note: Outcomes are the following: (1) Parent monitoring; (2) Teacher attendance; (3) Parental beliefs; (4) Corporal punishment; (5) Student effort; (6) Numeracy std. test score.

- Teachers = 1 in schools where only teachers are nudged, and 0 otherwise.
- Parents (text) = 1 in schools where only parents are nudged by text messages, and 0 otherwise.
- Parents (audio) = 1 in schools where only parents are nudged by audio messages, and 0 otherwise.
- Both (text) = 1 in schools where both parents and teachers are nudged, with text messages for parents, and 0 otherwise.
- Both (audio) = 1 in schools where both parents and teachers are nudged, with audio messages for parents, and 0 otherwise.

Ind. FE No Yes Yes Yes Yes Yes
Obs. 280 4,260 4,058 4,893 4,975 4,975
R-squared 0.0138 0.6330 0.7331 0.6613 0.6396 0.9061
This Appendix presents heterogeneous treatment effects by students’ gender\(^{25}\). Table F1 and Figure F1 restrict attention to girls, while Table F2 and Figure F2 focus on boys. With the exception of corporal punishment, which decreases to a greater extent for boys when teachers are nudged independently, the effects of the interventions are very similar for both genders.

Table F1—Treatment effects on parents’, teachers’ and children’s inputs for girls

<table>
<thead>
<tr>
<th></th>
<th>Teach. att.</th>
<th>Par. beliefs</th>
<th>Corp. punish.</th>
<th>Stud. effort</th>
<th>Num. std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>0.0327</td>
<td>0.1482</td>
<td>-0.1196</td>
<td>0.3284***</td>
<td>0.0648</td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.1470)</td>
<td>(0.1396)</td>
<td>(0.1161)</td>
<td>(0.0722)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0208</td>
<td>0.0299</td>
<td>-0.2187*</td>
<td>0.3559***</td>
<td>-0.0517</td>
</tr>
<tr>
<td></td>
<td>(0.0256)</td>
<td>(0.1403)</td>
<td>(0.1159)</td>
<td>(0.1192)</td>
<td>(0.0518)</td>
</tr>
<tr>
<td>Both</td>
<td>-0.0082</td>
<td>-0.0345</td>
<td>-0.1853</td>
<td>0.3696***</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td>(0.0251)</td>
<td>(0.1501)</td>
<td>(0.1192)</td>
<td>(0.0941)</td>
<td>(0.0794)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indiv. FE</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>2,148</td>
<td>2,026</td>
<td>2,462</td>
<td>2,501</td>
<td>2,501</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.6176</td>
<td>0.7075</td>
<td>0.6562</td>
<td>0.6259</td>
<td>0.9055</td>
</tr>
</tbody>
</table>

Note: Outcomes are the following: (1) Teacher attendance; (2) Parental beliefs; (3) Corporal punishment; (4) Student effort; (5) Numeracy std. test score. Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Effect sizes are reported for intention-to-treat estimates. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this table since this measure was collected at end line with the question “In your opinion, how often did the caregiver of the typical child in your classroom last year showed up in school?”, thus without specifying the gender of the child. Standard errors clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.

\(^{25}\)Parent monitoring is not included in the estimates since it was only collected in the extra follow-up wave by asking teachers about the frequency at which the caregiver of the typical child showed up in school, hence, unmatched to data on student’s gender.
Figure F1. Treatment effects on parents’, teachers’ and children’s inputs for girls

Note: Effect sizes are reported for intention-to-treat estimates, with student fixed-effects for all outcomes, restricting attention to the sub-sample of girls. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this figure since this measure was collected at end line with the question “In your opinion, how often did the caregiver of the typical child in your classroom last year showed up in school?”, thus without specifying the gender of the child. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. Standard errors clustered at the classroom level.
Table F2—Treatment effects on parents’, teachers’ and children’s inputs for boys

<table>
<thead>
<tr>
<th></th>
<th>Teach. att.</th>
<th>Par. beliefs</th>
<th>Corp. punish.</th>
<th>Stud. effort</th>
<th>Num. std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents</td>
<td>0.0112</td>
<td>0.1346</td>
<td>-0.2021</td>
<td>0.3961***</td>
<td>0.0647</td>
</tr>
<tr>
<td></td>
<td>(0.0227)</td>
<td>(0.1492)</td>
<td>(0.1553)</td>
<td>(0.1272)</td>
<td>(0.0798)</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0113</td>
<td>0.0593</td>
<td>-0.3479***</td>
<td>0.3548***</td>
<td>0.0140</td>
</tr>
<tr>
<td></td>
<td>(0.0191)</td>
<td>(0.1158)</td>
<td>(0.1182)</td>
<td>(0.1119)</td>
<td>(0.0695)</td>
</tr>
<tr>
<td>Both</td>
<td>-0.0130</td>
<td>-0.1419</td>
<td>-0.0298</td>
<td>0.2760***</td>
<td>-0.0259</td>
</tr>
<tr>
<td></td>
<td>(0.0258)</td>
<td>(0.1482)</td>
<td>(0.1087)</td>
<td>(0.0945)</td>
<td>(0.0685)</td>
</tr>
</tbody>
</table>

Indiv. FE: Yes
Obs.: 2,112 2,032 2,431 2,474 2,474
R-squared: 0.6474 0.7569 0.6634 0.6540 0.9055

Note: Outcomes are the following: (1) Teacher attendance; (2) Parental beliefs; (3) Corporal punishment; (4) Student effort; (5) Numeracy std. test score. Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; and Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Effect sizes are reported for intention-to-treat estimates. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this table since this measure was collected at end line with the question “In your opinion, how often did the caregiver of the typical child in your classroom last year showed up in school?”, thus without specifying the gender of the child. Standard errors clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.
Figure F2. Treatment effects on parents’, teachers’ and children’s inputs for boys

Note: Effect sizes are reported for intention-to-treat estimates, with student fixed-effects for all outcomes, restricting attention to the sub-sample of boys. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics. Parent monitoring is absent from this figure since this measure was collected at end line with the question “In your opinion, how often did the caregiver of the typical child in your classroom last year showed up in school?”, thus without specifying the gender of the child. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. Standard errors clustered at the classroom level.
Online Appendix G - Lee Bounds

As discussed in Appendix A, treatment assignment is systematically correlated with non-response at our end-line survey with students. For this reason, this Appendix computes bounds for treatment effects on outcomes based on end-line survey data. We follow Lee (2009)'s bounding procedure, trimming observations under extreme assumptions for the nature of selection to equalize response rates across treatment arms. Table GI reports the results.
Table G1—Lee Bounds for the treatment effects on parents’, teachers’ and children’s inputs

<table>
<thead>
<tr>
<th></th>
<th>Teacher attendance</th>
<th>Parental beliefs</th>
<th>Corporal punishment</th>
<th>Numeracy std. test score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Parents</td>
<td>0.0003</td>
<td>0.0087</td>
<td>-0.0574</td>
<td>0.0150</td>
</tr>
<tr>
<td>Teachers</td>
<td>0.0008</td>
<td>0.0038</td>
<td>-0.1014</td>
<td>-0.0434</td>
</tr>
<tr>
<td>Both</td>
<td>-0.0129</td>
<td>-0.0061</td>
<td>-0.1494</td>
<td>-0.1058</td>
</tr>
</tbody>
</table>

Note: Effect sizes reported for intention-to-treat estimates. Each row represents a different regression. Bounds computed following Lee (2009)'s procedure.

Parents = 1 in schools where only parents are nudged, and 0 otherwise; Teachers = 1 in schools where only teachers are nudged, and 0 otherwise; Both = 1 in schools where both parents and teachers are nudged, and 0 otherwise. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Parental beliefs are measured with respect to their child’s grade in mathematics.
Online Appendix H - Spillovers

This Appendix documents spillover effects of nudges to parents, leveraging on GPS data collected for all households surveyed at end line. The estimation procedure follows [Miguel and Kremer (2004)], taking advantage of random saturation of treated units across space as a by-product of random assignment. Table H1 estimates the effects of the share of treated units among study participants living at different radius from each student in the control group (within 5km, from 5 to 10km, from 10 to 20km, and further than 20km) on their outcomes.

Results highlight that spillovers are concentrated at close proximity: the higher the share of treated students living within 5km, the higher is teacher attendance reported by control students, the lower is the extent to which their parents rely on corporal punishment, and the better their performance in standardized numeracy tests.
Table H1—Spillover effects of nudges to parents on outcomes of control students

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of treated units &lt; 5km</td>
<td>0.0154</td>
<td>2.2174*</td>
<td>2.5657</td>
<td>-10.8191*</td>
<td>-0.5839</td>
</tr>
<tr>
<td>Share of treated units 5-10km</td>
<td>0.0064</td>
<td>0.7795</td>
<td>3.4536</td>
<td>5.2520</td>
<td>-0.8296</td>
</tr>
<tr>
<td>Share of treated units 10-20km</td>
<td>0.0321</td>
<td>-1.4758</td>
<td>-1.0420</td>
<td>-0.4526</td>
<td>1.3443</td>
</tr>
<tr>
<td>Share of treated units &gt; 20km</td>
<td>0.5861</td>
<td>-0.0150</td>
<td>0.4105</td>
<td>-0.1151</td>
<td>0.4757</td>
</tr>
<tr>
<td>Constant</td>
<td>0.8954***</td>
<td>-0.0047</td>
<td>0.1063***</td>
<td>-0.1761***</td>
<td>-0.0295*</td>
</tr>
</tbody>
</table>

Note: OLS estimates for the sub-sample of control group students. Shares of treated units at different radii computed by dividing the number of treated students (those whose parents are nudged) by the total number of students in our study within the circumference whose centroid lies on each control student's household. Standard errors are clustered at the classroom level. *** p < 0.01, ** p < 0.05, * p < 0.1.