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 only parents are nudged, numeracy and literacy test scores improve by an additional school quarter, and student dropouts decrease by over 50%. In contrast, in schools where both are nudged, there is no effect on either learning outcomes or dropouts – even though the latter also fall by nearly 50% where only teachers are nudged. In those schools, teachers show up less frequently, allocate less time to career development, and target instruction to top students to a greater extent than in schools where only parents are nudged. Monitoring backfires precisely for teachers who were most motivated at baseline, consistent with monitoring intensity eventually crowding out intrinsic motivation. Our results have implications for the design of accountability programs above and beyond education.

JEL: C93, D23, D91, I25

Keywords: Moral Hazard; Monitoring; Accountability; Education

This version: February 19, 2021

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.[1,2] 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 is a key hurdle to human capital formation in developing countries. Across seven Sub-Saharan African countries, Bold et al. (2017) documents 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). Even when teachers are with students, they often spend little time on task; in an audit study in India, 21% of teachers were found just 'minding the classroom' instead of actually teaching (Abadzi, 2007). And even when teachers actually teach, they tend to target instruction to the top students, leaving those who struggle behind (Duflo, Duplas and Kremer, 2011).

Widespread absenteeism and shirking by teachers in low-income countries have been dubbed a 'motivation crisis' (Moon, 2007). Linking under-provision of effort by teachers to their lack of intrinsic motivation reflects a moral hazard problem between parents and teachers. 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 leads incentives to be misaligned: teachers are often not motivated enough (intrinsically or extrinsically) to exert the same level of effort that parents would have optimally set.

In face of that problem, having parents monitor teachers more closely should increase effort and improve learning outcomes. Having said that, evidence from lab experiments suggests that monitoring might actually crowd out intrinsic motivation and effort (Fehr and Gächter, 2000), e.g. if agents perceive it as an indication of distrust (Frey, 1993). Is it also the case outside the lab, when it comes to monitoring teachers? If so, which effect dominates? Does too much monitoring *backfire*, with parents and teachers 'arm-wrestling' over children's education?

We study these questions 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.[3] Nudges to teachers aim at increasing their effort directly,

---

[1]https://www.pinterest.com/pin/108649409733682544/, accessed on January 28, 2021.

[2]Changes in parenting styles over the last decades, in response to several economic and societal changes, might actually have led parents to monitor both students and teachers to a greater extent; see Doepke and Zilibotti (2017).

[3]For many public school students in developing countries, primary caregivers are not the parents but,

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. Both teachers and parents receive nudges twice a week through text messages (SMS) over the course of the whole school year.[4] 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.

While nudges to parents indeed affect the extent to which they monitor teachers, they *do not* affect other dimensions of parental engagement. In schools where parents are nudged, 24% of teachers report that the typical parent shows up at least once a week, compared to only 13% in control schools. Where parents are nudged, students also report a higher likelihood that their parents show up in school to talk to teachers or participate in meetings, relative to control schools, but no systematic differences when it comes to the extent to which parents help students with academic activities, talk to them about school, motivate them to study harder, or resort to corporal punishment. This allows us to interpret the effects of nudges to parents exclusively as those of monitoring teachers indirectly.

Results are as follows. Starting with administrative data on student achievement – only available for school dropouts and grade repetition, aggregated at the classroom level –, we find that only nudges to parents systematically improve educational prospects. Effects are driven by student dropouts. We find that nudges to parents decrease dropouts across all grades by 2.48 p.p. – a 53% reduction relative to control schools. In contrast, in schools where both parents and teachers are nudged, there is *no impact* on dropouts relative to the control group.[5] This is the case even though dropouts also fall by nearly 50% in schools where only teachers are nudged.

Why do nudges dramatically improve the prospects that children remain in school when parents are nudged, but not when the intervention targets both parents and teachers? We start by analyzing treatment effects on teacher effort. Nudging only parents increases teacher attendance, significantly so among those who had below-median attendance at baseline – for whom the intervention decreases absenteeism by nearly 25%. In turn, nudging both parents and teachers actually *increases* teacher absenteeism relative to control schools. The effect is driven by low-attendance teachers, precisely those with more wiggle room to react

---

rather, close relatives (e.g. grandparents). In this paper, we use parents and caregivers interchangeably for ease of exposition.

[4]We also randomized whether parents received Eduq+ over SMS or automated phone calls (Interactive Voice Response units, IVR).

[5]This is not a matter of precision; the effect size of nudging both parents and teachers on student dropouts is only -0.31 p.p.

to higher-intensity monitoring – and, in fact, those whose attendance increases the most when only parents are nudged. Teachers' time use and the extent to which they 'teach to the top' also respond to the interventions in similar ways. In schools where both parents and teachers are nudged, teachers dedicate less time to career development than in those where only parents are nudged. Analyzing the set of students whom we are unable to track at end line (which we use as a proxy for school dropouts at the student level), the share of low-achievers who drop out in those schools is significantly *higher* than where only teachers are nudged – consistent with too intense monitoring inducing teachers to focus on struggling students to a lesser extent.

Next, how do parents and students respond to changes in teacher effort? Using survey data, we find that nudges make parents more pessimistic about students' grades relative to those in control schools, just as in Bettinger et al. (2020), where nudges bring over-optimistic parents' beliefs closer to the ground truth.[6,7] Having said that, their effect size is much larger where teachers are nudged concurrently, consistent with parents becoming *too pessimistic* about their children's school standing in those schools. In effect, where both teachers and parents are nudged, the latter are less likely to say they have engaged in dialogue or incentives as means to discipline children, relative to where only parents are nudged.[8] As such, while student effort increases similarly across treatment conditions, they only learn significantly more than control students in schools where only parents are nudged: students' numeracy and literacy tests scores leap by about 1 school quarter ahead in those schools, but experience *no improvement* where both parents and teachers are nudged.

Strikingly, nudging both parents and teachers backfires for the most intrinsically motivated teachers. Among those with above-median baseline motivation, nudges to both parents and teachers decreases their attendance and increases student dropouts.[9] In contrast, among those with below-median motivation at baseline, nudging both teachers and parents neither decreases teacher attendance nor increases student dropouts. These patterns are consistent with too intense monitoring crowding out teachers' intrinsic motivation.

Is non-monotonicity an artifact of intervening on both parents and teachers concurrently? In particular, sending text messages to both could trigger backlash because it violates social norms about the division of labor (Görges and Nosenzo, 2020). We rule out that concern in three ways. First, we take advantage of an additional experiment embedded in our research design, which randomized some

---

[6]We match observations based on their baseline characteristics to account for selective attrition induced by the interventions; see Section II.C.

[7]Different from Bettinger et al. (2020), Bergman (2021) and Dizon-Ross (2019), we cannot assess the extent to which parents are more or less accurate about their children's school performance because there is no administrative data on grades in Ivory Coast.

[8]Findings are consistent with parents in those schools either communicating less with their children or resorting to physical or psychological violence to a greater extent; see Section IV.B.

[9]That is the case even though nudging only parents or only teachers increases teacher attendance and decreases student dropouts in that case.

parents to nudges via *text* messages and others to nudges via *audio* messages – holding fixed everything else about the intervention, such as its content and the frequency of messaging. Consistent with parents seeking teachers' help to understand text messages from the school (as 41% of parents in our sample never attended school), we document that teachers systematically know when parents are nudged via SMS, but *not* when they are nudged via audio.[10] We show that nudging both teachers and parents indeed backfires to a greater extent when the latter receive text messages; having said that, nudges to both also backfire *even when teachers do not know* that parents are nudged: in schools where teachers are nudged and parents receive audio messages, teacher attendance also decreases, and student dropouts decrease to a much lesser extent – less than 1/4 the effect size of nudges in schools where only parents are nudged.

The second way we document that our results are not an artifact of our experimental design is by estimating how student test scores react to the share of parents who frequently show up in school (reported by students) across different treatment cells. In control schools, the higher the share of parents who monitor teachers, the higher are numeracy and literacy test scores. In contrast, in schools where only teachers are nudged, test scores increase, at first, with the share of parents who show up in school, but eventually stall, *decreasing* as that share exceeds 75% – even though no parent is nudged in those schools.

The third way we show that non-monotonicity is a general property of how teachers respond to monitoring intensity is by eliciting teachers' best-response function: their optimal effort level as a function of parents' monitoring levels. To do so, we survey teachers about their planned effort in the classroom in response to community monitoring, varying the *frequency* at which parents show up in school as well as the *share* of parents who monitor at that level. We find that teachers' planned effort at first increases with community monitoring but, in fact, decreases when *too many* parents show up *too often* in school, consistent with the 'arm-wrestling' mechanism underscored by our experimental results.

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, Gertler, Patrinos and Rubio-Codina (2012) finds that mobilizing parents fails to decrease grade repetition in extremely poor communities, 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

---

[10]Audio messages are delivered through automated calls (IVR), and cannot be played again once the call is over. We document that, when only parents are nudged, treatment effects do not systematically vary by text or audio.

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), and in regions where teachers are likely to be positively selected on intrinsic motivation (such as extremely poor communities).

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 have 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 or substitutability (Greaves et al., 2019; Chang, Cobb-Clark and Salamanca, 2020).[11] This is the first paper to show that teachers' best-response function is actually 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 promoting accountability 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.[12,13] While lab experiments document that monitoring above a certain threshold can decrease work effort (e.g. Falk and Kosfeld, 2006; Dickinson and Villeval, 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 I.A, followed by details about our interventions in subsection I.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 nearly 8 million Ivorian school children, most have to rely

---

[11]Related (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.

[12]Among accountability programs that successfully induced higher effort and improved downstream outcomes, see e.g. Gaduh et al. (2020); Duflo, Hanna and Ryan (2012); Björkman and Svensson (2009) in education, Christensen et al. (forthcoming); Nyqvist, de Walque and Svensson (2017) in health, and Olken (2007) in infrastructure.

[13]Among accountability programs that failed to induce higher effort or improve downstream outcomes, see e.g. Barrera et al. (2020); Raffler, Posner and Parkerson (2019); Wolf et al. (2019); Gertler, Patrinos and Rubio-Codina (2012); Banerjee et al. (2010). In particular, on how monitoring might displace effort to other margins, see e.g. Lichand and Fernandes (2019); Rasul and Rogger (2018); Gerardino, Litschig and Pomeranz (2017); Shi (2008); Yang (2008).

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, 30-40% 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 school progression, 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 classroom at CP2, 46 at CE2 and 40 at CM2. Primary completion rates are low (73.3%), with the majority of children dropping out before completing the three primary cycles.[14] In our sample, average grade repetition rates across all primary grades are over 15%, and yearly dropout rates average 4.7%.

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.[15] 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).[16]

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 directly to parents' and teachers' mobile phones.[17] Nudges are organized in thematic sequences – comprised of four messages –, with two messages delivered each week. Inspired

---

[14]https://data.worldbank.org/indicator/SE.PRM.PRSL.ZS?locations=CI.

[15]https://www.bbc.com/news/world-africa-15681986.

[16]In a companion paper, we study the experiment to study the extent to which child labor and education are really substitutes in this setting (Wolf and Lichand, 2020).

[17]Movva (http://movva.tech) delivers nudges to engage parents in their children's education across Brazil and Ivory Coast, as of January 2021. One of the authors (Guilherme) is Movva's co-founder and chairman.

by READY4K! (York, Loeb and Doss, 2019), 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 monitors whether teachers are complying with suggested activities. Because the program was ran in partnership with the Ministry of Education, it carried an official weight.

Nudges to parents aim at boosting parental engagement. 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. Because parents in developing countries often delegate their children's education to the school (e.g. Wolf, 2020), nudges did not systematically induce parents to engage in their children's school life other than monitoring teachers more frequently: in Section III, we show that, in schools where parents are nudged, both teachers and students report that parents show up in school and attend parent-teacher meetings to a greater extent that in control schools, but find no other systematic differences in the extent to which they help children in academic activities, incentivize them to study harder, engage in dialogue about school, or resort to corporal punishment as a disciplining strategy.

As Section II explains, we randomize treated parents to either nudges via text messages (SMS) or audio messages – delivered through automated phone calls (Interactive Voice Response units; IVR). Other than the delivery mode, the content and the frequency of the nudges is identical across the two groups.

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 selective attrition, 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 H presents the pre-analysis plan in full.

We cross-randomize nudges to teachers and parents across 296 classrooms (2nd,

4th and 6th grades) in 100 Ivorian public schools.[18] Randomization was undertaken at the school-level, to optimize capacity-building of school staff for enrolling parents into the nudge program. 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.[19] Within the group of schools assigned to nudges to parents (concurrently or not with nudges to teachers), half were assigned to have parents receive nudges via text messages, and the other half to audio messages.

<div align="center">

Nudges to Parents

		No	Yes	
			Text	Audio
Nudges to Teachers	No	25 schools 74 classrooms	12 schools 35 classrooms	12 schools 35 classrooms
	Yes	25 schools 74 classrooms	13 schools 39 classrooms	13 schools 39 classrooms

</div>

At each school, enrollment was voluntary: teachers and parents 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 in sampled schools consented to participate. All primary caregivers with phones were also enrolled, either directly, in school meetings to sensitize them about the nudge program, or indirectly, via the school principal. In both cases, they could opt-out of the program at any point, at no cost, via SMS.

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 student dropouts and grade repetition, shared by school districts upon request of the Ministry of Education. These variables were only made available at the classroom level; concretely, we were informed about how many students started the year in CP2, CM2 and CE2 classrooms, how many finished the year in school, and how many progressed to the next cycle. We compute grade repetition from the difference between those who finished the school year enrolled and those who progressed to the next school cycle in each classroom, and compute dropouts

---

[18]4 schools in our sample only covered the first two cycles of primary education.

[19]Numbers are uneven because local partners were particularly interested in the arm at which both parties were nudged.

from the difference between the number of students who started and those who finished the year in each classroom.

Because some students transfer across schools throughout the year, this measure of dropouts is not perfect; in particular, it is negative for 9 out of 296 classrooms, where more students were enrolled by the end of the school year than when it started.[20] For this reason, we also rely on a *proxy* for school dropouts at the student level: we define proxy dropouts equal to 1 if we cannot track a student between our baseline and end-line surveys, and 0 otherwise. 9.4% of students surveyed at baseline could not be tracked at our end-line survey. Since we are unable to track students for a variety of reasons other than dropouts, this is larger than the incidence of student dropouts in the administrative data (4.7%); nevertheless, the latter is highly correlated with our survey proxy, averaged at the classroom level (the slope coefficient is 0.98, with adjusted R^2=0.23; see Figure B1).[21] We use this proxy to validate our classroom-level estimates (see Appendix B), and to investigate the extent to which the interventions induce teachers to target instruction to top students – estimating whether the characteristics of those whom we cannot track at end line vary systematically across treatment cells (see Section IV.B).

The second set of outcomes we have access to is survey data. Baseline data collection took place at the beginning of the school year, in October/2018, and end-line data collection, at the end of it, in June/2019. We elicited school characteristics, measures of teacher attendance and parental engagement (both reported by students), teachers' self-reported motivation and career investments, the extent to which parents show up in school (as reported by students and teachers), parental beliefs about their children's school performance (captured by their best guess of their child's math and French grades), parents' disciplining strategies (reported by students and parents), student self-reported effort, and standardized numeracy and literacy test scores.[22] For test scores, we adapted PASEC's and IDELA's items, after extensive piloting to prevent floor effects.[23] Within each school, we randomly drew 13 CP2 students and 12 CE2 students to be sur-

---

[20]Appendix F shows that results are robust to imputing dropout rates as zero for those classrooms where our computations yield negative dropouts.

[21]We did not ask parents whether their children had abandoned school or stayed in the same cycle at the end-line survey because, at that point, school official decisions on grade promotion had not yet been released. We did try to collect teacher recall data on student-level grade repetition and dropouts in an additional follow-up survey after the new school year had already started. Those outcomes are not analyzed in the paper because, as nudges to teachers affect teacher attendance and instruction targeting (see Section IV.B), the interventions likely induce differential recall across treatment cells.

[22]In companion papers, 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 (Wolf, Lichand and Deambrosi, 2020) and executive functions Finch, Wolf and Lichand (2020). We also document treatment effects on child labor, showcasing that the child labor and education are not necessarily substitutes Wolf and Lichand (2020).

[23]For details on the Programme for the Analysis of Education Systems (PASEC); see https://www.worldbank.org/en/topic/education/publication/the-programme-for-the-analysis-of-education-systems-2014. For details on International Development and Early Learning Assessment (IDELA), see Pisani, Borisova and Dowd (2018).

veyed at each wave. As such, our sample comprises 200 teachers, and 1,285 CP2 students and 1,190 CE2 students surveyed at baseline along with their primary caregivers. We are able to track all teachers, 1,157 CP2 students (90.0%) and 1,086 CE2 students (91.3%) at end line. We assigned replacement households in case the ones drawn could not be tracked by enumerators.

Tables A1 and A2 compile summary statistics of schools', students', parents' and teachers' baseline characteristics. Schools have on average 258 students enrolled across all primary cycles, featuring an average of 6 teachers – each solely responsible for one of the grades within a primary cycle. 58% of primary caregivers in our sample are male; as discussed, 41% never attended school. 53% of students in our sample are boys. The average teacher is 41 years old and has worked as a teacher for almost 15 years, approximately half of this experience acquired at the current school. Appendix A documents that baseline characteristics of schools, teachers, parents and students are balanced across treatment arms.

Figure 1 displays the correlation between variation in teacher attendance between baseline and end line (reported by students at each point in time) and variation in numeracy test scores (Panel A) and literacy test scores (Panel B). While average student performance increases over time for most of the classrooms, numeracy and literacy test scores tend to increase proportionally to increases in teacher attendance. Such correlations are consistent with the idea that teacher effort is a key constraint to human capital production in this setting.

The third set of outcomes we have access to is data from an additional follow-up survey we undertook at the beginning of the following school year, in October/2019. We elicited teachers' future career plans, teacher perceived monitoring intensity (captured by their recall of how frequently the typical caregiver in her classroom showed up in school over the previous school year), and parents' and teachers' best-response functions with respect to the other party's effort levels. We tried to revisit all 200 CP2 and CE2 teachers interviewed in previous waves, and to survey the 96 CM2 teachers for the first time. We successfully surveyed 273 of those (92.2%).[24] We also re-surveyed 1 caregiver per classroom (randomly drawn, given budget constraints), across all 296 classrooms in our sample.[25]

C. Estimation

Throughout the paper, we estimate average treatment effects (ATE) using OLS regressions based on treatment assignment. Because of issues such as phone sharing and network availability, not all subjects received nudges as assigned. Appendix C estimates treatment effects on the treated (ATT), whereby treatment assignment is used as an instrument for whether parents or teachers in the treat-

---

[24]Out of the 23 teachers whom we were unable to survey in this additional follow-up, 1 refused to participate. The others were either on professional training activities in Abidjan, on sick leave, retired or had been transferred to another region.

[25]We again assigned replacement households in case the ones drawn could not be tracked by enumerators.

Figure 1. Variation in teacher attendance and std. test scores between baseline and end line

Panel A: Numeracy test scores
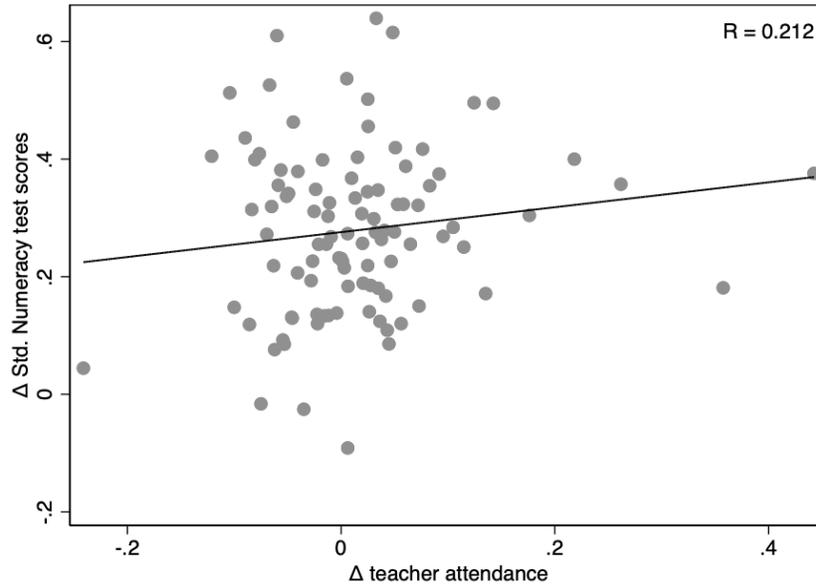
Panel B: Literacy test scores
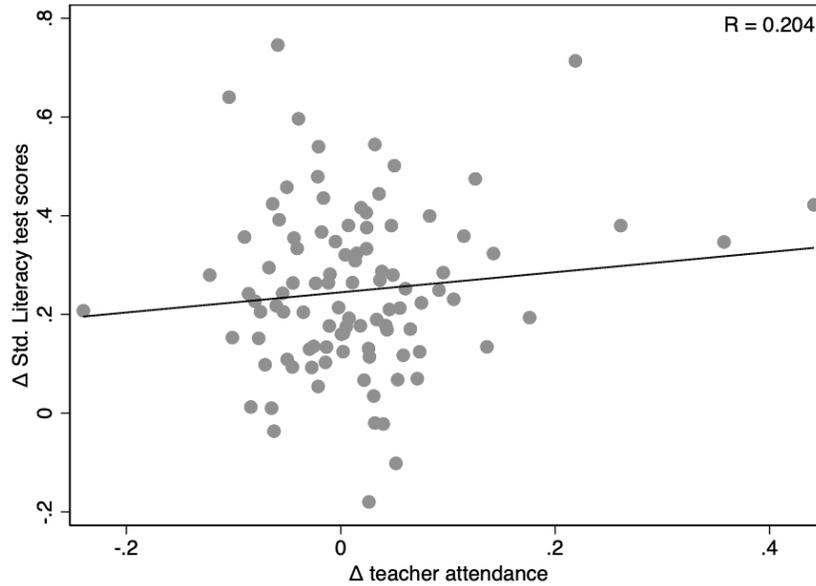
Note: Bin-scatter plots of the variation in students' std. numeracy test scores (Panel A) and literacy test score (Panel B) between baseline and end line as a function of the variation in teacher attendance between baseline and end line. Teacher attendance stands for the share of days on which teachers were in the classroom over the 2 weeks prior to the survey, according to students. Test scores and teacher attendance are averaged at the classroom level.

ment groups acknowledge, at the end-line survey, receiving text messages from the school over the course of the school year.

For administrative outcomes (only available at classroom level), we estimate the following equation:

$$(1) \qquad Y_{sc} = \alpha + \beta_1 Parents_s + \beta_2 Teachers_s + \beta_3 Both_s + \theta_c + \varepsilon_{sc},$$

where Y_{sc} stands for the dropout rate or grade repetition rate in classroom c at school s; $Parents_s = 1$ if school s was assigned to have only parents nudged, and 0 otherwise; $Teachers_s = 1$ if school s was assigned to have only teachers nudged, and 0 otherwise; $Both_s = 1$ if school s was assigned to have both teachers and parents nudged, and 0 otherwise; θ_c stands for grade fixed-effects; and ε_{sc} is an error term. We are interested in testing $\beta_1 \leq 0$, $\beta_2 \leq 0$ and $\beta_3 \geq \beta_1$.

To deal with family-wise error rates from multiple hypotheses testing, we combine student dropouts and grade repetition into a summary measure, following Kling, Liebman and Katz (2007).

For survey outcomes, we estimate an ANCOVA specification, as follows:

(2)

$$Y_{sci,1} = \alpha + \beta_0 Y_{sci,0} + \beta_1 Parents_s + \beta_2 Teachers_s + \beta_3 Both_s + \theta_c + \sum_{k=1}^{K} \gamma_k X_{scik,0} + \varepsilon_{sci},$$

where $Y_{sci,t}$ stands for the survey outcome (e.g. standardized test scores) for subject i in classroom c at school s, measured at wave t; and $X_{sci,0}$ is a vector of individual baseline characteristics. For positive outcomes, such as teacher attendance, parental engagement or student test scores, we are interested in testing $\beta_1 \geq 0$, $\beta_2 \geq 0$ and $\beta_3 \leq \beta_1$ (and the opposite for negative outcomes, such as corporal punishment).

We cluster standard errors at the classroom level, allowing outcomes to be arbitrarily correlated among students (and their parents) under the same teacher. Even though the intervention was randomized at the school level, this makes sense as no teacher in our sample is responsible for more than one classroom, and as nudges are meant to mobilize parents to monitor their children's teacher – not those of other grades.[26] Moreover, evidence from social contact patterns within schools in Sub-Saharan Africa suggests that students interact mostly within classroom, particularly at young age (Stehlé et al., 2011).

For survey outcomes with multiple measurements (e.g. we ask multiple questions about parental engagement and disciplining strategies), we again combine different components into summary measures, following Kling, Liebman and Katz (2007). We analyze treatment effects on specific combinations of outcomes as specified in our pre-analysis plan.

---

[26] Appendix F shows that clustering standard errors at the school level leads to only slightly less precisely estimated coefficients.

Since our interventions affect student dropouts (Section IV), treatment assignment influenced our ability to track students at end line relative to the control group. To deal with selective attrition, we predict the probability of successfully tracking respondents at the end-line survey – from a logistic regression based on baseline characteristics in the control group – and control flexibly for this *propensity score* in all regressions involving survey outcomes.[27,28] This matching procedure ensures that our estimates capture causal treatment effects (rather than compositional changes across treatment cells) under the assumption that selection is independent of potential outcomes once conditioning on observable baseline characteristics. We do not adjust standard errors to account for generated regressors because we restrict attention to average treatment effects (see Abadie and Imbens, 2016).

III. Perceived monitoring intensity

We start by documenting that parents and teachers indeed pay attention to the nudges, and, most importantly, that nudges to parents induce teachers to experience higher monitoring intensity, captured by their perceptions of the extent to which parents show up in school over the course of the experiment.

Table 1 estimates treatment effects on whether parents (columns 1-3) and teachers (columns 4-6) recall receiving messages from the school, breaking down the sample into schools where parents were nudged via text messages (column 2 and 5) and those where parents were nudged via audio messages (columns 3 and 6) to test whether parents' and teachers' recall depends on delivery mode. Last, column (7) estimates treatment effects on whether teachers report that the caregiver of the typical child in her classroom showed up in school at least once a week throughout the year.

The table shows that 52.2% of parents nudged independently, and 48.3% of those nudged concurrently with teachers, recall receiving messages from the school over the course of the experiment. While those figures should be 100% under perfect compliance with treatment assignment, several factors – from phone network availability issues in rural areas to phone sharing and frequent changes in SIM cards among low-income users – could explain why they fall substantially short of it. In any case, these figures are much higher than in control schools, where only 18% of parents recalls receiving messages from the school over the course of the experiment; in both cases, differences are statistically significant at the 1% level. Reassuringly, parents are no more likely to recall receiving messages if teachers were nudged, but not them.

---

[27]Appendix D compiles the details of the matching procedure, and shows that results are robust to matching observations based on their baseline characteristics through a non-parametric, nearest-neighbor algorithm.

[28]In the case of CM2 teachers, whom we only survey in the additional follow-up wave, selective non-response cannot me remedied through matching because we have no data on their baseline characteristics. For this reason, whenever using follow-up data from teacher surveys, we bound treatment effects following Lee (2009); see Appendix E.

Just like parents, teachers pay attention to the nudges: while 24.5% recall messages from the school in the control group, 82.9% of them do so when they are nudged (significant at the 1% level; column 4). Different from parents, teachers recall messages from the school *even when the other party is nudged*. Columns (5) and (6) show that this is driven by text messages: when parents are nudged via audio messages, teachers are no more likely to recall receiving messages than those in control schools (column 6); in contrast, when parents are nudged via SMS, teacher recall raises from 24% to 67% (a difference significant at the 1% level; column 5). In turn, parents' recall of messages from the school when they are nudged independently is statistically identical across text messages (52.2%; column 2) and audio messages (51.6%; column 3). Such interesting pattern in differential recall is consistent with parents seeking teachers' help for interpreting nudges via text messages, but not audio.[29]

Another insightful finding is that, when nudges are delivered via audio, parents' recall is just the same when they are nudged alone or concurrently with teachers (column 3); in contrast, recall is *much lower* in the latter case when nudges are delivered via SMS (from 52.2% to 41%, a significant difference at the 5% level; column 2). This is consistent with teachers being substantially less willing to help parents read and understand text messages when they are also nudged.

Last, while only 12.7% of teachers in control schools state in the additional follow-up survey that the typical parent in her classroom showed up in school at least once a week over the course of our experiment, that figure was 23.9% in schools where parents are nudged, and 25% where both teachers and parents are nudged (significant at the 10% level; column 7).[30] Consistent with the idea that nudges to both lead to perceptions of higher-intensity monitoring, teacher recall rates are even higher when they are nudged concurrently with parents (92.2% when via SMS, and 87.5% via audio; columns 3 and 4).

All in all, teacher recall and perceived monitoring patterns showcase that nudges led parents to show up in school to a greater extent, as intended, and that this affected teachers' perceived monitoring intensity – to a greater extent when both parents and teachers were nudged concurrently.

---

[29]We leverage this difference in Section V.B to estimate whether higher-intensity monitoring backfires even when teachers are not aware that parents are nudged.

[30]Appendix E shows that these results are not sensitive to selective non-response at the additional follow-up survey.

Table 1—Manipulation checks

	Parent recalls receiving messages			Teacher recalls receiving messages			Teacher reports parents show up
	Text+audio	Text only	Audio only	Text+audio	Text only	Audio only	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Parents	0.3459***	0.3454***	0.3399***	0.2320**	0.4204***	0.0260	0.1122
	(0.0330)	(0.0433)	(0.0390)	(0.0963)	(0.1164)	(0.1127)	(0.0663)
Teachers	-0.0270	-0.0276	-0.0289	0.5937***	0.5937***	0.5932***	-0.0001
	(0.0278)	(0.0279)	(0.0275)	(0.0792)	(0.0795)	(0.0793)	(0.0674)
Both	0.3069***	0.2318***	0.3859***	0.6543***	0.6768***	0.6296***	0.1233*
	(0.0383)	(0.0459)	(0.0483)	(0.0746)	(0.0823)	(0.0906)	(0.0652)
Control group mean		0.1764			0.2449		0.1268
Parents = Both [p-value]	0.3437	0.0437	0.4027	0.0000	0.0249	0.0000	0.8666
Teachers = Both [p-value]	0.0000	0.0000	0.0000	0.3574	0.2667	0.6615	0.0675
Baseline controls	Yes	Yes	Yes	No	No	No	No
Predicted propensity scores	Yes	Yes	Yes	No	No	No	No
Observations	2,176	1,644	1,615	196	149	146	280
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.1363	0.1153	0.1665	0.3300	0.3586	0.4002	0.0140

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. In column (1), the outcome variable is a dummy equal to 1 if the caregiver recalled receiving text- or audio-messages on her/his phone in the previous year when we surveyed her/him at end line, and 0 otherwise; in columns (2) and (3), we use the same outcome but restrict our analyses to the sub-sample of parents who received text-messages on their phone in column (2), and to those who received audio-messages on their phone in column (3); in column (4), the outcome variable is a dummy equal to 1 if the teacher recalled receiving text-messages on her/his phone in the previous year when we surveyed her/him at end line, and 0 otherwise; in columns (5) and (6), we use the same outcome but restrict our analyses to the sub-sample of parents who received text-messages on their phone in column (5), and to those who received audio-messages on their phone in column (6); in column (7), the outcome variable is a dummy equal to 1 if the teacher reported that the caregiver of the typical child in his classroom last year showed up in school at least once a week, and 0 if less. In columns (1) to (3), we control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. The sample is restricted to students who were tracked at end line. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level except in column (7). * p<0.1, ** p<0.05, *** p<0.01.

IV.   Impacts of monitoring teachers

This section documents treatment effects on educational outcomes. We start by investigating the impacts of the monitoring interventions on administrative outcomes, in subsection IV.A. Next, subsection IV.B analyzes treatment effects on several dimensions of teacher effort, followed by their effects on parental engagement, student effort and learning, in subsection IV.C. Last, subsection IV.D discusses heterogeneous treatment effects by grade and gender.

A.   *Effects on school dropouts and grade repetition*

We estimate the effects of nudges to parents and teachers on classroom-level measures of school progression. To avoid the issue of multiple comparisons, we compute a summary measure (following Kling, Liebman and Katz, 2007), averaging across dropout rates and grade repetition rates, each standardized with respect to the mean and standard deviation of the control group. In effect, dropouts and grade repetition are often treated as interconnected phenomena in developing countries' primary education, especially at young ages: children who abandon school are often those who are not progressing at the right pace (Kabay, 2016; Brophy, 2006). Table 2 presents the results in column (1), as well as treatment effects on each of its components in columns (2) and (3).

The table shows that nudges to parents systematically impact school progression: the summary measure of dropouts and grade repetition falls by 0.24 standard deviation when parents are nudged, relative to the control group (significant at the 5% level; column 1). Strikingly, nudging both parents and teachers has *no effect* on school progression: its effect size is nearly 1/6 that of nudging parents alone. Columns (2) and (3) highlight that effects are driven by student dropouts. In schools where only parents are nudged, dropouts decrease by 2.48 p.p. – a huge effect size, over 50% the average dropout rate in the control group (significant at the 1% level). In schools where only teachers are nudged, dropouts also decrease (by 2.23 p.p., significant at the 5% level). In contrast, nudging both teachers and parents has a tiny and statistically insignificant effect size (-0.31 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.

Appendix B shows that the very same patterns hold for our student-level proxy of dropouts, although less precisely estimated: nudges to parents cut proxy dropouts nearly in half relative to the control group (from 12% to 8%, significant at the 10% level), while the effect size of nudges to both parents and teachers is almost 20% lower and no longer statistically insignificant. As the Appendix also shows, part of the imprecision comes from the fact that treatment effects are concentrated on intrinsically motivated teachers; among those with above-median motivation at baseline, nudging both parents and teachers backfires in strikingly similar ways across the administrative and proxy measures (see Section V.A).

Table 2—Treatment effects on student dropout and grade repetition

		Components	
	Summary measure	Dropout rate	Grade rep. rate
	(1)	(2)	(3)
Parents	-0.2393**	-0.0248***	-0.0107
	(0.1093)	(0.0095)	(0.0206)
Teachers	-0.1225	-0.0223**	0.0148
	(0.1078)	(0.0094)	(0.0203)
Both	-0.0428	-0.0031	-0.0047
	(0.1064)	(0.0092)	(0.0201)
Control group mean	0.0000	0.0468	0.1527
Parents = Both [p-value]	0.0698	0.0212	0.7682
Teachers = Both [p-value]	0.4546	0.0384	0.3331
Grade Fixed Effects	Yes	Yes	Yes
Observations	296	296	296
R-squared	0.0353	0.0462	0.0151

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. Data are from administrative records. In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are dropout rate (col. 2) and grade repetition rate (col. 3) at the classroom level. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Last, Appendix C estimates average treatment effects on the treated (ATT), using treatment assignment as an instrument for whether treated parents and teachers recall receiving messages from the school. Impacts on proxy dropouts are larger and very precisely estimated in that case: the effect size of nudging only parents is more than 2-fold that of nudging both parents and teachers (the difference significant at the 1% level; column 1 in Table C1).

B. Teacher effort

Why do nudges dramatically improve the prospects that children remain in school when parents are nudged, but not when the intervention targets both parents and teachers? This section documents treatment effects on teacher effort.

We analyze three dimensions of effort: teacher absenteeism , teachers' time use, and the extent to which teachers target instruction to top students in the classroom. All those dimensions should matter for learning and school dropouts. Students cannot learn if teachers are not in the classroom, and we have shown that variation in teacher attendance correlates with improvements in tests scores in our sample (Figure 1); willingness to invest in professional development activities should map into higher-quality instruction; and 'teaching to the top' (characteristic of teaching practices in Sub-Saharan Africa; Duflo, Duplas and Kremer, 2011, e.g.) tends to leave students who struggle behind.

Table 3 displays treatment effects on the first two dimensions. Attendance is captured by the average share of classes in the two weeks prior to each survey for which the teacher was in the classroom, according to students, and time use is captured by the time teachers report they are willing to allocate weekly to professional development activities, collected in the follow-up survey.

Average teacher attendance was high at the time prior to the baseline survey – students state that teachers were absent only about 10% of the time. There is significant heterogeneity across classrooms, however, as in some teacher attendance was as low as 50% at that point (see Figure G1). As such, while the table shows that the interventions do not significantly affect teacher attendance on average (column 1), the full sample masks substantial heterogeneity. In effect, if nudges barely move teacher effort for teachers who were above-median baseline attendance (column 2), nudges to parents systematically increase teacher attendance (significant at the 10% level; column 3). The effect size is relevant, equivalent to a 23.4% decrease in teacher absenteeism relative to the control group, or over 6 additional periods of classes over the course of the school year.[31] Strikingly, nudges to both parents and teachers actually *increase* absenteeism among those teachers, by nearly the same magnitude – even though those low-attendance teachers are

---

[31]Students report on how many classes teachers missed out of 10 periods in the 2 weeks prior to the survey. With morning and afternoon classes every weekday, teachers are absent on average for 1 period every 2 weeks, or 20 periods over the course of the school year. An effect size of -2.5 p.p. stands for a 23.4% lower probability that the teacher misses 1 period of classes over that period, or 4.7 additional periods of classes over the course of the school year.
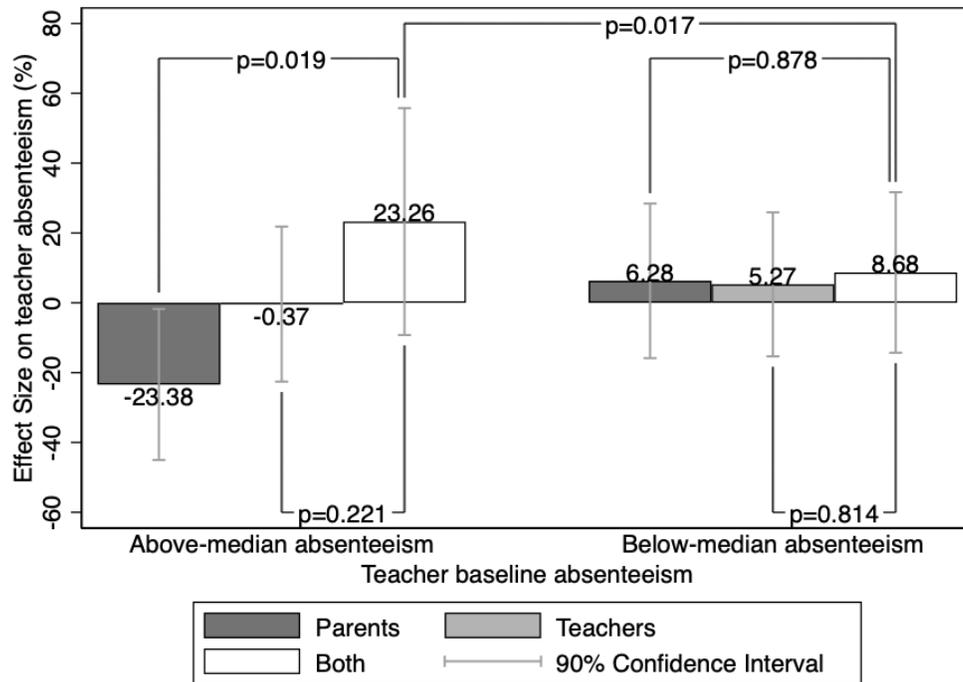
Table 3—Teacher absenteeism and time use

| | Teacher absenteeism | | | |
	All teachers (1)	Above-median (2)	Below-median (3)	Professional development (4)
Parents	-0.0080	-0.0253*	0.0051	-0.2772
	(0.0087)	(0.0141)	(0.0107)	(0.2015)
Teachers	0.0032	-0.0004	0.0042	-0.6219***
	(0.0088)	(0.0145)	(0.0100)	(0.2048)
Both	0.0191	0.0252	0.0070	-0.5885***
	(0.0123)	(0.0212)	(0.0111)	(0.1979)
Control group mean	0.0927	0.1081	0.0804	0.0000
Parents = Both [p-value]	0.0276	0.0191	0.8781	0.1222
Teachers = Both [p-value]	0.2063	0.2209	0.8136	0.8704
Grade Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	No
Predicted propensity scores	Yes	Yes	Yes	No
Observations	1,940	949	991	273
R-squared	0.0137	0.0369	0.0194	0.0659

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. In columns (1) to (3), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. In column (4), the dependent variable is the standardized value of teacher-reported willingness to invest time in professional development weekly, collected in the follow-up. In columns (1) to (3), we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses, clustered at the classroom level in columns (1) to (3). * p<0.1, ** p<0.05, *** p<0.01.

precisely the ones would presumably require (and react the most to) community monitoring. We can reject that this effect size is equal to that of nudges to parents within this sub-sample and in the full sample (both at the 5% level, columns 1 and 3). This is not merely the results on intervening on teachers; nudging them only does not affect teacher attendance across the board.

Figure 2 summarizes the discussion of heterogeneous treatment effects, normalizing effect sizes with respect to the control group mean within each sub-sample to give a sense of the economic significance of treatment effects.

Figure 2. Treatment effects on teacher effort, by baseline absenteeism

Note: Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students. Median of teachers' baseline absenteeism is 9.52%. We use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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.

When it comes to teachers' time use, we find that all nudges decrease the extent to which they are willing to dedicate time to professional development activities (column 4), but effects are much larger when teachers are monitored directly (concurrently with parents or not, significant at the 1% level). Effect sizes are large: while the median teacher states that she is willing to dedicate 2-4 hours a week to professional development activities, nudges to teachers would be nearly enough to bring her down to the previous bracket (1-2 hours a week). Nudging parents and teachers simultaneously does not magnify the negative effects of nudges to teachers in that case.

Next, we turn to the issue of instruction targeting: do nudges lead teachers to focus on top students to a greater extent? To study that question, we analyze whether the characteristics of students who abandon school vary systematically across treatment arms; in particular, we are interested in whether students at the bottom of the class drop out to a greater extent when parents or teachers are nudged, relative to the control group, and also in whether such patterns vary when both parents and teachers are nudged relative to nudging only parents. Because our administrative measure of school dropouts is only available at the classroom level, we rely on our student-level proxy for this analysis.

We start by plotting the distribution of baseline standardized tests scores for students whom we are unable to track at end line in Figure 3, with numeracy on the left-hand side, and literacy on the right-hand side. In each column, the dashed line highlights the bottom decile of baseline test scores (computed for the control group). The figure suggests that nudges to parents or nudges to teachers induce low-achieving students to drop out to a lesser extent; in both columns (particularly for numeracy, on the LHS), there is a lower mass of the histogram to the left of the dashed line relative to control schools. In contrast, in schools where both parents and teachers are nudged, if anything the mass of low-achieving students among dropouts seems to *higher* than in control schools.

Table 4 takes those insights to a regression framework, estimating treatment effects on whether student baseline test scores were in the bottom decile of the distribution. To deal with multiple comparisons, column 1 uses a summary measure, averaging across indicator variables of whether students' numeracy and literacy standardized test scores lie in the bottom decile (p10), while columns 2 and 3 present treatment effects on each component.

The table shows that nudges to parents and nudges to teachers in fact decrease the share of bottom students among those who abandon school relative to the control group (significant at the 10% level in the case of nudges to teachers; column 1), while nudging both parents and teachers does *exactly the opposite*. We can reject that the effects of nudging only teachers and that of nudging both parents and teachers are the same (at the 5% level; column 1). The table also shows that treatment effects on instruction targeting are similar across numeracy (column 2) and literacy (column 3), but particularly large for the former: the share of students in the bottom decile of numeracy at baseline among those who

Figure 3. Instruction targeting: distribution of baseline standardized test scores within students who we are unable to track at end line, by treatment cell
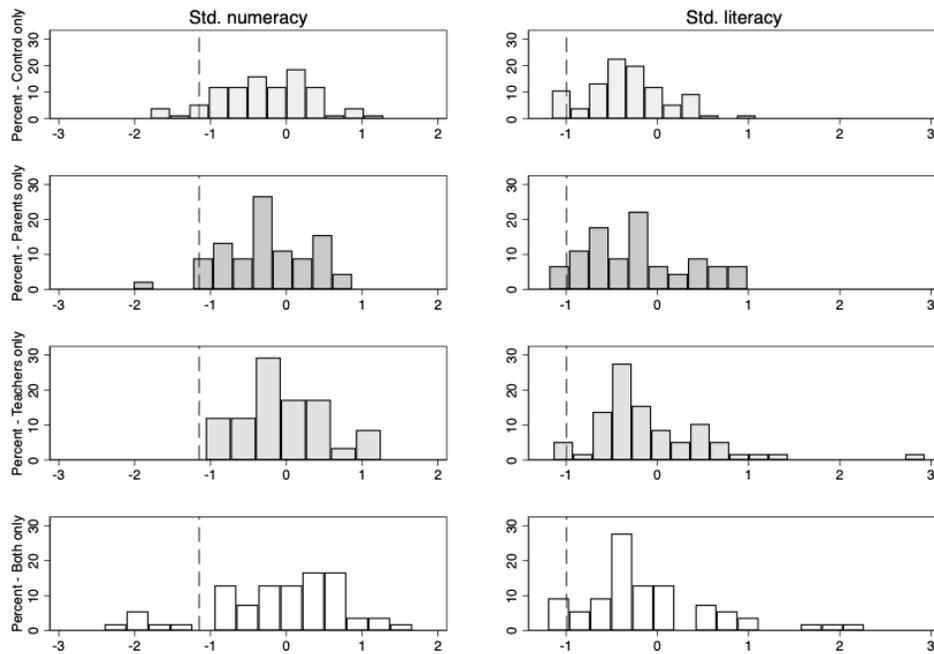
Note: First row for students in control schools; second row for students in schools where parents were nudged independently; third row for students in schools where teachers were nudged independently; and fourth row for students in schools where parents and teachers were nudged concurrently. Sample restricted to students who were not tracked at end line. In both columns, dashed line at the bottom decile of the respective baseline standardized test score in the control group among students whom we were not able to track at end line.

Table 4—Instruction targeting

	Bottom-decile summary measure (1)	Components	
		Numeracy (p10) (2)	Literacy (p10) (3)
Parents	-0.2129	-0.0224	-0.0911
	(0.1493)	(0.0536)	(0.0595)
Teachers	-0.2511*	-0.0872**	-0.0451
	(0.1272)	(0.0367)	(0.0513)
Both	0.0940	0.0353	0.0141
	(0.1872)	(0.0708)	(0.0508)
Control group mean	0.0926	0.0933	0.1067
Parents = Both [p-value]	0.1248	0.4624	0.0578
Teachers = Both [p-value]	0.0430	0.0468	0.2017
Grade Fixed Effects	Yes	Yes	Yes
Observations	232	232	232
R-squared	0.1560	0.0739	0.1240

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. In column (1), the dependent variable averages across the standardized value of outcome variables in columns (2) and (3), which are equal to 1 if the participant had a baseline standardized test score in either numeracy (col. 2) or literacy (col. 3) below the 10th percentile of the control group of the sample considered in each regression, and 0 otherwise. Sample restricted to students who were not track at end line. Standard errors clustered at the classroom level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

abandon school increases from 0.6 p.p. where only teachers are nudged to 12.86 p.p. where nudges target both parents and teachers (a difference significant at the 5% level; column 2). When it comes to literacy, this share is over 10 times higher in schools where both teachers and parents are nudged relative to those where only parents are nudged (a difference significant at the 5% level; column 3).

C. Effects on parental engagement, student effort and learning

Next, we turn to the effects of the interventions on parents' behavior and students' time use and standardized test scores. Because nudges affect school dropouts – and, consequently, our ability to track students and their caregivers at end line –, we control flexibly for each observations' propensity score throughout (see Section II.C).

First, we analyze treatment effects on parental engagement, captured through several questions posed to students about their caregivers. Following our pre-analysis plan, we compute indices of *school participation* (a summary measure of the frequency at which parents (i) participate in school meetings; and (ii) talk to teacher), help with *academic activities* (a summary measure of the frequency at which parents (i) help with homework; and (ii) help organize school material), *motivation* (a summary measure of the frequency at which parents (i) incentivize school attendance; and (ii) incentivize studying), and *dialogue* (a summary measure of the frequency at which parents (i) ask about homework; ask about school grades; and ask about how school was). Table 5 presents the results.

In schools where parents are nudged, students report a higher likelihood that their parents show up in school to talk to teachers or participate in meetings, relative to control schools. Other than that, we find no systematic differences when it comes to the extent to which parents help students with academic activities, motivate them to study harder or talk to them about school. This is different from the effects of nudges on parental engagement documented in Brazil (Bettinger et al., 2020), where the intervention significantly increases the extent to which parents help with academic activities, incentivize children to study harder, engage in dialogue about their children's school life, and aspire that their children eventually makes it to college.

Differences are presumably linked to different social norms; in Sub-Saharan Africa, children's education is considered primarily a responsibility of the school. As a leading example, Wolf (2020) collects qualitative evidence from structured interviews with parents of 0-6 year-old children in Ghana, concluding that "*Teachers felt parents were partly involved, with some parents engaged in both forms of involvement, others who only provided material resources to the school, and others who did neither. Parents, on other hand, mostly described their involvement as paying school fees and providing donations and resources to the school*" (p.3). For our purposes, that provides an empirical opportunity to use nudges to parents as an intervention that affects children's education exclusively through higher-

Table 5—Treatment effects on parental engagement

	School participation	Academic activities	Motivation	Dialogue
	(1)	(2)	(3)	(4)
Parents	0.1128*	-0.0330	-0.0084	-0.0096
	(0.0631)	(0.0627)	(0.0542)	(0.0519)
Teachers	0.0328	-0.0759	0.0213	-0.0010
	(0.0610)	(0.0599)	(0.0560)	(0.0529)
Both	-0.0652	0.0126	-0.0341	0.0102
	(0.0633)	(0.0593)	(0.0566)	(0.0465)
Parents = Both [p-value]	0.0055	0.4831	0.6308	0.6848
Teachers = Both [p-value]	0.1126	0.1525	0.3223	0.8248
Grade Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Predicted propensity scores	Yes	Yes	Yes	Yes
Observations	2,176	2,176	2,176	2,176
R-squared	0.0172	0.0238	0.1255	0.0736

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. In column (1), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents participate in school meetings; and frequency at which parents talk to teacher. In column (2), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents help with homework; frequency at which parents help organize school material. In column (3), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents incentivize school attendance; frequency at which parents incentivize studying. In column (4), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents ask about homework; frequency at which parents ask about school grades; frequency at which parents ask how school was. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. In all columns, we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

intensity monitoring of teachers.

In schools where both parents and teachers are nudged, it is not just that parental engagement does not systematically increase; if anything, school participation actually *decreases* relative to the control group, according to children. In those schools, students feel their parents are less actively involved in their education even though teachers perceive *exactly the opposite* (Section III). We can reject that the coefficients of nudges to parents and nudges to both are equal at the 1% level (column 1). That is not merely an effect of nudging teachers, as its effect size is small but positive (p-value for the difference is 0.113).

Next, we analyze treatment effects on the use of violence as a disciplining strategy, captured through questions posed to caregivers about whether they find appropriate to punish children across difference instances (or had recently done so). Following our pre-analysis plan, we average across indices of discipline *without violence* (a summary measure of whether someone in the household (i) took away privileges, forbade something the child likes, or did not allow the child to leave the house in the past month; (ii) explained why the child's behavior was wrong in the past month; and (iii) gave the child something else to do in the past month), *physical violence* (a summary measure of whether someone in the household (i) shook the child in the past month; (ii) hit or spanked the child with bare hand in the past month; (iii) spanked the child or hit him/her elsewhere on the body with something like a belt, hairbrush, stick or other hard object in the past month; (iv) hit or slapped the child on the face, head or ears in the past month; (v) hit or slapped the child on the hand, arm, or leg in the past month; (vi) hit the child over and over as hard as one could in the past month; and (vii) believes that in order to bring up, raise, or educate a child properly, the child needs to be physically punished) and *psychological violence* (a summary measure of whether someone in the household (i) shouted, yelled at or screamed at the child in the past month; and (ii) called the child dumb, lazy, or another name like that in the past month). Naturally, the *without violence* index enters negatively into the violence summary measure. Table 6 presents the results.

The table shows that nudging either parents or teachers does not systematically affect the extent to which parents resort to violence in disciplining children. Having said that, nudging both parents and teachers *decreases* the extent to which parents resort to non-violent disciplining strategies, relative to schools where either parents or teachers are nudged independently (significant at the 1% and 5% levels, respectively; column 2). As we find no treatment effects of nudges to both on corporal or psychological violence, such decrease in non-violent disciplining strategies in those schools means that either those parents do not disclose that they resort to violence to a greater extent as a disciplining strategy, or that they merely communicate less with their children overall – both of which are adversarial to child development (Wolf and Suntheimer, 2020*b*).

Table 6—Treatment effects on the use of violence as a disciplining strategy

	Summary measure (1)	Without violence (2)	Physical violence (3)	Psychological violence (4)
		Components		
Parents	-0.0065	0.0877*	0.0487	0.0194
	(0.0254)	(0.0492)	(0.0506)	(0.0579)
Teachers	-0.0197	0.0463	0.0087	-0.0216
	(0.0286)	(0.0465)	(0.0611)	(0.0560)
Both	0.0240	-0.0491	0.0056	0.0175
	(0.0281)	(0.0438)	(0.0517)	(0.0582)
Parents = Both [p-value]	0.2469	0.0036	0.3604	0.9726
Teachers = Both [p-value]	0.1390	0.0283	0.9561	0.4605
Grade Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Predicted propensity scores	Yes	Yes	Yes	Yes
Observations	2,176	2,176	2,176	2,176
R-squared	0.0080	0.0168	0.0112	0.0068

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. In column (1), the outcome averages across the outcomes of columns (2) to (4) (with the outcome of col. (1) set to negative). In column (2), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household took away privileges, forbade something the child likes (e.g., watching TV, playing with friends) or did not allow the child to leave the house in the past month; whether someone in the household explained why the child's behavior was wrong in the past month; whether someone in the household gave the child something else to do in the past month. In column (3), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household shook the child in the past month; whether someone in the household hit or spanked the child with bare hand in the past month; whether someone in the household spanked the child or hit him/her elsewhere on the body with something like a belt, hairbrush, stick or other hard object in the past month; whether someone in the household hit or slapped the child on the face, head or ears in the past month; whether someone in the household hit or slapped the child on the hand, arm, or leg in the past month; whether someone in the household beat the child up, that is hit him/her over and over as hard as one could in the past month; whether the caregiver believes that in order to bring up, raise, or educate a child properly, the child needs to be physically punished. In column (4), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household shouted, yelled at or screamed at the child in the past month; whether someone in the household called the child dumb, lazy, or another name like that in the past month. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. In all columns, we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Last, we estimate treatment effects on parents' beliefs, students' time use and standardized test scores. Parents' beliefs are a summary measure of their best estimate of their children's (i) current math GPA (1-10); (ii) current French GPA (1-10); (iii) current overall GPA (1-10); and (iv) future overall GPA (1-10). Students' time use (self-reported) is a summary measure of whether they (i) study at home on a general weekday; (ii) listen to the radio on a general weekday; and (iii) read a book, a newspaper or magazines on a general weekday.[32] When it comes to test scores, as discussed in Section II.B, we adapt PASEC's and IDELA's items to assess learning in math and French at baseline and end line. Table 7 presents the results.

Different from Bettinger et al. (2020), Bergman (2021) and Dizon-Ross (2019), we cannot assess the extent to which parents are more or less accurate about their children's school performance because there is no administrative data on grades in Ivory Coast. Having said that, the table shows that nudges make parents more pessimistic about students' grades relative to those in control schools, consistent with nudges making over-optimistic parents' beliefs closer to the ground truth, as in Bettinger et al. (2020). Effect sizes are larger and more precisely estimated when teachers are nudged, but particularly so in schools where parents and teachers are nudged concurrently (significantly different from nudges to parents only, at the 5% level; column 1). Patterns are consistent with parents becoming *too pessimistic* about their children's school standing in those schools relative to the control group, and could help explain why parents in those schools communicate to children to a lesser extent and are perceived by children as less engaged in their school life than those in other schools.

As a result, even though student effort increases similarly across all treatment cells relative to the control group (imprecisely estimated, in column 2), children learn more only where only parents are nudged. Standardized test scores increase by 0.07 (p=0.109; column 3). This is a large effect size, as if those students had spent approximately an additional quarter in school than those in control schools (benchmarking to the average difference in test scores between the baseline and end line assessments in each subject). Results are driven equally by numeracy (column 4) and literacy (column 5). In turn, the effect size of nudging both parents and teachers is actually negative, and we can reject that it is equal to that of nudging parents alone (at the 5% level; column 3).

---

[32]We include listening to radio as part of the engagement summary measure in face of the evidence 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.

Table 7—Treatment effects on parental beliefs, student effort, and learning

				Components	
	Parents' beliefs	Student effort	Test scores summary measure	Numeracy (std.)	Literacy (std.)
	(1)	(2)	(3)	(4)	(5)
Parents	-0.0654	0.0552	0.0674	0.0698*	0.0649
	(0.0589)	(0.0474)	(0.0418)	(0.0407)	(0.0498)
Teachers	-0.1203**	0.0531	-0.0340	-0.0010	-0.0670
	(0.0532)	(0.0463)	(0.0361)	(0.0376)	(0.0438)
Both	-0.2050***	0.0437	-0.0139	0.0033	-0.0311
	(0.0635)	(0.0415)	(0.0353)	(0.0384)	(0.0406)
Parents = Both [p-value]	0.0346	0.7571	0.0303	0.0962	0.0285
Teachers = Both [p-value]	0.1730	0.7888	0.5124	0.9068	0.3222
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Predicted propensity scores	Yes	Yes	Yes	Yes	Yes
Observations	2,013	2,243	2,243	2,243	2,243
R-squared	0.0695	0.0799	0.5714	0.4829	0.5350

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. In column (1), the outcome variable is constructed from the following individual components asked to caregivers: parental beliefs about performance in mathematics; parental beliefs about performance in French; parental beliefs about overall school performance; parental beliefs about future school performance. In column (2), the outcome variable is a summary measure constructed from the following individual components asked to students, equal to 1 if answer is sometimes or more more often, and 0 otherwise: whether they study at home on a general weekday; whether they listen to the radio on a general weekday; whether they read a book, a newspaper or magazines on a general weekday. In column (3), the summary measure averages across numeracy and literacy standardized test scores. In column (4), the outcome variable is a summary measure constructed from individual components consisting in mathematics exercises completed by students. In column (5), the outcome variable is a summary measure constructed from individual components consisting in French exercises completed by students. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. In all columns, we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. Standard errors in parentheses clustered at the classroom level. * p<0.1, ** p<0.05, *** p<0.01.

D.   *Heterogeneous treatment effects by grade and gender*

Appendix G documents that the patterns for treatment effects on students dropouts 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.

When it comes to gender, Appendix G shows that treatment effects on teacher effort do not vary systematically across males and females. In contrast, there is substantial heterogeneity when it comes to treatment effects by students' gender. Just as in Bettinger et al. (2020), boys benefit to a greater extent when it comes to impacts on test scores. Having said that, it is girls who benefit more in terms of student dropouts (differences are not very precisely estimated as we can only resort to proxy dropouts to analyze student-level differences). Those patterns are consistent with additional hurdles for girls to remain in school beyond academic performance.[33] Incidentally, too intense monitoring does *not* backfire for girls' dropout rates, suggesting that important forces that push girls out of school are concentrated on within-household dynamics above and beyond parental beliefs about girls' educational standing.[34]

V.   Mechanisms

Why does monitoring backfire when both parents and teachers are nudged? This section turns to the mechanisms behind non-monotonic effects of monitoring teachers. Subsection V.A starts by investigating the extent to which accountability triggered by nudges crowds out teachers' intrinsic motivation. Next, subsection V.B documents that too-high monitoring intensity backfires even when teachers do not know that parents are nudged, and even when we do not intervene on the latter. Last, subsection V.C showcases that teachers' optimal effort levels in response to community monitoring is indeed non-monotonic with respect to monitoring intensity.

A.   *Crowding-out of teachers' intrinsic motivation*

We start by estimating heterogeneous treatment effects with respect to teachers' intrinsic motivation, cutting the sample according to whether teachers were above or below the baseline median. Teacher motivation is a summary measure of 16 questions, based on the extent to which teachers agree that they look for new

---

[33]In effect, dropout rates are higher among girls even though they have higher tests scores than boys at baseline; see Wolf and Lichand (2020); UNICEF (2019).

[34]One such particular force is the distribution of household chores; see UNICEF (2019) for a discussion of differences in time use between boys and girls.

ways to teach because (i) they think it is important not to always teach the same way, (ii) they feel it is important to keep up with innovations in teaching, (iii) they think it is fun, or (iv) they want to please parents; on the extent to which teachers agree that they would feel (v) ashamed or (vi) guilty if they would not put effort into teaching; on the extent to which teachers agree that motivations for doing so are (vii) making children feel like they care, (viii) helping people, the joy of (ix) finding unique solutions for various students, or that of (x) sustaining children's attention, (xi) creating connections with people, or (xii) keeping their independence from the principal; and on the extent to which teachers agree that they (xiii) feel proud, (xiv) learn from students, (xv) have parents appreciate their knowledge, or (xvi) like being in touch with children and adolescents when they dedicate time to help students individually.[35]

Figure 4 summarizes the estimates of heterogeneous treatment effects with respect to teacher' baseline motivation, normalizing effect sizes with respect to the control group mean within each sub-sample to give a sense of the economic significance of treatment effects.
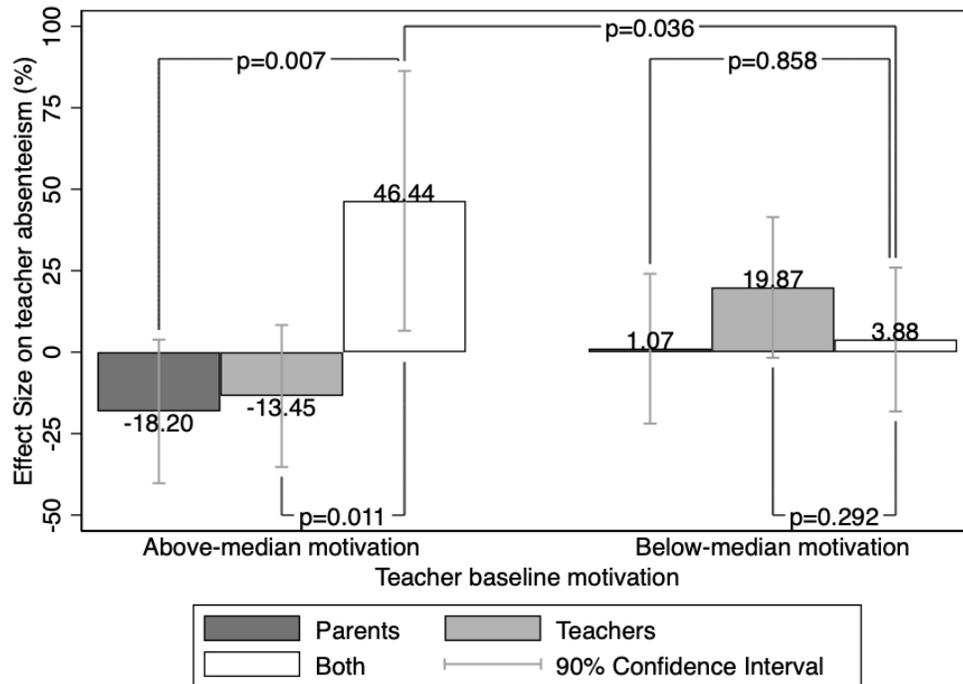
The figure collects two interesting findings. First, nudges do not help ensure teachers are more often in the classroom when they lack intrinsic motivation: no intervention decreases teacher absenteeism for the least motivated at baseline (on the RHS). Second, even though nudges to either parents or teachers decrease teacher absenteeism for those most intrinsically motivated at baseline (although imprecisely estimated), nudging both parents and teachers backfires precisely for those teachers: their absenteeism significantly *increases* in this case relative to the control group (significant at the 10% level, on the LHS). The effect size is huge, two-fold that of nudging both parents and teachers among those most absent at baseline (see Section IV.B), and statistically different from that of nudging parents only (at the 1% level, on the LHS) and from that of nudging both parents and teachers among those least motivated at baseline (at the 5% level, on the RHS).

Estimating heterogeneous treatment effects on student dropouts yields remarkably similar patterns for when nudges backfire. Figure 5 shows that nudges to parents induce sizeable effects on school dropouts across all teachers: dropout rates decrease by 42.5% within those with below-median baseline motivation (on the RHS), and by 41.1% within those most intrinsically motivated at baseline (on the LHS). In turn, nudging both parents and teachers *increases* dropouts amongst those most intrinsically motivated at baseline (p-value for the difference is 0.22, on the LHS). This is the case even though nudging teachers decreases student dropouts by 42.8% in that case (p-value for the difference is 0.23, on the LHS).

Using our survey proxy for student dropouts confirms those patterns – only more precisely estimated with student-level data. Appendix B shows that while

---

[35]What economists think of as intrinsic motivation (as opposed to extrinsic motivation) comprises multiple concepts in the education literature: intrinsic motivation, introjected motivation, identified motivation, and even extrinsic motivation; see (Roth et al., 2007).

Figure 4. Heterogeneous treatment effects of monitoring on teacher effort, by baseline motivation

Note: Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students. Teachers' motivation is constructed following Kling, Liebman and Katz (2007) by standardizing each motivation component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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.

nudging only parents decreases dropout rates by 29.3-37.8% across all teachers, once again nudging both parents and teachers *increases* dropouts amongst those most intrinsically motivated at baseline; the effect sizes are different at the 10% level (on the LHS). In turn, nudging parents and teachers decreases student dropouts by nearly 50% for the teachers least intrinsically motivated at baseline (the difference is significant at the 10% level, on the RHS).

Figure 5. Heterogeneous treatment effects of monitoring on student dropouts, by baseline teacher motivation
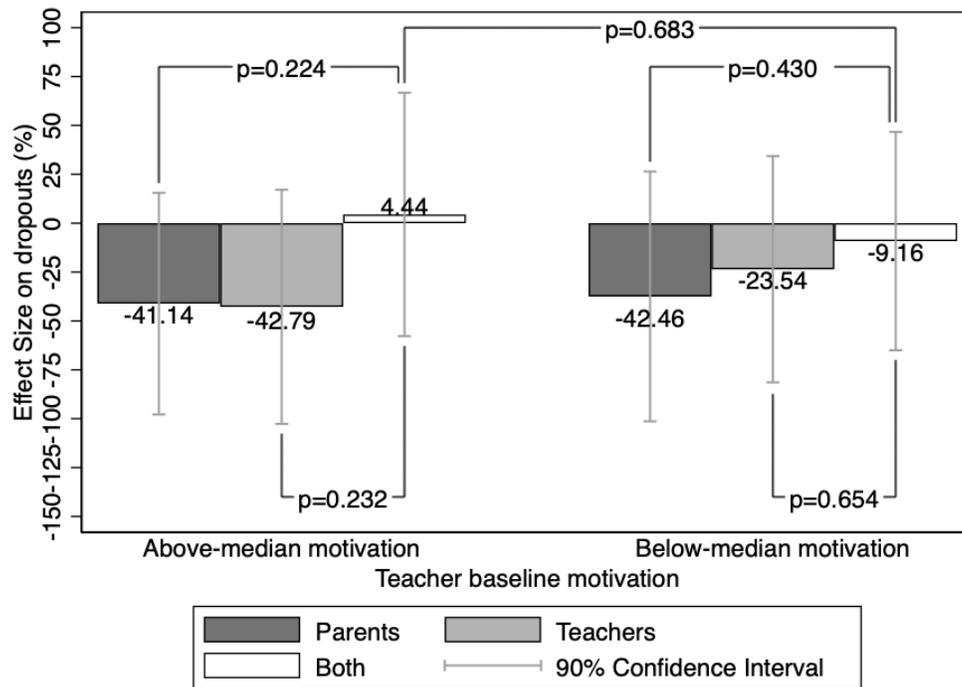
Note: Student dropouts available only at the classroom-level. Dark grey bars stand for treatment effect sizes of nudging parents independently; light grey bars, those of nudging teachers independently; and white bars, those of nudging both concurrently. We control for grade fixed effects. P-values shown in dark brackets.

All in all, results are consistent with external monitoring crowding out teachers' intrinsic motivation.

B. Conditional responses to community monitoring

Is non-monotonicity a general property of how teachers react to monitoring intensity, or is our finding an artifact of intervening on parents and teachers at

the same time? To answer this question, we first take advantage of the fact that teachers were systematically unaware that parents were nudged when messages were delivered via audio (see Section III).

Figure 6 summarizes the estimates of heterogeneous treatment effects with respect to whether parents receive nudged via text or audio messages on teacher absenteeism (Panel A) and student dropouts (Panel B), based on administrative data. Effect sizes are normalized with respect to the control group mean within each sub-sample to give a sense of the economic significance of treatment effects.

Panel A illustrates very interesting patterns, although imprecisely estimated. There is a clear ranking of effect sizes across text and audio: nudging only parents decreases absenteeism, while nudging only teachers has nearly zero effects. Most importantly, nudging both parents and teachers leads to sizeable increases in teacher absenteeism in both cases. Even though text messages magnify the extent to which nudging both parents and teachers backfires (when perceived monitoring intensity is even higher), nudges via audio messages also increase teacher absenteeism when teachers are nudged concurrently even though the latter do not know that parents are nudged as well.

Panel B shows similar patterns for treatment effects on student dropouts. While nudging either teacher or parents significantly decrease dropouts in this case (by roughly 50%, regardless of whether via text or audio), nudging parents and teachers does *not* significantly decrease it in both cases. Even though its effect is even smaller in magnitude (and even slightly positive) when nudges are delivered via text messages, it is also less than 1/4 the effect size of nudging only parents when they are targeted by audio messages – in which case teachers are systematically unaware. In sum, non-monotonic responses to monitoring intensity are not an artifact of teachers' knowledge of nudges to parents.

Next, we estimate heterogeneous treatment effects of community monitoring on learning outcomes, restricting attention to schools where parents were *not nudged*. We estimate a quadratic relationship between standardized test scores and the share of parents who monitor teachers, separately within control schools and within those where only teachers are nudged.[36] To do that, we leverage students' answers to the end-line survey question of how often their parents showed up in school to talk to their teachers over the course of the experiment to compute the classroom-level share of parents who show up sometimes or more often. Accordingly, we average test scores at the classroom level. Table 8 presents the results. Odd columns restrict attention to the control group, and even columns, to schools where only teachers were nudged. Columns (1) and (2) display estimates for the summary measure of standardized test scores, and the other columns to its components (numeracy in columns 3 and 4, and literacy in columns 5 and 6).

The table shows that, while there is no systematic relationship between commu-

---

[36] We use test scores in this analysis instead of dropouts because the latter are a relatively rare outcome; with dropout rates equal to zero across several schools, the statistical power to detect a quadratic relationship is greatly reduced.

Figure 6. Heterogeneous treatment effects of monitoring on teacher effort and student dropouts, for audio vs text messages

Panel A: Teacher effort

Panel B: Student dropouts
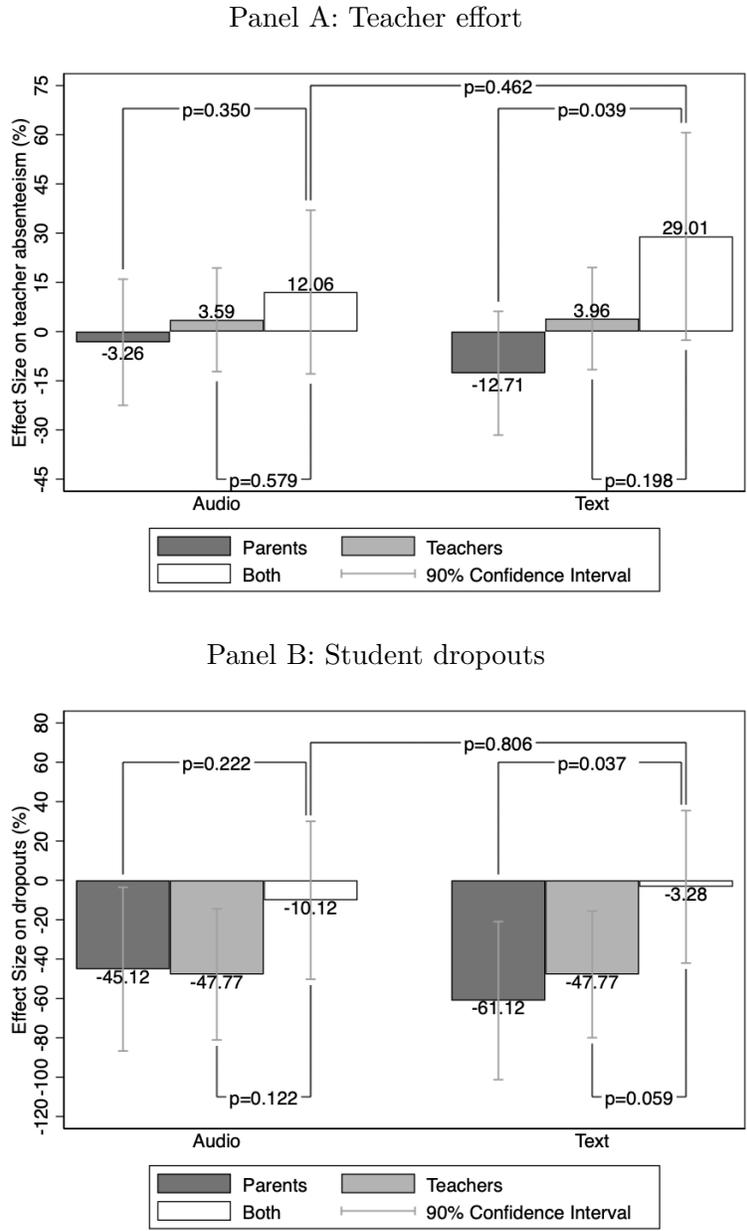
Note: In Panel A, teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students. In Panel B, student dropouts available only at the classroom-level. In Panel A, we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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.

Table 8—Heterogeneous treatment effects of monitoring on end line standardized test scores

	Summary measure		Std. numeracy		Std. literacy	
	Control only	Teacher only	Control only	Teacher only	Control only	Teacher only
	(1)	(2)	(3)	(4)	(5)	(6)
Share of parents who	-1.4937	3.3528*	-2.1532	3.5030*	-0.8342	3.2027
show up in school	(2.2523)	(1.8393)	(1.9974)	(2.0720)	(2.6022)	(2.0260)
(Share of parents who	1.3452	-2.3051*	1.6968	-2.3130	0.9936	-2.2971
show up in school)^2	(1.6978)	(1.2878)	(1.4867)	(1.4315)	(1.9730)	(1.4586)
Equality of square coefficients across samples	0.0756		0.0439		0.1642	
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	50	50	50	50	50	50
R-squared	0.0258	0.0545	0.0278	0.0653	0.0275	0.0425

Note: In columns (1) and (2), the outcome variable averages across numeracy and literacy standardized test scores. In columns (3) and (4), the outcome is a summary measure constructed from individual components consisting in mathematics exercises completed by students. In columns (5) and (6), the outcome is a summary measure constructed from individual components consisting in French exercises completed by students. For each outcome, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. Each outcome is averaged at the classroom level. In columns (1),(3) and (5), the sample is restricted to control schools only. In columns (2), (4) and (6), the sample is restricted to schools where only teachers received nudges. The share of parents who show up in school is the classroom-level average of children who answer "sometimes" or more often to the end-line survey question "How often do your parents or guardians talk to your teachers?". Standard errors clustered at the classroom level. * p<0.1, ** p<0.05, *** p<0.01.

nity monitoring and learning in the control group (column 1), there is a robust inverse U-shaped association between the two in schools where teachers are nudged (significant at the 10% level, column 2). The squared coefficient is significantly different across the two sub-samples (at the 10% level), and the relationship is very similar across numeracy (columns 3 and 4) and literacy (columns 5-6) test scores, even though more precisely estimated for the former.

Figure 7 shows that, if anything, test scores actually increase with community monitoring in the control group (Panel A); in contrast, in schools where teachers are nudged (Panel B), test scores first increase with the share of parents who show up in school, but then stall and eventually *decrease* as the share of parents who show up in school frequently surpasses 75%.

Figure 7. Heterogeneous treatment effects of monitoring on test scores

Panel A: Control group

Panel B: Teachers only
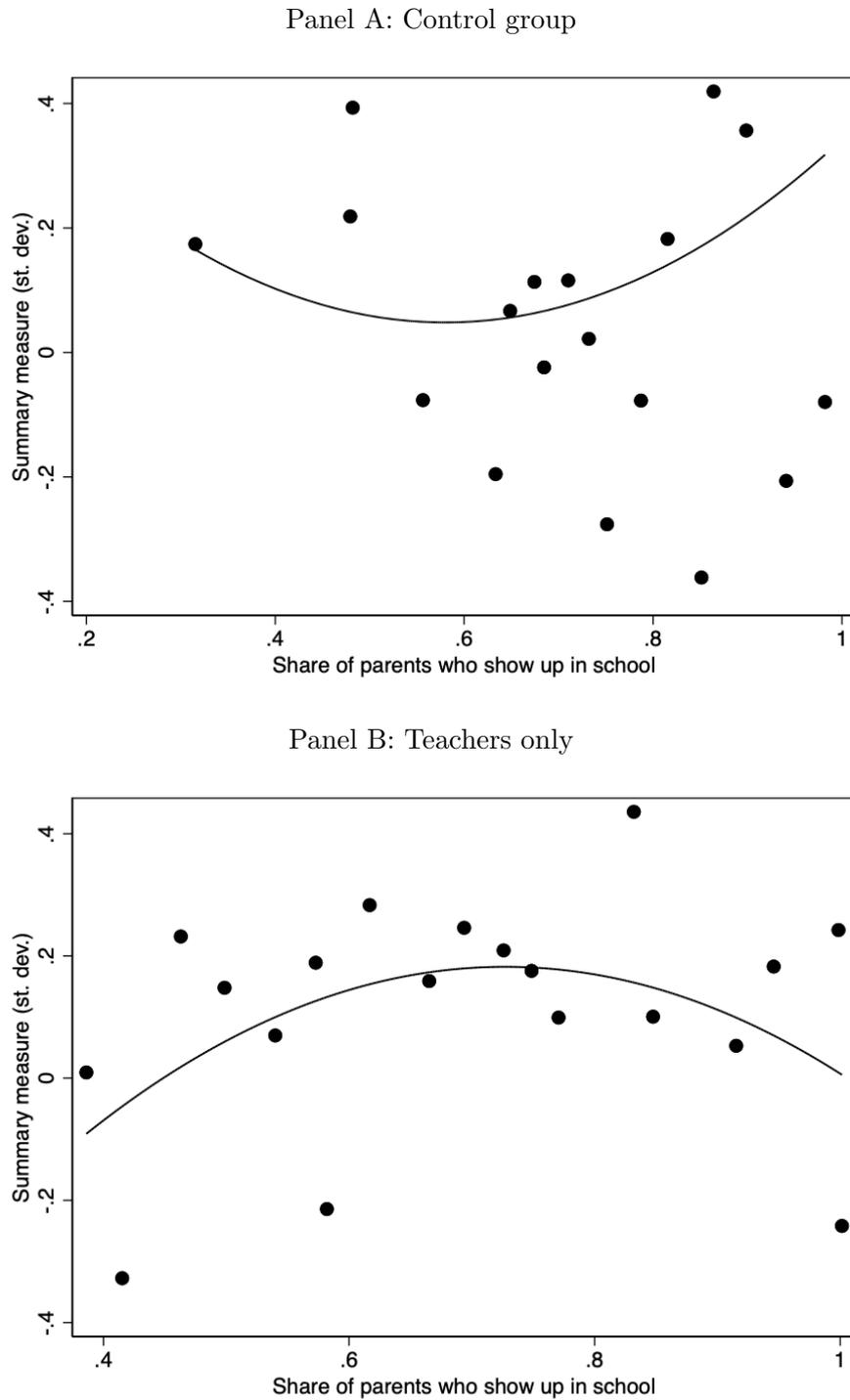
Note: Bin-scatter plots of end line standardized test scores as a function of the share of parents who show up in school regularly, within the sub-samples of control schools (Panel A) and those where only teachers receive nudges (Panel B). The dependent and independent variables are described in Table 8.

C. *Best-response functions*

Last, we elicit parents' and teachers' planned effort levels in response to different effort levels by the other party. For parents, we ask how they would change their effort in response to teacher requests to show up in school, and in response to teacher absenteeism, varying the frequency of teacher requests and teacher attendance in each case. For teachers, we ask how they would change their effort in response to parents showing up in school unannounced to talk about their children, varying both the *frequency* of community monitoring and its *intensity* (for 1 out of 10 parents or 10 out of 10 parents showing up at the specified frequency). Effort changes are measured in a scale from -2 ("decrease a lot") to 2 ("increase a lot"). Figure 8 presents the average responses for parents (Panel A) and teachers (Panel B), standardizing effort levels within each question.

From the way parents specify that they would change effort in response to teachers' requests, in Panel A, 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. Most importantly, the figure shows that parents' best-response function is monotonic: the extent to which parents state that they would engage in their children's education is increasing in the frequency at which teachers request them to come to school, and in the frequency at which teachers are in the classroom teaching.

In turn, Panel B shows that, while teachers' best-response function with respect to the frequency of community monitoring is monotonically increasing when it comes to a small number of parents, it bears an inverted U-shaped relationship with respect to the frequency of monitoring when all parents show up in school at that frequency. For low-frequency monitoring, the more parents show up in school, the better: the level of effort teachers specify increases with the number of parents who show up in school as frequently as once a week. Teachers' planned effort level, however, dramatically *decreases* if all parents were to show up in school every day – down to roughly the same effort level as if parents only showed up in school *once a year*.

Figure 8. Parents' and teachers' best-response functions

Panel A: Parents
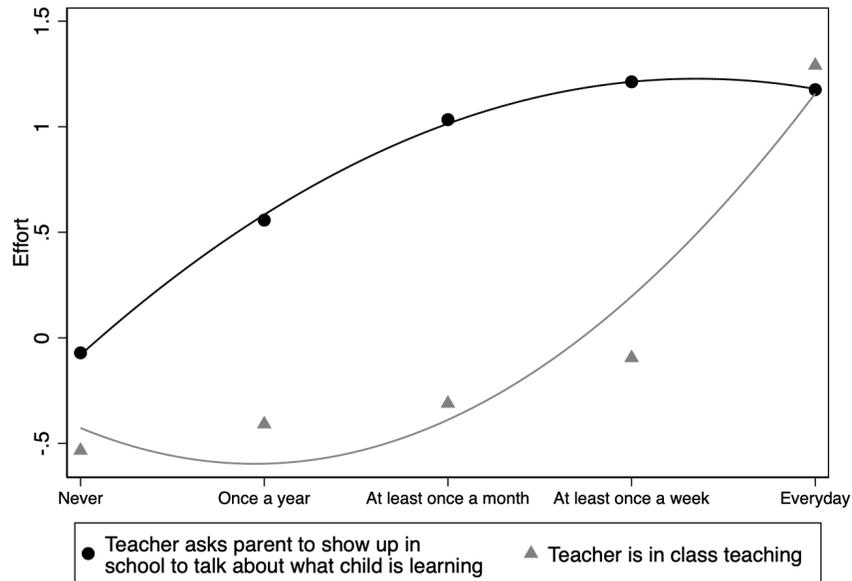
Panel B: Teachers
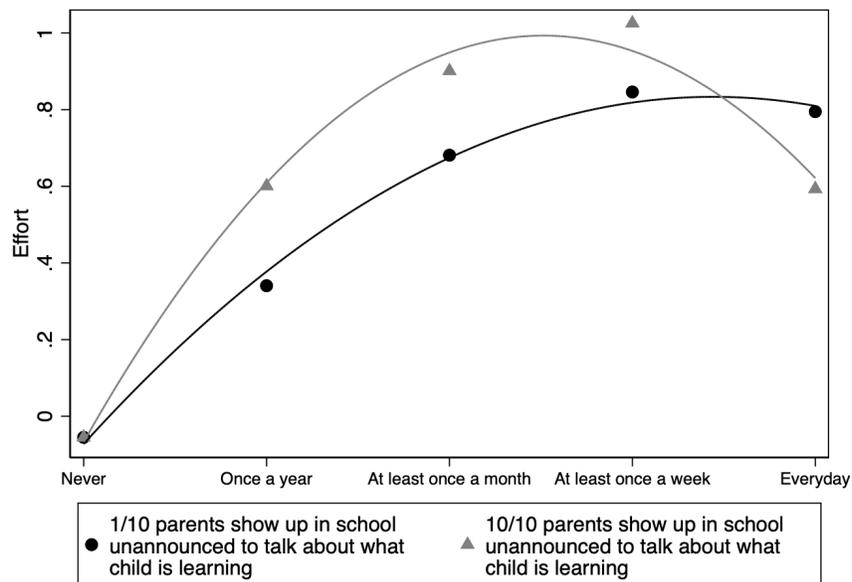
Note: Panel A showcases mean changes in effort by parents, computed from answers to the questions "During the last school year, how would your involvement in your child's education have changed if their teacher had asked you to come to school to talk about what s/he was learning at the following frequency?" and "During the last school year, how would your involvement in your child's education have changed if their teacher had been present in school teaching at the following frequency?", where frequency varies between "Never" and "Everyday". We plot parents' answer on a scale from -2 to 2, where -2 means "decrease a lot", 0 means "no change" and 2 means "increase a lot". Panel B showcases mean changes in effort by teachers, computed from answers to the questions "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 to talk about what their child is learning at the following frequency?", where frequency varies between "Never" and "Everyday". We plot parents' answer on a scale from -2 to 2, where -2 means "decrease a lot", 0 means "no change" and 2 means "increase a lot".

VI.   Concluding remarks

This paper documents first-hand that monitoring teachers can have non-monotonic effects. In our experiment, teacher accountability interventions massively decreased school dropouts when teachers were monitored either directly, through SMS nudges, or indirectly through community mobilization, but failed to improve educational outcomes when implemented together. *Too high* monitoring intensity backfires because it crowds out teachers' intrinsic motivation, leading to backlash: lower attendance, lower willingness to dedicate time to professional development activities, and instruction targeting to a greater extent. Lower teacher effort, in turn, triggered a chain reaction: in those schools, parents became overly pessimistic, parent-child dialogue was affected, and learning outcomes no longer improved. As a result, students dropped out of school to the same extent as if no intervention had been in place.

Evidence from heterogeneous treatment effects when parents are nudged via audio, and from the relationship between learning outcomes and community monitoring when only teachers are nudged, attest that results are not an artifact of our experimental design, but rather a general property of team production in children's education. In effect, we show that teachers' best-response function is indeed non-monotonic with respect to monitoring intensity.

Our findings match qualitative evidence about frustrated interactions between teachers and parents, documented in the education literature (e.g. Wolf and Suntheimer, 2020*a*; Chikutuma, 2017). They are also consistent with Frey (1993)'s 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 are already highly motivated. That risk might be particularly high in vulnerable communities, as teachers who accept those positions might be the ones most intrinsically motivated to start with (which would be consistent with Gertler, Patrinos and Rubio-Codina, 2012's results). As such, effectively intervening on teachers might require carefully eliciting their motivation and 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.

Those insights are by no means exclusive to education. From health to public works, it is also the case that accountability programs sometimes work (e.g. Nyqvist, de Walque and Svensson, 2017; Olken, 2007) but sometimes fail (e.g. Raffler, Posner and Parkerson, 2019; Lichand and Fernandes, 2019). Thinking carefully about monitoring intensity and the extent to which it can crowd-out bureaucrats' intrinsic motivation is central to the quality of governance agenda. In particular, we show that accountability can backfire even without displacing effort to margins less intensely monitored, and even in the absence of financial incentives.

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 highly motivated teachers find acceptable? Do financial incentives (in particular, pay for performance) prevent or magnify backlash in response to intense monitoring? Those are promising avenues for future research.

REFERENCES

Abadie, A, and G Imbens. 2016. "Matching on the Estimated Propensity Score." *Econometrica*, 84(2): 781–807.

Abadzi, H. 2007. "Predictors of Learning Outcomes in Rural Cote d'Ivoire." World Bank Policy Research Working Paper 4376.

Abimpaye, M, C Dusabe, J P Nzabonimpa, R Ashford, and L Pisani. 2019. "Improving parenting practices and development for young children in Rwanda: Results from a randomized control trial." *International Journal of Behavioral Development*, 44.

Agostinelli, F, M Saharkhiz, and M Wiswall. 2019. "Home and School in the Development of Children." National Bureau of Economic Research.

Banerjee, A, R Banerji, E Duflo, R Glennerster, and S Khemani. 2010. "Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India." *American Economic Journal: Economic Policy*, 2(1): 1–30.

Barrera, O, K Macours, P Premand, and R Vaki. 2020. "Texting Parents about Early Child Development: Behavioral Changes and Unintended Social Effects." World Bank Policy Research Working Paper 9492.

Bergman, Peter. 2021. "Parent-child information frictions and human capital investment: Evidence from a field experiment." *Journal of Political Economy*, 129(1): 199–222.

Bettinger, E, N Cunha, G Lichand, and R Madeira. 2020. "Are the Effects of Informational Interventions Driven by Salience?" University of Zurich, Department of Economics Working Paper 350.

Björkman, M, and J Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *The Quarterly Journal of Economics*, 124(2): 735–769.

Bold, T, D Filmer, G Martin, E Molina, C Rockmore, B Stacy, J Svensson, and W Wanea. 2017. "Monitoring works: Getting teachers to come to school." World Bank Policy Research Working Paper 7956.

Brophy, J. 2006. "Grade Repetition." nternational Academy of Education and International Institute for Educational Planning.

Chang, S, D A Cobb-Clark, and N Salamanca. 2020. "Parents' Responses to Teacher Qualifications."

Chikutuma, T. 2017. "Parental involvement in early childhood development (ECD) programmes: A case of Harare primary schools in Zimbabwe." *Zimbabwe Journal of Educational Research*, 29.

Christensen, D, O Dube, J Haushofer, B Siddiqi, and M Voors. forthcoming. "Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak." *The Quarterly Journal of Economics.*

Das, J, S Dercon, J Habyarimana, P Krishnan, K Muralidharan, and V Sundararaman. 2013. "School inputs, household substitution, and test scores." *American Economic Journal: Applied Economics*, 5(2): 29–57.

Dickinson, D, and M-C Villeval. 2008. "Does monitoring decrease work effort? The complementarity between agency and crowding-out theories." *Games and Economic Behavior*, 63: 56—-76.

Dizon-Ross, R. 2019. "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments." *American Economic Review*, 109(8): 2728–65.

Doepke, M, and F Zilibotti. 2017. "Parenting with style: Altruism and paternalism in intergenerational preference transmission." *Econometrica*, 85(5): 1331–1371.

Duflo, E, P Duplas, and M Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5): 1739–1774.

Duflo, E, R Hanna, and S Ryan. 2012. "Incentives work: Getting teachers to come to school." *American Economic Review*, 102(4): 1241–1278.

Falk, A, and M Kosfeld. 2006. "The Hidden Costs of Control." *American Economic Review*, 96: 1611–1630.

Fehr, E, and S Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Literature*, 14: 159–181.

Finch, J, S Wolf, and G Lichand. 2020. "The Importance of Executive Functions and Social-Emotional Skills for Students' Academic Success in Sub-Saharan Africa."

Frey, B. 1993. "Does monitoring increase work effort? The rivalry between trust and loyalty." *Economic Inquiry*, 31: 663—-670.

Fryer Jr, R, T Devi, and R Holden. 2016. "Vertical versus horizontal incentives in education: Evidence from randomized trials." National Bureau of Economic Research.

Gaduh, A, M Pradhan, J Priebe, and D Susanti. 2020. "Scores, Camera, Action? Incentivizing Teachers in Remote Areas."

Gerardino, M P, S Litschig, and D Pomeranz. 2017. "Can Audits Backfire? Evidence from Public Procurement in Chile." NBER Working Paper 23978.

Gertler, P, H Patrinos, and M Rubio-Codina. 2012. "Empowering parents to improve education: Evidence from rural Mexico." *Journal of Development Economics*, 99(1): 68–79.

Glewwe, P, and M Kremer. 2006. "Schools, teachers, and education outcomes in developing countries." *Handbook of the Economics of Education*, 2: 945–1017.

Görges, L, and D Nosenzo. 2020. "Social Norms and the Labor Market." *Handbook of Labor, Human Resources and Population Economics*, 1–26.

Greaves, E, I Hussain, B Rabe, and I Rasul. 2019. "Parental responses to information about school quality: Evidence from linked survey and administrative data." ISER Working Paper Series.

Kabay, S. 2016. "Grade Repetition and Primary School Dropout in Uganda." *Harvard Educational Review*, 86(4): 580–606.

Kling, J, J Liebman, and L Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica*, 75(1): 83–119.

Kremer, M, and A Holla. 2009. "Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?" *Annual Review of Economics*, 1: 513—542.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies*, 76(3): 1071–1102.

Lichand, G, and G Fernandes. 2019. "The Dark Side of the Contract: Do Government Audits Reduce Corruption in the Presence of Displacement By Vendors?"

Moon, B. 2007. "School-Based Teacher Development in Sub-Saharan Africa: Building a New Research Agenda." *Curriculum Journal*, 18(3): 355–371.

Nyqvist, M, D de Walque, and J Svensson. 2017. "Experimental Evidence on the Long-Run Impact of Community-Based Monitoring." *American Economic Journal: Applied Economics*, 9(1): 33–69.

Olken, B. 2007. "Monitoring corruption: Evidence from a field experiment in Indonesia." *Journal of Political Economy*, 115(2): 200—-249.

Pisani, L, I Borisova, and A J Dowd. 2018. "Developing and validating the International Development and Early Learning Assessment (IDELA)." *International Journal of Educational Research*, 91: 1–15.

Raffler, P, D Posner, and D Parkerson. 2019. "The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector."

Rasul, I, and D Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal*, 128: 413–446.

Roth, G, A Assor, Y Kanat-Maymon, and H Kaplan. 2007. "Autonomous Motivation for Teaching: How Self-Determined Teaching May Lead to Self-Determined Learning." *Journal of Educational Psychology*, 99(4): 761—-774.

Shi, L. 2008. "The Limit of Oversight in Policing: Evidence from the 2001 Cincinnati Riot." *Journal of Public Economics*, 93: 99–113.

Stehlé, J, N Voirin, A Barrat, C Cattuto, L Isella, J-P Pinton, M Quaggiotto, W Broeck, C Régis, B Lina, and P Vanhems. 2011. "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments." *PLoS One*.

UNICEF. 2019. "Harnessing the Power of Data for Gender Equality: Introducing the 2019 EM2030 SDG Gender Index." United Nations Children's Fund: Surrey, UK.

Wolf, S. 2020. ""Me I don't really discuss anything with them": Parent and teacher perceptions of early childhood education and parent-teacher relationships in Ghana." *International Journal of Educational Research*, 99.

Wolf, S, and G Lichand. 2020. "Are Child Labor and Education Really Substitutes?"

Wolf, S, and N M Suntheimer. 2020*a*. "'Me I don't really discuss anything with them': Parent and teacher perceptions of early childhood education and parent-teacher relationships in Ghana." *International Journal of Educational Research*, 99.

Wolf, S, and N M Suntheimer. 2020*b*. "Predictors of parental disciplinary practices and associations with child outcomes among Ghanaian preschoolers." *Child and Youth Services Review*, 112.

Wolf, S, G Lichand, and M Deambrosi. 2020. "Predictors of Learning Outcomes in Rural Cote d'Ivoire."

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

World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise.* Washington, DC:World Bank.

Yang, D. 2008. "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines." *The Review of Economics and Statistics*, 90: 1–14.

York, Benjamin N, Susanna Loeb, and Christopher Doss. 2019. "One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers." *Journal of Human Resources*, 54(3): 537–566.

ONLINE APPENDIX A - BALANCE TESTS

This Appendix compiles balance tests, to show that randomization was in fact properly implemented. Tables A1 and A2 display sample means for each cell when it comes to school characteristics (Panel A), caregiver characteristics (Panel B), student characteristics (Panel C) and teacher characteristics (Panel D). The tables display p-values for ANOVA tests of joint significance for differences in means across cells. We find few statistical differences, all small in magnitude and likely to emerge by chance given the large number of tests. In effect, F-tests for joint significance rule out systematic imbalances across cells in each panel.

Table A1—Descriptive statistics and balance tests

	Sub-sample means				ANOVA test p-value	# of obs.
	Control	Parents Only	Teachers Only	Both		
Panel A: School chars						
Number of students enrolled	258.44	281.83	283.48	273.35	0.714	99
Number of teachers	6.20	5.57	5.72	6.19	**0.073**	99
Hot days	13.48	15.43	14.36	10.62	0.158	99
Meals provided	0.56	0.52	0.60	0.31	0.187	99
Chalks	2.52	2.52	2.76	2.58	0.743	99
Blackboards	2.16	2.26	2.16	2.00	0.756	99
Textbooks	3.48	3.35	3.52	3.50	0.798	99
Notebooks	2.20	2.52	2.60	2.77	0.180	99
Pens	2.44	2.52	2.84	2.88	0.354	99
Monthly teacher absenteeism	1.64	1.35	1.32	1.35	0.524	99
Daily students' absence	7.12	11.22	11.16	8.38	0.653	99
F-test for nudges to parents					0.524	99
F-test for nudges to teachers					0.793	99
F-test for nudges to both					0.226	99
Panel B: Caregiver chars						
Female share	0.42	0.37	0.45	0.40	0.190	2,471
Relationship with child	3.08	3.30	3.03	3.24	0.323	2,471
Marital status	1.98	1.98	2.05	2.02	0.254	2,471
Years as primary caregiver	5.89	6.20	6.52	5.94	0.110	2,441
Source of income	3.65	3.76	3.79	3.89	0.901	2,471
Ever attended school	1.38	1.41	1.42	1.43	0.575	2,471
Std. parental engagement	-0.01	-0.04	-0.05	-0.07	0.628	2,471
Std. mindset wrt children	-0.07	-0.05	0.00	-0.01	0.627	2,471
Std. failure mindset	-0.02	-0.07	-0.09	-0.06	0.578	2,471
Std. parental aspirations	-0.01	0.02	-0.11	-0.03	0.314	2,471
Std. parental expectations	-0.04	-0.01	-0.07	-0.03	0.794	2,471
Beliefs child attendance	0.61	0.61	0.57	0.65	0.887	2,454
Std. parents' mental health	0.00	-0.08	-0.06	-0.08	0.106	2,471
Std. physical punishment	0.04	0.09	0.11	0.01	0.104	2,471
Std. child labor	0.01	0.07	0.02	0.09	0.503	2,471
Std. audio trust	0.00	0.02	0.03	0.07	0.833	2,408
Std. audio comprehension	0.00	-0.06	0.00	0.02	0.599	2,408
Std. visual trust	0.00	-0.07	-0.05	0.04	0.427	2,408
Std. visual comprehension	0.00	-0.02	0.04	0.07	0.601	2,408
Std. audio memory	0.00	-0.04	-0.02	-0.02	0.675	2,475
Std. visual memory	0.00	-0.03	-0.12	-0.01	0.810	2,475
Std. audio attention	0.00	0.05	0.03	0.03	0.532	2,475
Std. visual attention	0.00	0.04	0.01	-0.03	0.224	2,475
F-test for nudges to parents					0.512	2,362
F-test for nudges to teachers					0.291	2,362
F-test for nudges to both					0.582	2,362

Note: P-values computed with standard errors clustered at the classroom level.

Table A2—Descriptive statistics and balance tests

	Sub-sample means				ANOVA test p-value	# of obs.
	Control	Parents Only	Teachers Only	Both		
Panel C: Child chars						
Std. literacy	-0.13	-0.12	-0.11	-0.04	0.369	2,475
Std. numeracy	-0.14	-0.12	-0.10	-0.08	0.770	2,475
Std. social-emotional	-0.03	0.01	-0.03	-0.03	0.869	2,475
Std. working memory	-0.11	-0.10	-0.09	-0.09	0.936	2,475
Female share	0.47	0.54	0.50	0.50	0.123	2,475
Std. child labor	-0.01	-0.05	-0.03	-0.07	0.689	2,475
Std. self-esteem	0.04	0.01	0.05	0.02	0.869	2,475
Std. mindset	-0.06	-0.11	-0.03	-0.08	0.649	2,475
Teacher absenteeism	0.11	0.11	0.12	0.11	0.936	2,104
Std. parental engagement	-0.08	-0.13	-0.03	-0.04	0.162	2,475
Std. student effort	-0.09	-0.06	-0.06	-0.03	0.539	2,475
F-test for nudges to parents					0.115	2,104
F-test for nudges to teachers					0.797	2,104
F-test for nudges to both					0.542	2,104
Panel D: Teacher chars						
Age	41.29	38.70	38.28	37.71	0.198	197
Years worked as teacher	14.88	11.48	10.74	10.25	**0.097**	197
Years worked at this school	7.04	5.22	5.06	5.52	0.423	197
Education	3.88	4.26	4.18	4.23	0.168	197
Completed the last level	1.53	1.43	1.42	1.37	0.408	197
Std. language skills	0.00	-0.34	-0.39	-0.19	**0.065**	197
Follow advice from SMS	3.92	3.67	3.80	3.83	0.134	197
Std. external motivation	0.02	0.24	0.01	0.00	0.280	197
Std. introjected motivation	-0.02	0.05	-0.19	-0.12	0.374	197
Std. intrinsic motivation	0.04	0.10	-0.04	-0.03	0.756	197
Std. identified motivation	-0.05	0.05	-0.13	-0.14	0.592	197
Std. failure mindset	-0.05	-0.14	-0.19	-0.01	0.445	197
Std. job satisfaction	0.10	0.09	-0.04	-0.05	0.635	197
Teacher self-reported attendance	2.35	2.33	2.50	1.98	0.776	196
Arrive late	1.06	0.80	0.72	0.79	0.890	197
Leave early	0.14	0.11	0.16	0.27	0.674	197
F-test for nudges to parents					0.466	196
F-test for nudges to teachers					0.870	196
F-test for nudges to both					0.743	196

Note: P-values computed with standard errors clustered at the classroom level.

Online Appendix B - Proxy Dropouts

This Appendix validates our estimates of treatment effects that use classroom-level administrative data on student dropouts by using a survey proxy, equal to 1 if we cannot track a student between our baseline and end-line surveys, and 0 otherwise. As Figure B1 shows, proxy dropouts and administrative dropouts are highly correlated. The remainder of the Appendix documents that the patterns for treatment effects using administrative dropouts discussed in the main text are similar to those found when we use proxy dropouts instead.

Figure B1. Binscatter plot of the survey proxy (averaged at the classroom level) and administrative data on student dropouts
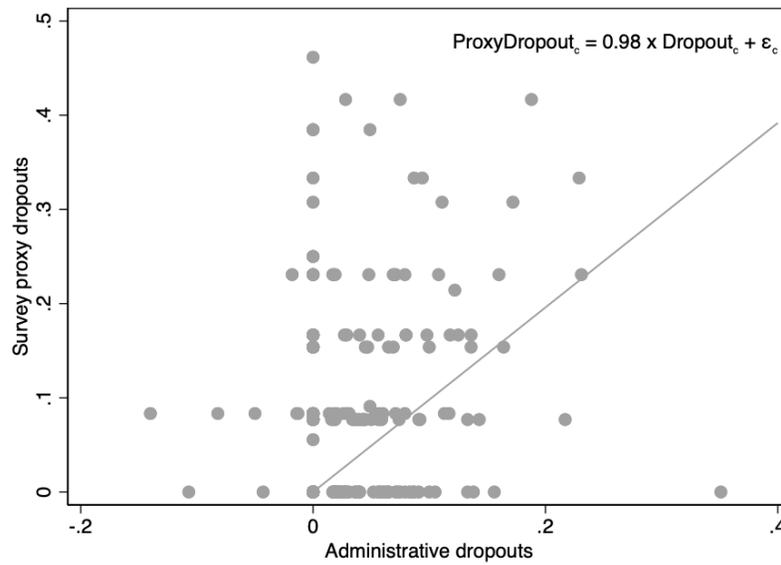

$$\text{ProxyDropout}_c = 0.98 \times \text{Dropout}_c + \varepsilon_c$$

Table B1—Treatment effects on proxy dropouts, by above- and below-median teacher motivation and above- and below-median teacher absenteeism

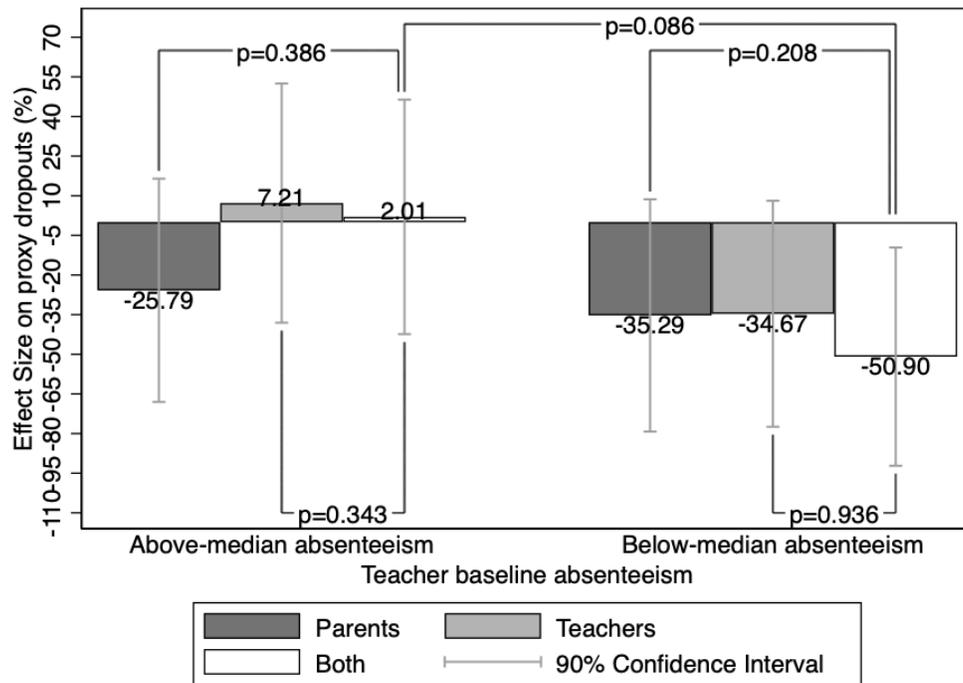| | Proxy dropouts | | | | |
| | | Median baseline motivation | | Median baseline absenteeism | |
| | All teachers | Above-median | Below-median | Above-median | Below-median |
	(1)	(2)	(3)	(4)	(5)
Parents	-0.0409*	-0.0438	-0.0373	-0.0270	-0.0470
	(0.0236)	(0.0305)	(0.0384)	(0.0266)	(0.0352)
Teachers	-0.0262	-0.0368	-0.0196	0.0075	-0.0462
	(0.0236)	(0.0322)	(0.0372)	(0.0285)	(0.0343)
Both	-0.0338	0.0056	-0.0629*	0.0021	-0.0678**
	(0.0227)	(0.0291)	(0.0355)	(0.0279)	(0.0331)
Control group mean	0.1209	0.1157	0.1273	0.1045	0.1331
Parents = Both [p-value]	0.7077	0.0856	0.3492	0.2493	0.4574
Teachers = Both [p-value]	0.6925	0.1620	0.0970	0.8328	0.4091
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Observations	2,462	1,220	1,242	1,243	1,232
R-squared	0.0087	0.0156	0.0172	0.0221	0.0173

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. In all columns, the dependent variable equal to 0 if the participant surveyed at baseline was successfully tracked at end line, and 1 otherwise. In columns (2) and (3), teachers' motivation is constructed following Kling, Liebman and Katz (2007) by standardizing each motivation component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. In columns (4) and (5), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students. Median of teachers' baseline absenteeism is 9.52%. We control for baseline students' characteristics and grade fixed effects. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure B2. Heterogeneous treatment effects of monitoring on proxy dropouts, by baseline teacher motivation
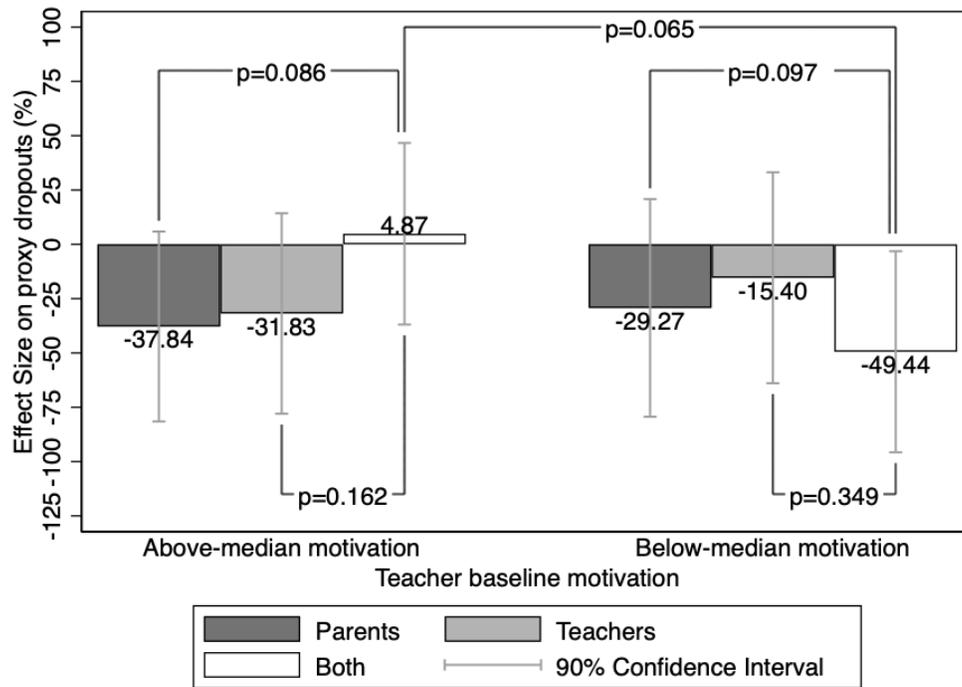
Note: Proxy dropouts equal to 0 if the participant surveyed at baseline was successfully tracked at end line, and 1 otherwise. Teachers' motivation is constructed following Kling, Liebman and Katz (2007) by standardizing each motivation component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We control for baseline students' characteristics and grade fixed effects. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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. We control for grade fixed effects. P-values shown in dark brackets.

Figure B3. Heterogeneous treatment effects of monitoring on proxy dropouts, by baseline teacher absenteeism

Note: Proxy dropouts equal to 0 if the participant surveyed at baseline was successfully tracked at end line, and 1 otherwise. Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students. Median of teachers' baseline absenteeism is 9.52%. We control for baseline students' characteristics and grade fixed effects. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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. We control for grade fixed effects. P-values shown in dark brackets.

ONLINE APPENDIX C - INSTRUMENTAL VARIABLES

This Appendix accounts for imperfect compliance with the interventions by estimating an instrumental variables' model. We show results for survey outcomes and 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). Table C1 reports IV estimates. Results are very similar to the patterns documented in the main text, only more precisely estimated.

Table C1—Instrumental variables estimates for treatment effects on proxy dropouts, teacher absenteeism and teachers' time use

	Proxy dropout (1)	Teacher absenteeism (2)	Professional development (3)
Parents	-0.5770***	-0.0196	-0.3247
	(0.0657)	(0.0167)	(0.2521)
Teachers	-0.3581***	0.0025	-1.0168***
	(0.0436)	(0.0106)	(0.3061)
Both	-0.2431***	0.0349	-0.5169**
	(0.0407)	(0.0250)	(0.2340)
Control group mean	0.1968	0.0908	-0.1644
Parents = Both [p-value]	0.0000	0.0528	0.4740
Teachers = Both [p-value]	0.0348	0.2438	0.0855
Grade Fixed Effects	Yes	Yes	Yes
Baseline controls	Yes	Yes	No
Weights	No	Yes	No
Observations	2,475	1,845	180
R-squared		0.0030	0.0331

Note: 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. In column (1), the dependent variable equal to 0 if the participant surveyed at baseline was successfully tracked at end line, and 1 otherwise. We control for grade fixed effects and baseline students' characteristics in this column. In column (2), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. In column (3), the dependent variable is the standardized value of teacher-reported willingness to invest time in professional development weekly, collected in the follow-up. In column (2), we use end-line data only and control for grade fixed effects, baseline students' characteristics, and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in grade level; child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors clustered at the classroom level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Online Appendix D - Matching

This Appendix compiles the details of the propensity score matching procedure, and shows that results in the main text are robust to matching observations based on their baseline characteristics through a non-parametric, nearest-neighbor algorithm. Table D1 reports estimates from a logistic regression with dependent variable equal to 1 if the participant surveyed at baseline was also present at end line, and 0 otherwise, regressed on baseline student characteristics, restricting the sample to those in control schools. These estimates are then used to predict propensity scores for each observation in the whole sample. The remainder of this Appendix then reproduces the results of the tables and figures that rely on end-line survey data, using a nearest-neighbour matching algorithm to handle selective attrition instead. In those estimates, each observation (in each of the treatment cell) is matched with its closest match in the control group, based on its predicted propensity score.

D1. Propensity score estimation

Table D1—Propensity score's estimation

| | Tracked at end line |
	(1)
Grade	0.1533
	(0.3740)
Child gender	-0.1540
	(0.2852)
Std. numeracy	0.1283
	(0.2483)
Std. literacy	0.9650**
	(0.4132)
Std. parental engagement	0.2070
	(0.1294)
Std. student effort	0.2118
	(0.2359)
Std. child labor	-0.2330
	(0.2119)
Std. social-emotional	-0.6276**
	(0.2599)
Std. working memory	-0.0812
	(0.3032)
Std. visual attention	-0.1961
	(0.2983)
Std. impulsivity	0.2188
	(0.2115)
Std. self-esteem	-0.0638
	(0.1798)
Std. mindset	-0.1499
	(0.1828)
Observations	625

Note: Logistic regression model with dependent variable equal to 1 if the participant surveyed at baseline was also present at end line, and 0 otherwise. Independent variables are observed baseline students' characteristics. The sample is restricted to the control group. Standard error clustered at the classroom level. Significance levels are denoted by *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table D2—Nearest-neighbor matching strategy for manipulation checks

	Parent recalls receiving messages		
	Text+audio	Text only	Audio only
	(1)	(2)	(3)
Parents	0.3486***	0.3362***	0.3547***
	(0.0328)	(0.0412)	(0.0400)
Teachers	-0.0278	-0.0285	-0.0210
	(0.0290)	(0.0282)	(0.0289)
Both	0.2956***	0.2357***	0.3688***
	(0.0368)	(0.0422)	(0.0510)
Control group mean		0.1764	
Parents = Both [p-value]	0.1739	0.0441	0.8129
Teachers = Both [p-value]	0.0000	0.0000	0.0000
Grade Fixed Effects	Yes	Yes	Yes
Matching group Fixed Effects	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes
Observations	2,119	1,609	1,577
R-squared	0.3840	0.4248	0.4530

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. In column (1), the outcome variable is a dummy equal to 1 if the caregiver recalled receiving text- or audio-messages on her/his phone in the previous year when we surveyed her/him at end line, and 0 otherwise; in columns (2) and (3), we use the same outcome but restrict our analyses to the sub-sample of parents who received text-messages on their phone in column (2), and to those who received audio-messages on their phone in column (3). We control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. The sample is restricted to students who were tracked at end line. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table D3—Nearest-neighbor matching strategy for treatment effects on teacher absenteeism

	Teacher absenteeism		
	All teachers (1)	Above-median (2)	Below-median (3)
Parents	-0.0080	-0.0268	0.0084
	(0.0087)	(0.0170)	(0.0149)
Teachers	0.0037	-0.0132	0.0045
	(0.0081)	(0.0171)	(0.0135)
Both	0.0238**	0.0255	0.0118
	(0.0112)	(0.0228)	(0.0145)
Control group mean	0.0927	0.1081	0.0804
Parents = Both [p-value]	0.0053	0.0273	0.8444
Teachers = Both [p-value]	0.0723	0.0980	0.6293
Grade Fixed Effects	Yes	Yes	Yes
Matching group Fixed effects	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes
Observations	1,886	949	937
R-squared	0.3196	0.5451	0.5069

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. Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. Median of teachers' baseline absenteeism is 9.52%. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table D4—Nearest-neighbor matching strategy for treatment effects on parental engagement

	School participation	Academic activities	Motivation	Dialogue
	(1)	(2)	(3)	(4)
Parents	0.1275**	-0.0215	0.0049	0.0082
	(0.0607)	(0.0551)	(0.0523)	(0.0539)
Teachers	0.0247	-0.0940*	0.0266	-0.0057
	(0.0577)	(0.0526)	(0.0564)	(0.0553)
Both	-0.0671	0.0095	-0.0303	0.0079
	(0.0579)	(0.0555)	(0.0538)	(0.0500)
Parents = Both [p-value]	0.0006	0.6008	0.4906	0. 9949
Teachers = Both [p-value]	0.0866	0.0697	0.3046	0.7852
Grade Fixed Effects	Yes	Yes	Yes	Yes
Matching group Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Observations	2,119	2,119	2,119	2,119
R-squared	0.2920	0.2802	0.3575	0.3186

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. In column (1), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents participate in school meetings; and frequency at which parents talk to teacher. In column (2), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents help with homework; frequency at which parents help organize school material. In column (3), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents incentivize school attendance; frequency at which parents incentivize studying. In column (4), the outcome variable is constructed from the following individual components asked to parents: frequency at which parents ask about homework; frequency at which parents ask about school grades; frequency at which parents ask how school was. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table D5—Nearest-neighbor matching strategy for treatment effects on disciplining strategies

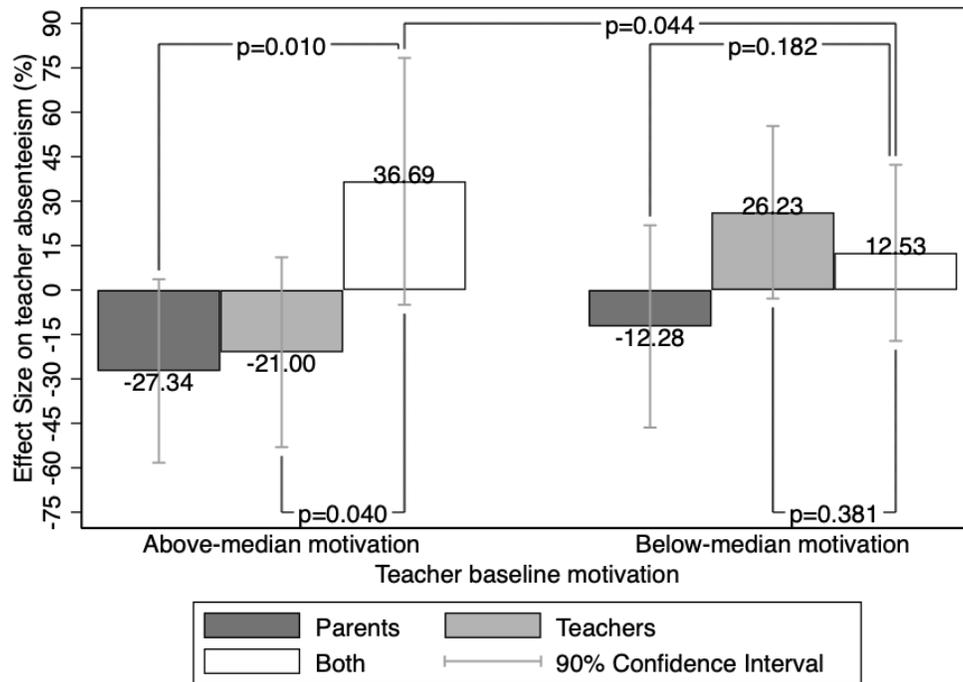|	Summary measure (1)	Components		
		Without violence (2)	Physical violence (3)	Psychological violence (4)
Parents	-0.0090	0.0964**	0.0460	0.0234
	(0.0248)	(0.0436)	(0.0434)	(0.0553)
Teachers	-0.0163	0.0608	0.0188	-0.0069
	(0.0263)	(0.0465)	(0.0538)	(0.0548)
Both	0.0365	-0.0423	0.0151	0.0520
	(0.0253)	(0.0435)	(0.0475)	(0.0542)
Parents = Both [p-value]	0.0775	0.0013	0.4946	0.5864
Teachers = Both [p-value]	0.0511	0.0227	0.9462	0.2540
Grade Fixed Effects	Yes	Yes	Yes	Yes
Matching group Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Observations	2,119	2,119	2,119	2,119
R-squared	0.2827	0.2810	0.2953	0.2569

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. In column (1), the outcome averages across the outcomes of columns (2) to (4) (with the outcome of col. (1) set to negative). In column (2), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household took away privileges, forbade something the child likes (e.g., watching TV, playing with friends) or did not allow the child to leave the house in the past month; whether someone in the household explained why the child's behavior was wrong in the past month; whether someone in the household gave the child something else to do in the past month. In column (3), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household shook the child in the past month; whether someone in the household hit or spanked the child with bare hand in the past month; whether someone in the household spanked the child or hit him/her elsewhere on the body with something like a belt, hairbrush, stick or other hard object in the past month; whether someone in the household hit or slapped the child on the face, head or ears in the past month; whether someone in the household hit or slapped the child on the hand, arm, or leg in the past month; whether someone in the household beat the child up, that is hit him/her over and over as hard as one could in the past month; whether the caregiver believes that in order to bring up, raise, or educate a child properly, the child needs to be physically punished. In column (4), the outcome variable is constructed from the following individual components asked to caregivers: whether someone in the household shouted, yelled at or screamed at the child in the past month; whether someone in the household called the child dumb, lazy, or another name like that in the past month. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table D6—Nearest-neighbor matching strategy for treatment effects on parental beliefs, student effort, and learning

|				Components	
	Parents' beliefs	Student effort	Test scores summary measure	Numeracy (std.)	Literacy (std.)
	(1)	(2)	(3)	(4)	(5)
Parents	-0.0583	0.0574	0.0675*	0.0675*	0.0674
	(0.0616)	(0.0445)	(0.0388)	(0.0376)	(0.0471)
Teachers	-0.1158**	0.0546	-0.0279	0.0048	-0.0606
	(0.0542)	(0.0423)	(0.0353)	(0.0340)	(0.0458)
Both	-0.2357***	0.0409	-0.0150	-0.0056	-0.0244
	(0.0629)	(0.0391)	(0.0340)	(0.0355)	(0.0411)
Parents = Both [p-value]	0.0079	0.6614	0.0255	0.0496	0.0422
Teachers = Both [p-value]	0.0405	0.6890	0.7028	0.7629	0.4124
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes
Matching group Fixed Effects	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Observations	1,960	2,180	2,180	2,180	2,180
R-squared	0.3251	0.3237	0.6753	0.6142	0.6441

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. In column (1), the outcome variable is constructed from the following individual components asked to caregivers: parental beliefs about performance in mathematics; parental beliefs about performance in French; parental beliefs about overall school performance; parental beliefs about future school performance. In column (2), the outcome variable is a summary measure constructed from the following individual components asked to students, equal to 1 if answer is sometimes or more more often, and 0 otherwise: whether they study at home on a general weekday; whether they listen to the radio on a general weekday; whether they read a book, a newspaper or magazines on a general weekday. In column (3), the outcome variable is a summary measure constructed from individual components consisting in mathematics exercises completed by students. In column (4), the outcome variable is a summary measure constructed from individual components consisting in French exercises completed by students. In column (5), the summary measure averages across numeracy and literacy standardized test scores. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure D1. Nearest-neighbor matching strategy for heterogeneous treatment effects of monitoring on teacher absenteeism, by baseline teacher motivation

Note: Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. Teachers' motivation is constructed following Kling, Liebman and Katz (2007) by standardizing each motivation component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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.

Figure D2. Nearest-neighbor matching strategy for heterogeneous treatment effects of monitoring on teacher absenteeism, by audio vs text messages
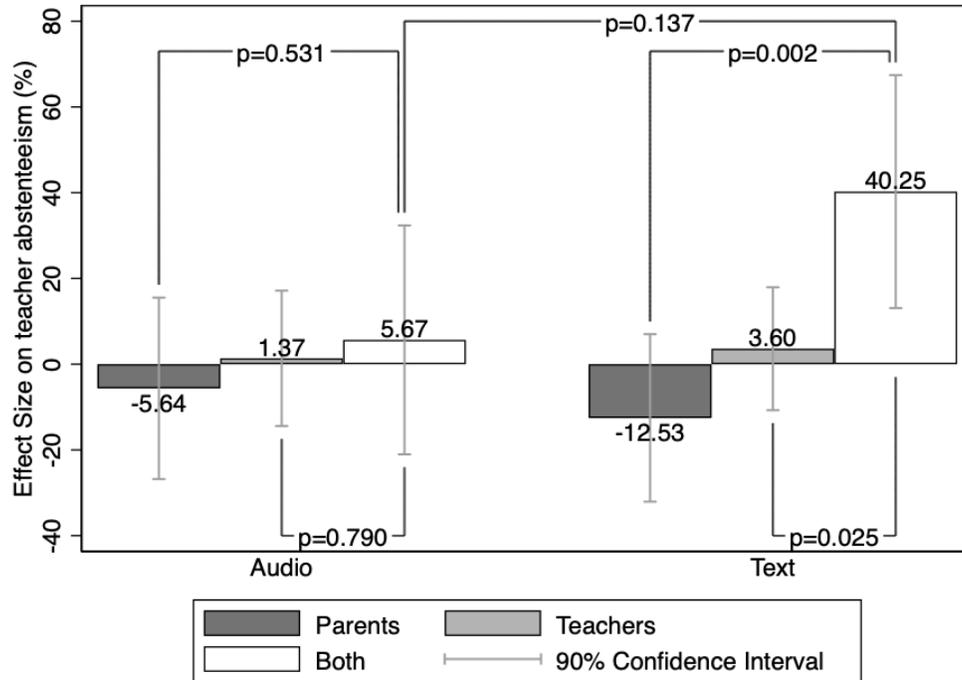
Note: Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. We use end-line data only and control for baseline students' characteristics, grade fixed effects, and matching group fixed effects. To control for selective attrition, we use a nearest-neighbor matching strategy in all columns, where each observation in the control sub-sample is matched with the closest observation in each of the treatment groups based on the predicted probability of students to be tracked at end line, predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. 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.

Online Appendix E - Teachers' selective non-response

Treatment assignment is systematically correlated with non-response at our follow-up survey with teachers, but we do not have baseline characteristics for all of them since this was when CM2 teachers were surveyed for the first time. For this reason, this Appendix computes bounds for treatment effects on outcomes based on follow-up 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 E1 reports the results, showing that our estimates are not sensitive to teachers' selective non-response at the additional follow-up survey.

Table E1—Lee Bounds for treatment effects on monitoring and teachers' time use

	Selective non-response	Teacher reports parents show up		Professional development	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
Parents	0.0016	0.0038	0.2208	-0.5815	0.0253
	(0.0441)				
Teachers	0.1081**	-0.0982	0.0947	-0.9382	-0.1729
	(0.0435)				
Both	0.0370	0.0158	0.2320	-0.9189	-0.2616
	(0.0429)				
Control group mean	0.0139	0.0981	0.1232	0.1121	0.1177
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	296				
R-squared	0.0468				

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. In column (1), the outcome is a dummy variable where dependent variable = 1 if we had no answer for that teacher in the additional follow-up follow-up survey, and 0 otherwise. In columns (2) to (4), we report the upper and lower limits of the 90% confidence interval of respectively the upper and lower bounds computed following Lee (2009). Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

ONLINE APPENDIX F - ADDITIONAL ROBUSTNESS CHECKS

This Appendix presents additional robustness checks for the main results of the paper. As discussed in the main text, our classroom-level measures of student dropouts and grade repetitions are imperfect because we cannot account for student transfers across schools. In particular, this leads those measures to assume negative values in a few instances. Table F1 shows that our estimates for treatment effects on student dropouts and grade repetition using administrative data are robust to setting negative values to 0. The remainder of the Appendix reproduces our estimates of treatment effects in the main text but clustering standard errors at the school level instead.

Table F1—Treatment effects on student dropout and grade repetition

		Components	
	Summary measure (1)	Dropout rate (2)	Grade rep. rate (3)
Parents	-0.1788*	-0.0182**	-0.0086
	(0.1063)	(0.0088)	(0.0198)
Teachers	-0.1001	-0.0195**	0.0148
	(0.1048)	(0.0087)	(0.0195)
Both	-0.0217	-0.0030	0.0005
	(0.1035)	(0.0086)	(0.0193)
Control group mean	0.0000	0.0468	0.1527
Parents = Both [p-value]	0.1357	0.0810	0.6397
Teachers = Both [p-value]	0.4493	0.0550	0.4596
Grade Fixed Effects	Yes	Yes	Yes
Observations	296	296	296
R-squared	0.0320	0.0376	0.0169

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. Data are from administrative records, corrected by setting to 0 the negative values of dropout (9 values) and grade repetition rate (3 values). In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are dropout rate (col. 2) and grade repetition rate (col. 3) at the classroom level, corrected by setting negative values to 0. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table F2—Treatment effects on student dropouts and grade repetition with standard errors clustered at the school level

	Summary measure (1)	Components	
		Dropout rate (2)	Grade rep. rate (3)
Parents	-0.2393	-0.0248**	-0.0107
	(0.1463)	(0.0114)	(0.0242)
Teachers	-0.1225	-0.0223**	0.0148
	(0.1325)	(0.0100)	(0.0213)
Both	-0.0428	-0.0031	-0.0047
	(0.1388)	(0.0107)	(0.0244)
Control group mean	0.0000	0.0468	0.1527
Parents = Both [p-value]	0.1102	0.0329	0.7972
Teachers = Both [p-value]	0.4496	0.0237	0.3395
Grade Fixed Effects	Yes	Yes	Yes
Observations	296	296	296
R-squared	0.0353	0.0462	0.0151

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. Data are from administrative records. In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are dropout rate (col. (2)) and grade repetition rate (col. (3)) at the classroom level. Standard errors in parentheses clustered at the school level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table F3—Treatment effects on teacher absenteeism and time use with standard errors clustered at the school level

| | Teacher absenteeism | | | |
	All teachers (1)	Above-median (2)	Below-median (3)	Professional development (4)
Parents	-0.0080	-0.0253*	0.0051	-0.2772
	(0.0093)	(0.0130)	(0.0113)	(0.2167)
Teachers	0.0032	-0.0004	0.0042	-0.6219***
	(0.0097)	(0.0138)	(0.0109)	(0.1906)
Both	0.0191	0.0252	0.0070	-0.5885***
	(0.0130)	(0.0199)	(0.0126)	(0.2097)
Control group mean	0.0927	0.1081	0.0804	0.0000
Parents = Both [p-value]	0.0433	0.0169	0.8865	0.2048
Teachers = Both [p-value]	0.2506	0.2190	0.8281	0.8807
Grade Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	No
Predicted propensity scores	Yes	Yes	Yes	No
Observations	1,940	949	991	273
R-squared	0.0137	0.0369	0.0194	0.0659

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. In columns (1) to (3), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. In column (4), the dependent variable is the standardized value of teacher-reported willingness to invest time in professional development weekly, collected in the follow-up. In columns (1) to (3), we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of being tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses, clustered at the school level. * p<0.1, ** p<0.05, *** p<0.01.

ONLINE APPENDIX G - ADDITIONAL RESULTS

This Appendix compiles additional tables and figures to complement the analyses presented in the main text. Figure G1 displays histograms for the distribution of teacher absenteeism 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 absenteeism is quite low in our sample (below 10%), 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.The remainder of the Appendix estimates heterogeneous treatment effects by grade and gender.

Figure G1. Histogram of teacher absenteeism
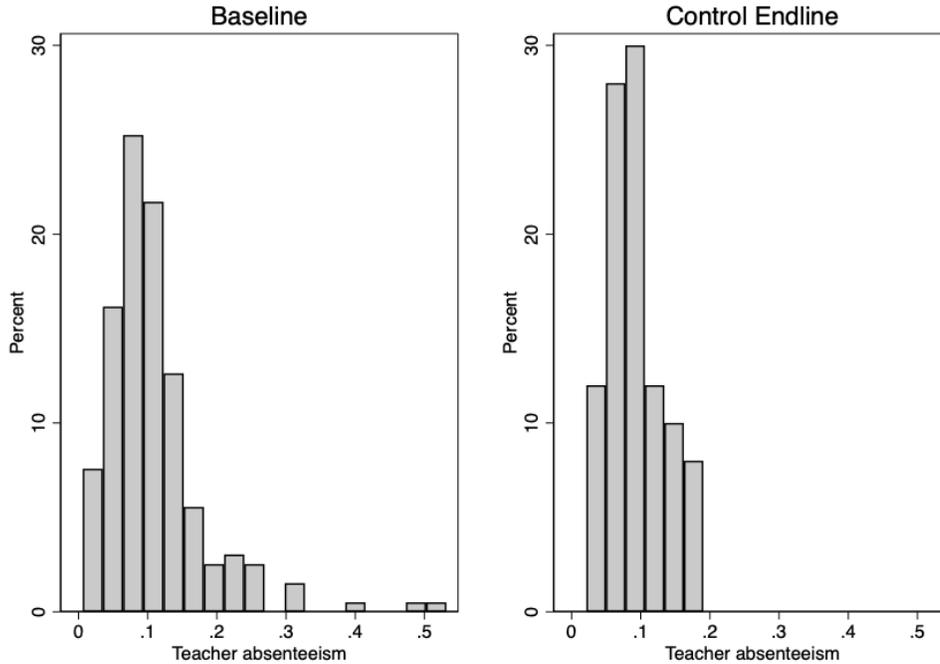
Note: Histogram of teacher absenteeism at the classroom level. Teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the 2 weeks prior to the survey, according to students.

G1.   Heterogeneous treatment effects by grade

Table G1—Treatment effects on student dropout and grade repetition for CP2 students

	Administrative data		
		Components	
	Summary measure	Dropout rate	Grade rep. rate
	CP2	CP2	CP2
	(1)	(2)	(3)
Parents	-0.0917	-0.0147	0.0068
	(0.2258)	(0.0153)	(0.0267)
Teachers	0.1668	-0.0086	0.0405
	(0.2235)	(0.0151)	(0.0264)
Both	0.1385	-0.0017	0.0255
	(0.2214)	(0.0150)	(0.0261)
Control group mean	0.0000	0.0458	0.1358
Parents = Both [p-value]	0.3060	0.3928	0.4819
Teachers = Both [p-value]	0.8988	0.6453	0.5659
Baseline controls	No	No	No
Observations	100	100	100
R-squared	0.0178	0.0121	0.0293

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. The sample is restricted to CP2 students. In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are rates at the classroom level. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table G2—Treatment effects on student dropout and grade repetition for CE2 students

| | Administrative data | | |
| | | Components | |
	Summary measure CE2 (1)	Dropout rate CE2 (2)	Grade rep. rate CE2 (3)
Parents	-0.2502	-0.0286	0.0081
	(0.2272)	(0.0177)	(0.0346)
Teachers	-0.1979	-0.0270	0.0166
	(0.2249)	(0.0175)	(0.0343)
Both	-0.2711	-0.0038	-0.0535
	(0.2227)	(0.0173)	(0.0339)
Control group mean	0.0000	0.0515	0.1739
Parents = Both [p-value]	0.9264	0.1598	0.0754
Teachers = Both [p-value]	0.7433	0.1830	0.0416
Baseline controls	No	No	No
Observations	100	100	100
R-squared	0.0186	0.0441	0.0517

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. The sample is restricted to CE2 students. In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are rates at the classroom level. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table G3—Treatment effects on student dropout and grade repetition for CM2 students

	Administrative data		
		Components	
	Summary measure	Dropout rate	Grade rep. rate
	CM2	CM2	CM2
	(1)	(2)	(3)
Parents	-0.3317**	-0.0316*	-0.0498
	(0.1555)	(0.0166)	(0.0442)
Teachers	-0.2340	-0.0318*	-0.0138
	(0.1521)	(0.0163)	(0.0433)
Both	0.0122	-0.0041	0.0136
	(0.1492)	(0.0159)	(0.0424)
Control group mean	0.0000	0.0428	0.1481
Parents = Both [p-value]	0.0266	0.0948	0.1478
Teachers = Both [p-value]	0.1021	0.0859	0.5194
Observations	96	96	96
R-squared	0.0755	0.0676	0.0245

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. The sample is restricted to CM2 students. In column (1), following Kling, Liebman and Katz (2007), the summary measure averages across index components, each normalized by control group mean and standard deviation; in columns (2) and (3), index components are rates at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table G4—Teacher absenteeism and time use for CP2, CE2 and CM2 students

	Teacher absenteeism		Professional development		
	CP2	CE2	CP2	CE2	CM2
	(1)	(2)	(3)	(4)	(5)
Parents	-0.0081	-0.0089	-0.7505**	0.1201	-0.1848
	(0.0116)	(0.0127)	(0.3301)	(0.3791)	(0.3340)
Teachers	0.0048	0.0002	-0.8286**	-0.8926**	-0.1885
	(0.0109)	(0.0139)	(0.3425)	(0.3886)	(0.3302)
Both	0.0260	0.0094	-0.7457**	-0.3602	-0.6521**
	(0.0191)	(0.0152)	(0.3233)	(0.3748)	(0.3266)
Control group mean	0.0910	0.0945	0.1005	-0.2497	0.1388
Parents = Both [p-value]	0.0891	0.1917	0.9883	0.2086	0.1652
Teachers = Both [p-value]	0.2863	0.5435	0.8077	0.1744	0.1637
Baseline controls	Yes	Yes	No	No	No
Predicted propensity scores	Yes	Yes	No	No	No
Observations	975	965	92	88	93
R-squared	0.0248	0.0169	0.0864	0.0872	0.0465

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. In columns (1) and (3), the sample is restricted to CP2 students. In columns (2) and (4), the sample is restricted to CE2 students. In column (5), the sample is restricted to CM2 students. In columns (1) and (2), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students. In columns (3) to (5), the dependent variable is the standardized value of teacher-reported willingness to invest time in professional development weekly, collected in the follow-up. In columns (1) and (2), we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses, clustered at the classroom level in column (1) and (2). * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

G2. *Heterogeneous treatment effects by gender*

Table G5—Treatment effects on proxy dropouts for boys and for girls

	Survey data	
	Proxy dropouts	
	Boys	Girls
	(1)	(2)
Parents	-0.0242	-0.0569*
	(0.0292)	(0.0296)
Teachers	0.0039	-0.0586**
	(0.0323)	(0.0290)
Both	-0.0187	-0.0511*
	(0.0275)	(0.0290)
Control group mean	0.1145	0.1263
Parents = Both [p-value]	0.8261	0.7936
Teachers = Both [p-value]	0.4367	0.7232
[Boys]Parents = [Girls]Parents [p-value]	0.3691	
[Boys]Teachers = [Girls]Teachers [p-value]	0.1134	
[Boys]Both = [Girls]Both [p-value]	0.3476	
Grade Fixed Effects	Yes	Yes
Baseline controls	Yes	Yes
Observations	1,237	1,238
R-squared	0.0081	0.0172

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. In column (1), the sample is restricted to boys, while it's restricted to girls in column (2). The dependent variable equal to 0 if the participant surveyed at baseline was successfully tracked at end line, and 1 otherwise. We control for grade fixed effects and baseline students' characteristics, which consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table G6—Teacher absenteeism and time use by gender

	Teacher absenteeism		Professional development	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Parents	-0.0094	-0.0072	-0.1741	-0.6736
	(0.0102)	(0.0152)	(0.2760)	(0.5063)
Teachers	0.0084	0.0016	-0.8775***	-0.7573
	(0.0118)	(0.0121)	(0.2999)	(0.4868)
Both	0.0068	0.0328*	-0.4502	-0.5541
	(0.0142)	(0.0192)	(0.3033)	(0.4425)
Control group mean	0.0920	0.0946	-0.0146	-0.1910
Parents = Both [p-value]	0.2376	0.0626	0.3680	0.7932
Teachers = Both [p-value]	0.9141	0.0955	0.1942	0.6351
[Male]Parents=[Female]Parents [p-value]	0.9081		0.3846	
[Male]Teachers=[Female]Teachers [p-value]	0.4597		0.8362	
[Male]Both=[Female]Both [p-value]	0.2918		0.8476	
Grade Fixed Effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	No	No
Predicted propensity scores	Yes	Yes	No	No
Observations	1,177	763	108	72
R-squared	0.0189	0.0470	0.0846	0.0518

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. In columns (1) and (3), the sample is restricted to male teachers. In columns (2) and (4), the sample is restricted to female teachers. In columns (3) and (4), the dependent variable is the standardized value of teacher-reported willingness to invest time in professional development weekly, collected in the follow-up. In columns (1) and (2), teacher absenteeism stands for the share of days on which teachers were absent from the classroom over the two weeks prior to the survey, according to students, and we use end-line data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at end line (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at end line on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses, clustered at the classroom level in column (1) and (2). Standard errors in parentheses, clustered at the classroom level in columns (1) and (2). * p<0.1, ** p<0.05, *** p<0.01.

Table G7—Treatment effects on test scores by gender

	Summary measure		Numeracy (std.)		Literacy (std.)	
	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Parents	0.0974**	0.0350	0.1110**	0.0281	0.0838	0.0418
	(0.0488)	(0.0515)	(0.0490)	(0.0517)	(0.0583)	(0.0619)
Teachers	0.0201	-0.0916**	0.0685	-0.0756	-0.0282	-0.1076**
	(0.0422)	(0.0451)	(0.0448)	(0.0480)	(0.0531)	(0.0544)
Both	0.0237	-0.0573	0.0232	-0.0238	0.0243	-0.0909*
	(0.0414)	(0.0449)	(0.0454)	(0.0497)	(0.0485)	(0.0516)
Control group mean	0.1085	0.1754	0.1308	0.1524	0.0862	0.1983
Parents = Both [p-value]	0.1217	0.0372	0.0773	0.2815	0.2847	0.0130
Teachers = Both [p-value]	0.9290	0.3487	0.3154	0.2402	0.2908	0.7032
[Male]Parents=[Female]Parents [p-value]	0.2450		0.1539		0.5207	
[Male]Teachers=[Female]Teachers [p-value]	0.0206		0.0075		0.1959	
[Male]Both=[Female]Both [p-value]	0.0987		0.4048		0.0458	
Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Predicted propensity scores	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,109	1,134	1,109	1,134	1,109	1,134
R-squared	0.5688	0.5815	0.4863	0.4859	0.5343	0.5414

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. In columns (1) and (2), the outcome variable is a summary measure constructed from individual components consisting in mathematics exercises completed by students. In columns (3) and (4), the summary measure averages across numeracy and literacy standardized test scores. In columns (5) and (6), the outcome variable is a summary measure constructed from individual components consisting in French exercises completed by students. Sample restricted to boys only in odd columns, and to girls only in even columns. For all outcomes, we follow Kling, Liebman and Katz (2007) by standardizing each component within grade (normalizing by their mean and standard deviation among the control group) and then averaging across them in order to construct the variable. In all columns, we use endline data only and control for baseline students' characteristics, grade fixed effects and for the predicted probability of students to be tracked at endline (as well as the square and cubic terms of this variable), predicted using a logistic regression of a dummy indicating tracking at endline on baseline students' characteristics in the control group. Baseline characteristics consist in child gender; standardized baseline grades (numeracy and literacy); standardized parental engagement; standardized student effort; standardized child labor; standardized socio-emotional skills; standardized working memory; standardized visual attention; standardized impulsivity; standardized self-esteem; standardized mindset. Standard errors in parentheses clustered at the classroom level. * p<0.1, ** p<0.05, *** p<0.01.

ONLINE APPENDIX H - PRE-ANALYSIS PLANS

This Appendix compiles our pre-analysis plans in full. 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 H.H1 discusses the instances in which our analyses deviate from what had been pre-specified, followed by subsections H.H2 and H.H3, which present the pre-analysis plans for the baseline and end-line surveys, and for the additional follow-up wave, respectively.

H1. Deviations from pre-analysis plan

When we started this project, back in 2017, we did not anticipate that nudges to both parents and teachers would backfire. We were originally interested in whether nudging parents was enough to improve educational outcomes in this setting, or whether nudging teachers concurrently was necessary for that. The motivation for posing this question was two-fold. First, although the same SMS program had proven effective while nudging parents in Brazil (Bettinger et al., 2020), teachers are much more likely to be absent in Ivory Coast – potentially muting the effects of higher parental engagement, given joint production of children's education. Second, a much higher share of parents is illiterate in the latter – potentially muting the effects of the intervention on parental engagement (also the reason for why we experimented with audio vs. text messages). Upon learning about the large positive effects of nudges to parents and the absence of effects of the joint intervention in our experiment, we decided to look deeper into the data, and realized that our research design and the specifics of how nudges to parents had worked in this context allowed us to study a broader and more fundamental question.

We had 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, Appendix G compiles our main results broken down by grade. Next, we did not pre-register that we would estimate heterogeneous treatment effects, splitting the sample by median teacher attendance or median teacher motivation. Those estimates are, however, used only to shed further light on the mechanisms behind the main treatment effects – which are all estimated using analyses registered in the pre-analyis plan. We also did not specify that we would analyze treatment effects on instruction targeting. That hypothesis only came to mind after we plotted the distributions of test scores for students whom we could not track at the end-line survey, and we thought it would be important to document those effects even though the analysis was not pre-registered.

While we had not originally planned to analyze treatment effects on teachers' time use, or to elicit 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. Having said that, even this additional pre-analysis plan could

not anticipate every angle relevant to our research question before we looked at the data. We highlight two issues in particular. First, we had not specified how exactly we would control for students' and teachers' baseline characteristics in the analyses. Throughout the paper, we resort to ANCOVA specifications, rather than student fixed-effects, to maximize our statistical power to detect even small effect sizes, in instances like test scores. Second, we had not specified how we would handle selective attrition. Whenever we have access to baseline characteristics, we use matching techniques (propensity score matching in the main text, and nearest neighbor matching in Appendix D); otherwise, we bound treatment effects following Lee (2009).

Treatment effects on pre-registered outcomes not analyzed in this paper are discussed in Wolf, Lichand and Deambrosi (2020), Finch, Wolf and Lichand (2020) and Wolf and Lichand (2020).

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 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

		Eduq+ Parents		
		No **(arm A)**	**Yes (text)** **(arm B)**	**Yes (audio)** **(arm C)**
Eduq+ Teachers	**No** **(arm D)**	25 schools Students: 5,000 (survey data for 625) Teachers: 50	12 schools Students: 2,400 (survey data for 300) Teachers: 24	12 schools Students: 2,400 (survey data for 300) Teachers: 24
	Yes **(arm E)**	25 schools Students: 5,000 (survey data for 625) Teachers: 50	13 schools Students: 2,600 (survey data for 324) Teachers: 26	13 schools Students: 2,600 (survey data for 324) Teachers: 26

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 Assesment),
- 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]:

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:

$$Y^j_{m,i,t} = \beta_0 + \beta_1 Parents_m + u_{m,i,t}$$

Where:

- $Y^j_{m,i,t}$: 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 behavioural 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:

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

Where:

- $Y^j_{m,i,t}$: 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)):

$$Y^j_{m,i,t} = \beta_0 + \beta_1 Parents_m + \beta_2 Teacher_m + \beta_3 Parents_m \times Teacher_m + u_{m,i,t}$$

Where:

- $Y^j_{m,i,t}$: 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 carachteristics.

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

$$Y_{m,i,t}^{j} = \beta_0 + \beta_1 Parents_m + \sum \beta_2^{j}\left(Parents_m \times Y_{m,i,t-1}^{j}\right) + \sum \beta_3^{k}\left(Parents_m \times X_{m,i,t-1}^{k}\right) + \beta_4 Y_{m,i,t-1}^{j} + \beta_5 X_{m,i,t-1} + u_{m,i,t}$$

Where:

- $Y_{m,i,t}^{j}$: Outcome variable;

- $Parents_m$: Dummy variable equal to 1 if parents are assigned to receive the intervention (either text or audio), 0 otherwise;

- $Y_{m,i,t-1}^{j}$: Outcome variable at baseline;

- $X_{m,i,t-1}^{k}$: 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:

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

Where:

- $Y_{m,i,t}^{j}$: 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 $\beta_1 = 0$.

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?

➤ 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 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 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

---

[1] World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1096-1. License: Creative Commons Attribution CC BY 3.0 IGO

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 (β_2 and β_1 with different signs):

$$Y_{ck} = \alpha + \beta_1 X_{cj} + \beta_2 X_{cj}^2 + \varepsilon_{ck} \tag{1}$$

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 (β_2 and β_1 with different signs):

   $$Y_c = \alpha + \beta_1 S_c + \beta_2 S_c^2 + \varepsilon_c \tag{2}$$

   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*:

$$Y_{sci} = \alpha + \beta_1 Parents_s + \beta_2 Teacher_s + \beta_3 Parents_s \times Teacher_s + \varepsilon_{sci} \tag{3}$$

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$_s$*: 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:

$$Y_{sci} = \alpha + \beta_1 Treatment_s + \beta_2 X_{si} + \beta_3 Treatment_s \times X_{sci} + \varepsilon_{sci} \tag{4}$$

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.