What is It About Communicating With Parents?*

Nina Cunha[†], Guilherme Lichand[‡], Ricardo Madeira[§], Eric Bettinger[¶]

Sunday 1st October, 2017

Abstract

Informing parents about children's attendance and grades has been shown to significantly raise educational achievement. However, there is no evidence for *why* communication with parents works: is it mainly because information lowers monitoring costs, or is it mainly because it increases the salience of monitoring benefits? The distinction matters – if salience is the key driver behind those effects, nudging could potentially produce even larger impacts, and at much lower cost. To decompose the effects of communication into the two mechanisms, we run a field experiment with 19,300 ninth-graders in São Paulo, Brazil. Math teachers fill-in a platform with information about their students' behavior, and we randomly assign parents to different messages over SMS: some parents receive information provided by teachers, some just receive an awareness message emphasizing the importance of paying attention to that dimension of children's behavior, and others receive no message at all. We find that while communication has large impacts on attendance, test scores and promotion rates, most of the effects are driven by salience: awareness messages improve outcomes by 89-126% of the effects of information. Consistent with the behavioral mechanism, salience effects are larger for least attentive parents; moreover, higher-frequency communication and alternating delivery times significantly increase effect sizes. The optimal combination of features for nudging parents improves students' test scores by 0.33 standard deviation, almost 4-fold the effect of information alone.

[Click here to download the most recent version]

*We would like to thank comments from seminar participants at REAP, University of Toronto and Warwick University. This research was funded by Stanford University's Lemann Center, Itau BBA and the University of Zurich. Any remaining errors are ours.

†Stanford University; e-mail: *nina2012* at *stanford* dot *edu*

‡University of Zurich

§University of São Paulo

¶Stanford University

1 Introduction

Family socio-economic status is a key factor behind variation in children's educational achievement (Woessmann and Hanushek, 2011). While poor and rich families differ across many dimensions, few seem as easy to address as their differential monitoring of children's school performance. In fact, there is increasing evidence that informing parents about children's attendance and grades can significantly raise achievement (Bergman, 2017; Berlinski et al., 2016; Kraft and Dougherty, 2013; Rogers and Feller, 2016). This paper investigates the mechanisms behinds the effects of communicating with parents, decomposing them into those of lower monitoring costs and those of higher salience of monitoring benefits.

The rationale for informational interventions is that of a moral hazard problem between parents and children: as children grow older, their goals may drift increasingly apart from those of their future-oriented parents; moreover, it becomes progressively hard for parents to observe children's effort at school (Cunha and Heckman, 2007; Heckman and Mosso, 2014).[1]

While informing parents lowers monitoring costs – alleviating moral hazard and inducing better school outcomes, in line with parent's objectives –, communication may also increase the salience of monitoring benefits. Consider the following example: when Nina's parents get a message pointing out that she missed school yesterday, while they learn at no cost that her behavior was not in line with their expectations, they may also realize that attendance is an important dimension of their daughter's behavior to which they should attend moving forward. This is particularly relevant in face of the evidence on how poverty captures attention (Mani et al., 2013), and on how, given limited attention, individuals may fail to learn from dimensions they do not notice (Hanna et al., 2014).

Understanding the mechanisms behind communicating with parents matters for two reasons. First, in developing countries, real-time information systems are often absent, making informational interventions expensive.[2] If salience explains most of the effects of communication, similar effects could be achieved at much lower costs, as interven-

[1]To that effect, poor parents in Brazil prefer conditional cash transfers that mandate school attendance – such that parents get notified when students miss over 15% of classes – to unconditional ones (Bursztyn and Coffman, 2012). Consistent with the moral hazard mechanism, such preference disappears when schools systematically share information about their children's attendance.

[2]Vitória da Conquista, a municipality in a poor Brazilian State, spent over USD 700,000 in 2012 placing microchips in students' uniforms, hoping to cut truancy by informing parents immediately when students missed classes. Read more: http://www.bbc.com/news/world-latin-america-17484532

tions to capture attention do not require such information systems. Second, and most importantly, if salience is the key driver of the effects of communication, potentially, the effects of communication could be *much larger*. While informational interventions are constrained by the frequency at which information is available, nudging can be implemented at much higher frequency. Moreover, once the objective is capturing parents' attention, sharing information is just one tool in a much richer toolbox; nudging also allows manipulating other features, such as time of delivery and interactivity.[3]

To decompose the effects of communication into lower monitoring costs and salience of monitoring benefits, we run a field experiment with 19,300 ninth graders across 287 schools in São Paulo, Brazil. In the experiment, over the course of 18 weeks, Math teachers have to weekly fill-in a platform with information about their students' behavior (attendance, tardiness and assignment completion). We randomly assign parents to different messages, shared by the platform over SMS: some parents receive information that the teacher filled in (e.g.: "Nina missed less than 3 classes over the last 3 weeks"), some receive an awareness message, emphasizing the importance of paying attention to that dimension (e.g.: "It is important that Nina attends class every day"), while others receive no message at all (the control group). Because we anticipate that parents' or peer interactions may generate large spillovers, we include a pure control group, which we use as counterfactual in most of our analyses.

In line with previous findings, we find that weekly communication has large impacts on attendance (2.1 percentage points, or about 5 additional classes a year), Math GPA and standardized test scores (0.09 standard deviation) and promotion rates (3.2 percentage points). We find that treated parents ask their children systematically more about school, incentivize studying to a greater extent, and have higher aspirations about their children's making it to college. Children in treated households report engaging in academic and reading activities to a greater extent.

Interestingly, most of the effects are driven by salience: awareness messages improve outcomes by 89-126% of the effects of information. We also estimate heterogeneous treatment effects by parents' attention and willingness to receive information, both according to our baseline phone survey, treating as inattentive those parents with above-median average response times.[4] Consistent with the mechanism, the effects of

[3]It is also worth noting that certain pieces of information may not be as effective in raising perceived returns to monitoring as nudges. Consider a parent who thinks their kid is missing more classes than s/he actually is; information may induce him or her to monitor even less.

[4]Response times are used by cognitive psychologists as a measure of cognitive control, see Mani et al. (2013).

awareness messages are larger for inattentive parents, and are positive even for parents with low willingness to receive information – for whom communication presumably should have had no effects if it was only about reducing monitoring costs.

Is information really unnecessary, or did the experiment convey too coarse information to produce additional effects? To test whether finer-grain information matters, for a subsample of the information treatment group we placed children's metrics in relative terms to the median behavior of their peers (e.g.: "Nina missed less than 3 classes over the last 3 weeks, while most of her colleagues missed between 3 and 5"). Similar to Rogers and Feller (2016), relative information has larger point estimates, but it is still the case that awareness messages amount to at least 68% of the effects of information. While more frequent or finer-grain information could promote larger effect sizes, our information intervention provides an appropriate counterfactual as it resembles the typical structure of school-parents communication campaigns in developing countries (e.g.: Berlinski et al. (2016), which also finds a 0.09 effect size of an SMS information program on students' standardized test scores).

The richness of our data allows us to say more about mechanisms. Parents in our sample have mixed beliefs: in what comes to GPA, the sample is about equally distributed across optimistic, pessimistic and accurate parents. This provides an opportunity to test whether beliefs are indeed the mediating mechanism for the effects of communicating with parents, as Bergman (2017) claims. To test whether this is the case, we start by analyzing treatment effects on parents' beliefs. First, because of the way our messages were framed (attendance over the last 3 weeks, rather than over the last school quarter), information ended up making parents *less accurate* about students attendance (measured over the school quarter at endline), in comparison to both the pure control and the salience groups; nevertheless, information significantly improved school performance, and to the same extent that salience messages. Second, we analyze heterogeneous treatment effects, breaking down our sample according to parents' baseline beliefs. We find that communication leads to positive effects across all categories; even parents accurate at baseline change behavior and see better school performance at endline. Altogether, our findings suggest that parent's beliefs do not play a central role in the behavioral change leading to better school performance. Rather, parents' *engagement* seems to be the key mediating factor.

Spillover effects from communication are substantial: within-classroom control students experience almost as large effects on attendance and GPA, and statistically identical effects on standardized test scores and promotion rates. For this reason – since we have to rely on the pure control group as a counterfactual –, an important concern is whether our results are driven by differences in teacher behavior, induced by requiring them to weekly fill-in a platform with information about their students.[5]

To investigate whether such requirement may drive our results, for a sub-sample of those enrolled we deliver a nudge program instead, reaching parents directly, without informational requirements or the need to involve teachers at all. Such program shares weekly suggestions of activities for parents to do with their children, over SMS. The main challenge of using that sub-sample is that its students were not statistically identical at baseline to those of our main sample.[6] Even though we can control for a wide array of students' and parents' characteristics, one may still worry that students of different profiles could have evolved differentially over time due to unobservable factors that cannot be controlled for.

To deal with this concern, we take advantage of the fact that our program was ran only during the second half of the school year, comparing the evolution of the different sub-samples, before and after the program was introduced. The differences-in-differences strategy estimates the causal effects of the nudge program as long as student outcomes in different sub-samples would have evolved identically on the absence of the program. While the identification assumption cannot be tested, we can test whether the different sub-samples were evolving differentially within the first half of the school year, even before the onset of the program. Results are as follows. Comparing students in nudge schools to those in the pure control group, we find effects of the exact same magnitude to those of communication on standardized test scores (0.09 standard deviation). Using the first quarter as the reference period, different trends across sub-samples only become significant after the program was introduced, dismissing concerns with differential pre-trends due to different baseline characteristics. Last, comparing the nudge program to the salience-only sample (which experienced even larger impacts on attendance and grades), we can rule out that teacher behavior explains more than 1/3 of the treatment effects. All in all, results suggest that of our findings do not stem from differential teacher behavior in treatment schools and can be generalized beyond

[5]There are no other differences across the treatment and pure control groups: (i) sub-samples are balanced across a rage of observable characteristics, (ii) students in pure control schools were enrolled through the same process as those in treatment schools, and (iii) principals of all schools, even in the pure control group, are allowed to use the platform to send monthly communication to parents about school events.

[6]The reason is that the Education Secretariat required us to work in a different region whenever the communication platform was not made available to principals.

this setting.

A final concern is whether the effects of salience messages may die out over time. If parents infer poor performance from salience messages, and if such inference is systematically biased, then parents may realize this over time and stop reacting to communication. To test whether that is the case, we look again at heterogeneous treatment effects with respect to parents' baseline beliefs. We do not find that parents in the salience group become systematically more pessimistic about their children's behavior.[7] What is more, we find that the gap between salience and the pure control groups increases over time both with respect to attendance and GPA. At least within the 6-month length of our study, not only it is not the case that the effects of salience messages die out; they even increase over time.

Based on our findings, a nudge program targeted at capturing parents' attention could potentially have larger effects on educational achievement. Expanding on the design of the SMS program that suggests weekly activities for parents to do with their children, we test whether the combination of different features can deliver higher impacts. We cross-randomize 4 sets of features to test whether SMS frequency, time of delivery, consistency of SMS delivery time, and interactivity impact students' outcomes in line with the behavioral mechanism. Consistent with inattention, higher frequency and alternating delivery times significantly increase effect sizes. As one would expect, however, there are decreasing returns to getting parenting to the top of mind: for attendance, we find evidence that saturation kicks-in beyond 2 messages a week. Also consistent with the mechanism, nudging affects Math and Portuguese standardized test scores to the same extent, whereas the effects of information are mostly confined to the subject it targets.[8] Strikingly, the optimal combination of features increases students' test scores by 0.33 standard deviation, almost 4-fold the effect of the informational intervention. Interactivity, however, had the opposite effects from what we expected in this setting: students whose parents received weekly follow-up questions asking whether they engaged with their child in that week's suggested activity fared significantly worse in both Math and Portuguese GPA. One possible explanation is that parents experienced negative reinforcement from those questions whenever they failed to engage in

[7]Moreover, if parents systematically inferred poor performance from salience messages, increasing monitoring in response, the ratio between the effects of salience and information messages should be equal to 1 for low-performing students, but higher than 1 for high-performing students; however, we do not find evidence that such ratio varies across students' profiles.

[8]For information and salience messages, which were targeted at students' behavior in Math classes, spillover effects on Portuguese standardized test scores are only about half their effect sizes on Math.

the suggested activity in any given week, but this puzzling finding deserves further investigation by future research.

This is the first paper to investigate the mechanisms behind the effects of communication with parents. Our findings challenge previous results about the drivers of the effects of communication. In particular, different from Bergman (2017), which attributes most of the effects of an intensive communication campaign to more accurate beliefs, we show that our effects are not driven by belief updating, but rather by higher parental engagement in response to changes in the salience of monitoring benefits.

Changing poor families' dynamics is daunting: little is known about the intra-household education production function. Low-cost interventions, such as asking parents to study with their children, are often unfeasible, as in poor families many children's educational attainment is equal to or higher than that of their parents (69% of the ninth graders in our sample). While communication interventions have showed promise, scaling them up to developing countries involve cost challenges. This paper suggests that nudge programs can be both cheaper and more effective to alleviate the moral hazard problem between parents and children.

The remainder of the paper is structured as follows. Section 2 provides background on education in Brazil and São Paulo State's education system. Section 3 discusses the rich dataset we draw upon, from administrative data to baseline and end-line survey data. Section 4 introduces our empirical strategy. Section 5 presents our main results, followed by robustness tests. Section 6 presents heterogeneous treatment effects, providing further evidence for the attention mechanism. Section 7 exploits heterogeneity of baseline beliefs to investigate whether parent's accuracy is a key mechanism behind treatment effects. Section 8 introduces the nudge program, our experimental design, and results for students' outcomes. Section 9 concludes the paper.

2 Education in Brazil and São Paulo State

Like most Latin American countries, while Brazil has achieved significant progress over the last 20 years in making basic education universal (over 98% of 7-14 year-olds are enrolled), it still struggles with educational quality.[9] The eight LAC countries that participated in the 2015 PISA exam scored at the bottom of the 65-country distribution, and were outscored even by countries with much lower per capita income. Brazilian 15

[9] 2015 National Household Survey (PNAD), Brazilian Institute for Geography and Statistics (IBGE). It is mandatory for children to go to school from age 6 to 14 (primary school).

year-olds scored 121 points below the OECD average in Math, implying a two-year lag in Math skills.[10]

Education in Brazil is supervised by government offices across municipal, state and federal levels. Municipalities are responsible for providing and regulating early childhood education. State governments are responsible for the provision and regulation of primary and secondary education. The federal government is responsible for the provision of education through federal institutions, and for the regulation of private institutions.

São Paulo is the richest and most populous state in Brazil, and its education system encompasses that largest number of students in the country. According to the Educational Census from the Brazilian Ministry of Education, enrollment corresponded to 5.3 million primary and middle school pupils in 2015. Among those, 700,000 were enrolled in ninth grade, 63% of which served by schools directly administered by the state authority. Despite responding for 40% of the country's GDP, São Paulo also features high inequality in access to education: while wealthy families typically enroll their children in higher-quality private schools, public schools concentrate students from disadvantaged backgrounds. In our sample, over 50% of households earn less than 3 minimum wages (about 900 USD as of September, 2017), within the income range of slum dwellers in the State capital. Such challenges translate into low performance: in 2015, São Paulo State's public middle school students scored 4.7 out of 10 in the Ideb – the national index for de development of basic education, which averages Math and Portuguese standardized test scores, penalizing that average by retention rates –, falling short of its 5.0 target for the final years of primary school.

Poor educational outcomes reflect not only poor infrastructure and low teacher value-added, but also low family engagement in students' school life. While, across OECD countries, 20% of students reported that they had skipped a day of school or more in the two weeks prior to the PISA test, that figure was 48% among Brazilian students. Family engagement is low among public school students: according to the 2015 National Survey of Students' Health, about 1 in every 4 parents do not know if children skip classes, about 1 in every 3 parents do not systematically ask children about their problems, and about 1 in very 2 parents do not regularly ask about homework. In

[10]The Programme for International Student Assessment (PISA) is an ongoing triennial survey that assesses the extent to which 15 year-olds students near the end of compulsory education have acquired key knowledge and skills that are essential for full participation in modern societies. Around 540,000 students completed the assessment in 2015, representing about 29 million 15 year-olds across 72 participating countries.

fact, teachers often cite low family engagement as the leading cause of students' poor school performance in focus groups.

Engaging parents is hard in this setting: notes sent through students' notebooks are an inefficient way to reach busy parents whose children often do not want aware of their behavior in school, and school landlines are often blocked to call cellphones.[11] What is more, most school systems do not have real-time information systems to track students' attendance or school behavior: while teachers pencil down daily records about students, those records are only entered into the centralized school system at the end of the school year.

3 Empirical strategy

This section summarizes our empirical strategy. First, we lay out our experimental design in subsection 3.1, providing details about the platform that teachers had to fill-in weekly in subsection 3.2, and about the intervention timeline in subsection 3.3. Next, subsection 3.4 describes the data for parents' and students' outcomes, followed by descriptive statistics and randomization checks in subsection 3.5. Last, in subsection 3.6 we describe our data on parents' beliefs, a central element of our analyses of heterogeneous treatment effects.

3.1 Experimental Design

To decompose the effects of communication on parents' beliefs, aspirations and behavior, and on students' behavior and educational outcomes, we randomly assign parents to different messages – information, awareness, or no communication (the control group). Because we can only measure outcomes such as standardized test scores and retention rates at the final quarter, we keep the assignment fixed over the course of the experiment, and stratify randomization by an array of students' baseline characteristics, including first quarter's attendance and grades in Math and Portuguese.[12]

[11] Less than 30% of Brazilian households own landlines, while 93.4% of them own cellphones, according to the 2015 National Household Survey (PNAD). While cellphone penetration is high in Brazil, that of internet and smartphone apps is not: about 55% of active lines are not systematically connected (Regional Study Center to Information Society Development, CETIC).

[12] The alternative would have been randomizing communication every week, stratifying the lottery by teachers' weekly inputs in our online platform. Besides the issue of the frequency at which we can measure the outcomes of interest, one may think that students' behavior – and, hence, platform scores – should be responsive to treatment assignments in previous weeks, what would have compromised

Communication is delivered through weekly text messages (SMS) over the course of 18 weeks in the second half of the school year. Content alternates across three dimensions of children's effort that the online platform requires teachers to fill-in—attendance, lateness and assignment completion. We included those dimensions because teachers already measure them weekly (even if on paper), because the Secretariat thought it was important to inform parents about all of them (rather than just about attendance), and because we thought it would be less likely that teachers' usage of the platform would die out over time if they had to alternate across behaviors rather than just replicate the same records they already do on paper every week over the course of 4 months.

We restricted communication to student's behavior in Math classes. The reason is that standardized tests only cover Math and Portuguese, and the Education Secretariat pointed out that Math teachers keep records to a greater extent, and would have an easier time using the online platform, compared to Portuguese teachers.

In order to collect cell phone numbers and baseline data for parents in the control group as well, both treatment and control schools are offered access to the platform for sharing notifications about school events, up to two school events per month. Once an event is registered through principal's login, the system sends two SMS notifications to parents, respectively one week and one day prior to the event.

Our experiment encompasses 287 schools. Whenever there are multiple ninth-grade classrooms in a given school, we include all of them in the experiment. For a first set of schools, we ranodmly assign students within classrooms to each of the following groups:

0. **Control:** No messages.
1. **Information:** Messages with information about child attendance, lateness and assignment completion.
2. **Salience:** Messages with statements raising awareness about school attendance, punctuality and assignment completion.

Comparing *information* and *salience* students to *control* students allows separating the effects of lower monitoring costs from those of higher salience of monitoring benefits.

One concern is that parents may already have (to a reasonable extent) information about their child behavior; the key piece of information missing may be how to place

balance across different groups.

their child relatively to his or her classmates (Rogers and Feller, 2016).[13] To tackle that issue, for a different set of schools we frame information on child behavior relative to the median of their classroom.[14] [15]

3. **Relative information:** Messages with information about child attendance, lateness and assignment completion framed relatively to classroom's median behavior

Teachers and schools are not aware of their assignment, nor of parents' assignment. For the *relative information* arm, the platform computes the class median once the teacher submits all students' information every week. As for the *salience* arm, although teacher will fill in child-level information every week, parents will only receive general information aimed at raising salience about that dimension of children's effort.

Parents of all treatment arms only receive the text message if the teacher fills in the platform that week. This is true even for the salience group, in order to avoid confounding treatment effects with teachers' non-compliance. After teachers have filled the platform until Sunday of each week, parents receive the message on the following Tuesday, according to their treatment status, as showed in the table below. The content of the messages is simple and clear, and messages across treatment arms were designed to match number of characters as close as possible.

Salience	**Individual Information**	**Relative Information**
For a good school performance, it is important that Guilherme doesn't miss school for no reason.	According to the information registered by the teacher in the system the past 3 weeks, Eric missed less than 3 classes.	In the past 3 weeks, Nina missed less than 3 classes. In his class, most of the students didn't miss any class.

For salience messages, we change the wording of the messages only slightly every cycle, so as to prevent triggering spam-avoiding behavior by parents. For the full script of messages sent for each treatment arm, see Appendix A.

Because we worry about the possibility that spillovers coming from peer effects, contamination across parents or teacher effects may bias downwards any differences

[13] Rogers and Feller (2016) convey information relative to the classroom modal behavior, using child-level information as a placebo, across US schools.

[14] We thought that the median behavior (e.g.: "most students in Nina's class missed less than 3 classes in the previous 3 weeks") was much easier for parents to understand than the mode, which was graphically conveyed through letters in Rogers and Feller (2016).

[15] We also survey parents at baseline about their best guess for their child's attendance, so as to investigate heterogeneity of treatment effects by the accuracy of parents' beliefs.

across groups, we include another set of schools in which all students were assigned to the control group. We call this subsample *pure control* group. If on the one hand we stratified assignment so as to ensure that students in this group were statistically identical with respect to baseline characteristics to students assigned to the interventions, on the other hand in the pure control group teachers do not weekly fill in the platform. The reason was to avoid deception, and since we expected no or very few teachers to fill in the platform if they were aware that no information would be shared with their students' parents. There are no other differences across the treatment and pure control groups: students in pure control schools were enrolled through the same process as those in treatment schools, and principals of all schools, even in the pure control group, are allowed to use the platform to send monthly communication to parents about school events.

While relying on this group as a counterfactual rules out spillovers, it also brings about potential concerns with teacher effects, since feeling in the platform may have induced teachers to think they were being monitored by the school system or by their students' parents. To deal with this concern, we include another subset of schools for which we deliver a nudge program instead, reaching parents directly, without informational requirements or the need to involve teachers at all. Such program is inspired by READY4K (York et al., 2017), sharing weekly suggestions of activities for parents to do with their children, over SMS. This intervention is also randomized within classrooms within this subsample.

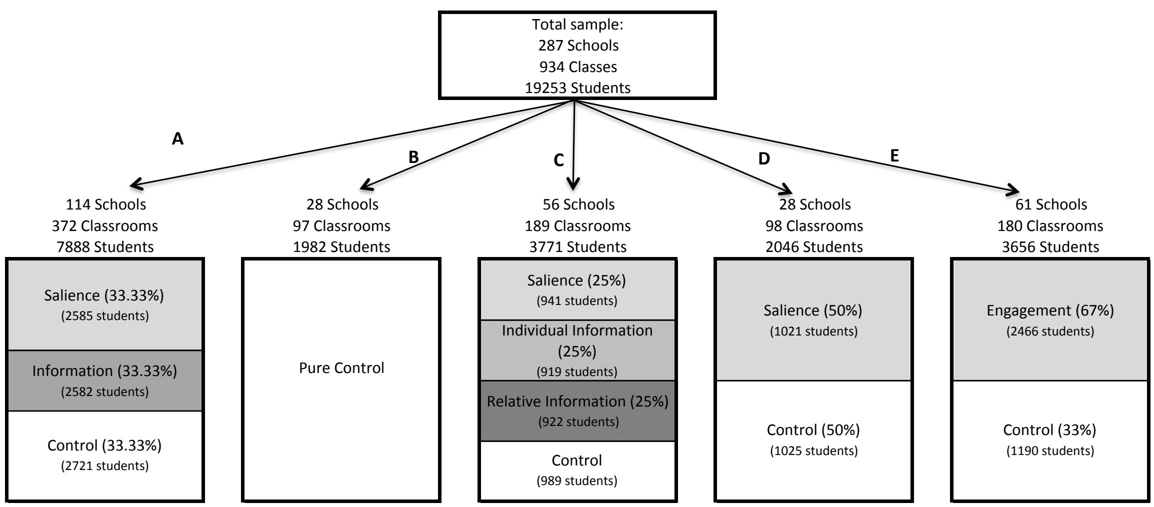
4. **Engagement:** Messages with suggestions of non-curricular activities to do with their children

There are two relevant features of this subset of schools. First, they were not offered the possibility of sending monthly communication to parents, since in some schools Math teachers also handled this activity (delegated by principals) and we want to preclude any teacher effects. Second, their students are not statistically identical at baseline to those in our other subsamples. The reason is that the Education Secretariat required us to work in a different region of the State whenever the communication platform was not made available to principals, where students were relatively low-performing at baseline.

We tackle this issue by taking advantage of the fact that our program was ran only during the second half of the school year, comparing the evolution of the different subsamples, before and after the program was introduced. The differences-in-differences strategy estimates the causal effects of the nudge program as long as student outcomes in different sub-samples would have evolved identically on the absence of the program. We discuss this strategy in greater detail in subsection 5.

In order to provide this group a proper comparison to pin down the magnitude of teacher effects, we include a final subset of schools, within which students were assigned to either the salience or the control groups – but not to the information group –, making the sample distribution analogous. An additional advantage of this group is to rule out concerns with interactions between the information and salience groups, since it allows estimating the effects of salience on the absence of the former.

Randomization is performed in two steps. First, schools are randomly assigned to each of the 5 different subsamples, determining the treatment arms made available at each school. Second, students are randomized within-class to each treatment arm:



To summarize, subsamples A through C allow separating the effects of information and salience; subsample B allows a counterfactual for estimating spillovers; subsample D allows estimating the effect of salience without spillovers from information; and subsample E allows pinning down the extent of teacher effects.

We randomize assignment in two steps. In the first step, we stratify the assignment of schools to subsamples based on three variables: school average first quarter Math scores in the Education Secretariat's internal quarterly assessment, school average truancy rate, and share of parents enrolled in our program. In the second step, we stratify the assignment of students to groups within class based on the first quarter Math scores in the Education Secretariat's internal quarterly assessment.[16]

The design choice for subsamples A through D reflects power calculations accounting for the hypothesis of interest. In the case of subsample E, the sample reflects the demand of the Education Secretariat.

3.2 Teacher platform

A web-platform was created specifically to this project and was designed in a simple and intuitive way such that schools could easily manage it.[17] Math teachers from treatment schools were oriented to fill in the platform every week with that week's dimension of students' behavior: attendance, lateness or assignment completion, as shown in the table below. Teachers filled information regarding student behavior on each dimension considering the past three weeks.[18] The system requires teachers to fill in information for all students.

Attendance	**Tardiness**	**Assignment Completion**
1. Missed more than 5 classes	**1.** Was late for more than 5 classes	**1.** Did no complete any of the assignments
2. Missed 3 to 5 classes	**2.** Was late 3 to 5 classes	**2.** Completed less than half of the assignments
3. Missed less then 3 classes	**3.** Was late for less then 3 classes	**3.** Completed more than half of the assignments
4. Did not miss any class	**4.** Was not late for any class	**4.** Completed all the assignments

Each week teachers receive a text message, reminding them which dimension they should fill in that week. Teachers who miss a week receive an alert, stating that they did not fill in the platform that week and encouraging them to do so in the following week. Principals receive motivational messages, encouraging them to engage teachers in the program, as well as alerts in case the usage by teachers in the school is low.

There is perfect compliance with the randomization protocols, since our implementing partner (MGov Brasil) had full control over enrollment (all data had to be entered by teachers into its system, and assignment was conditional on enrollment), and over platform's outputs and messages ultimately sent to parents and guardians.

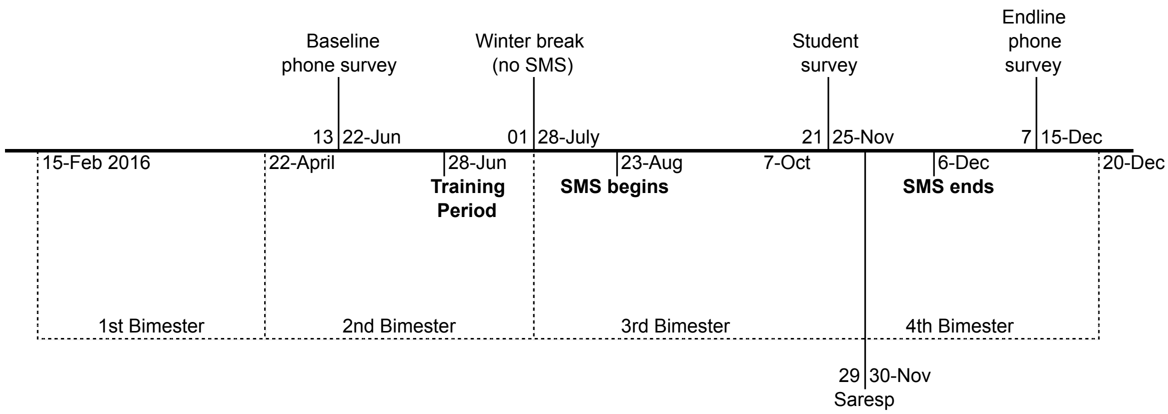
[16]Not all students take this test (which is not mandatory). For students with no scores, we predict their scores based on a simple linear regression using all baseline covariates, and then stratify based on predicted scores.

[17]60% of Brazilian schools have access to internet, although typically only with very limited bandwidth – typically below 4 mbps, shared across staff and all student computers, if any. The web-platform is very low-bandwidth, and can be accessed by principals and teachers from any computer or smartphone, even outside of school.

[18]Students have around 6 Math classes per week.

3.3 Timeline

Teachers began to fill the platform on the week of June 24th. The school year in Brazil runs from February to December and is divided in 4 quarters, with a winter break in July. Parents were exposed to the program during 6 months of the academic year, until the first week of December.



3.4 Data

In order to enroll in the program, parents had to provided informed consent through a registration form, in which they where requested to inform their cell phone number, their relationship with the student, gender, age, race, income bracket, education as well as their children gender and age.

Through our online platform, we have weekly records of teachers's inputs about their students, alternating across attendance, tardiness and assignment completion.

In what comes to parents, we surveyed those enrolled through automated phone surveys (Interactive Voice Response, IVR) at baseline and endline to collect self-reported parenting practices, parents' beliefs about their children, as well as parents demand for information. The baseline survey was conducted on the week of June 13th and the endline survey took on the weeks of December 5th and 12th.

We also conducted a face-to-face survey with enrolled students at the end of the intervention (December), through which we collected data on parents' participation, student's activities, values and aspirations, as well as students' social and emotional skills.

The São Paulo Education Secretariat provided quarterly data on student attendance, grades and retention status in 2016. According to official guidelines, all teachers assign numeric integer grades ranging from 0 to 10, with a passing grade set at 5 points for all disciplines. Attendance is recorded in percentage points (0–100 interval). Last, we draw upon data from SARESP (System of School Performance Evaluation of the State of São Paulo), the Education Secretariat's yearly standardized test, applied across all State schools.[19]

3.5 Sample and balance tests

Table 1 presents averages for students' and parents' baseline characteristics by treatment arm, along with p-values of a joint test for the null hypothesis of whether averages are equal across groups. Panel A displays baseline characteristics for students. Around 50% are girls, 40% are brown or black, and the average age is 14.7, within the range of the appropriate age for the ninth grade. Panel B shows parents' characteristics. 76% of those enrolled are mothers, at their early 40s; strikingly, 69% have educational achievement no greater than middle school, what means that, for 2/3 of our sample, children have progressing in school at least as far as their parents did. Together with the figure of 59% of families earning monthly less then 3 minimum wages, the table illustrates the low socioeconomic status of parents in our sample and the challenges associated with the most straightforward interventions, such as advising parents to work together with their children in homework assignments.

[Table 1]

Column (6) shows the p-values for the F-tests of joint equality of averages for each variable across the four treatment arms. The sample is balanced: across 17 variables, only for age differences are statistically significant at the 10% level – which is consistent with chance, and even in that case it is fair to say the difference is a precisely estimated zero.

66% of the almost 30,000 parents invited to participate enrolled in the program. Table 2 analyzes selection in opt-in. For parents who did not enroll, we only have student characteristics available from administrative records – gender, age, Math and Portuguese baseline attendance and grades, and status of participation in *Bolsa Familia*,

[19]All students in grades 1^{st}, 3^{rd}, 5^{th}, 7^{th}, 9^{th} of primary school and the 3rd (final) year of high school are tested on their knowledge of Mathematics and Portuguese.

Brazil's flagship conditional cash transfer. If there are systematic differences across those enrolled and those who are not in what comes to those characteristics, then one might be concerned about whether our results would generalize if the intervention was scaled-up to the whole school system.

[Table 2]

According to Table 2, parents who joined the program were less likely to benefit from the conditional cash transfer, and their children had statistically higher attendance and grades compared to those of parents who did not enroll in the program. Since assignment is randomized conditional on enrollment, selection does not bias our results. Having said that, one might still worry about generalizability. To that point, since any educational intervention that requires parents' consent is expected to have imperfect compliance, the relevant parameter should be the average treatment effect on the treated, which is captured by our estimates. Moreover, even if one were interested in the average treatment effect on the absence of selection, we can still re-weight observations by the inverse opt-in probability to gauge the extent to which results would change due to heterogeneous treatment effects (See Table C.5 in the Appendix).

3.6 Beliefs

Before we turn to the results, it is useful to introduce our data on parent's beliefs, since we rely on it extensively for estimating heterogeneous treatment effects.

Parents were asked at baseline to give their best estimate of how many times their child had missed Math classes over a typical three-week period. Their answers were then compared to administrative records on students' attendance over the first quarter, scaled for three weeks. Parents had to choose one out of four brackets over the phone survey (no absences; 1 to 2 absences; 3 to 5 absences; or more than 5 absences).[20] Parents were also asked to give their best estimate of their child's performance in Math classes. Again, parents had to choose one out of of four categories (below average; adequate; good; or very good). In the Brazilian school system, GPA ranges from 0 to 10, with 5 as the passing grade. Parents' answers were compared to administrative records for the first quarter: below average was determined as a GPA below 5, adequate as 5-6; good as 7-8 and very good as 9-10.

[20] See Appendix B for the full script.

Figure 1 illustrates the distribution of parents' beliefs at baseline and children's actual outcomes. Panel A overlays the distributions of parents' answers and administrative records, while Panel B documents the gap between the two, such that positives values indicate optimistic parents – for attendance, those who believe their children are less absent than they actually are, for GPA, that their kids are doing better than they actually are.

[Figure 1]

Overall, parents are optimistic about their children attendance: similar to Bergman (2017), most parents think that their kid miss less classes than they actually do. Interestingly, however, the same is not true for GPA: the sample is about evenly distributed across optimistic, accurate and pessimistic parents. We take advantage of that variation when teasing out the mechanisms behind the effects of communication.

Last, we repeat the same exercise at the endline survey with parents, asking them about attendance and grades over the last quarter. In particular, we are interested in whether communication affects accuracy at endline. Note the important change with respect to how we ask about attendance at endline – for the whole quarter, rather than for the last 3 weeks. The reason why we ask about it in this way is because by that time students were supposed to have been handed in their final scorecards. If communication increases the likelihood that parents learn about the content of the scorecards, then the right metric to track would be their knowledge about their child's overall absences, rather than the scaled version for the last three weeks.

4 Results

This section starts by presenting manipulation tests in subsection 4.1, followed by our main results in subsection 4.2. Next, subsection 4.3 presents findings for more demanding counterfactuals to salience effects: relative information and extreme messages, both of which are more likely to make parents update their beliefs. In face of our null result, we discuss power calculations in subsection 4.4 to tackle the issue of whether our design would have allowed detecting meaningful differences between salience and information effects. Last, we assess the robustness of our results to different concerns: in subsection 4.5 we investigate whether the lack of difference between salience and information is driven by the interaction of the two treatments; in subsection 4.6 we document the extent to which effects may be driven by differential behavior of teachers required to fill-in the platform; and in subsection 4.7 we investigate whether treatment effects are short-lived.

4.1 Manipulation tests

To begin with, if teachers did not weekly fill-in the platform with students' information, or if parents did not even acknowledge receiving text messages from the school, then there would be no hope that our experiment could allow us detecting the effects of interest. For this reason, we start by looking at these output measures. Figure 2 displays statistics for platform usage and receipt of text messages.

[Figure 2]

Over the course of the 18 weeks, 66% of teachers inputted students' information through the platform on a typical week. Since this figure was slightly lower for sub-samples A and C relative to sub-sample D, students assigned to the information treatment are associated with a 2 p.p. lower messaging rate. In the Appendix, we assess the robustness of our results to dropping observations from schools with the highest platform usage for the salience-only subsample (D), so as to equalize usage rates across treatment arms.

At the endline surveys, we asked parents whether they had received text messages from the school, and asked students whether they knew their parents were getting such text messages. While 46% of parents in the control group acknowledge receipt of text messages (principals could send up to two notifications a month about school events to *all* parents, even in the pure control group), that figure is 90% across treatment groups – close to the expected 100%, and statistically different from the control group. Meanwhile, 74% of students across treatment arms acknowledged their parents received text messages from the school, as opposed to 40% in the control group . Since over 50% of parents' reported a cell phone number for their kids at enrollment, this result is not just a mechanical artifact of sharing parents and children sharing the same cell phone, but rather hints at some form of communication between parents and children being triggered by the text messages.

4.2 Main results

To decompose the effects of communication into those of lower monitoring costs and those of higher salience of monitoring benefits, we estimate the following equation:

$$Y_{sci} = \alpha + \beta_1 Salience_{sci} + \beta_2 Info_{sci} + \beta_3 Control_{s=treated,ci} + \sum_{k=1}^{K} \gamma_k X_{scik} + \theta_s + \varepsilon_{sci} \quad (1)$$

where Y_{sci} denotes the outcome of interest for student i in classroom c of school s ; $Control_{s=treated,ci} = 1$ for the control group within treatment schools, and 0 otherwise – pure control schools stand for the reference category, the omitted indicator variable –; X_{scik} is a matrix of student's covariates, including students' gender, age and race, their attendance and GPA prior to the intervention, and their parents' or guardians' gender, age, race, income and education; θ_s is a randomization stratum fixed-effect; and ε_{sci} is a zero-mean error term. We cluster standard-errors at the classroom level. The share of the effects of *information* that could be accounted for by *salience* effects is computed from the ratio $\frac{\hat{\beta_1}}{\hat{\beta_2}}$.

Table 3 shows the results for fourth-quarter's attendance in Math classes, Math GPA, promotion status, and Math scores in SARESP, São Paulo State standardized test.[21]

[Table 3]

First, focusing on the estimates for the effects of *information*, even though average attendance on the control group is already quite high – in particular, because *Bolsa Familia*'s conditionality requires attendance 85% or higher – it is still the case that communication increases it by 2.1 percentage points, equivalent to attending five additional classes in the academic year. Information increases Math GPA by 0.071 standard deviation, similar to what has been found elsewhere (Berlinski et al., 2016). Counter to the worry that Math tests might be graded differentially by the teacher herself, effect sizes are about the same (0.107 standard deviation) when it comes to standardized test scores, graded centrally by external officers. We also find a significant and sizeable positive effect of information on the likelihood of being promoted to high school – a 2.6

[21] Only students with non-missing values for all outcomes and control variables are included in the analysis. Descriptive statistics and balance tests are shown in Tables C.1, C.2, (C.3) and (C.4) in the Appendix.

percentage-point effect size, even though the control mean is above 90% (partly because it is quite expensive for the State to fail students).

Second, and most strikingly, comparing those estimates to those of awareness messages, we find that *salience* can account for most of the effects of information: the ratio of coefficients is never lower than 89%, and salience point estimates are sometimes larger – up to 126% of information effects. Information and salience coefficients and are never statistically different at the 10% significance level, and, even considering the lower bound of 90% confidence intervals for the ratio between salience and information effects, the former would never account for less than 60% of the effects of information.

Going beyond averages, figure 3 displays the fourth-quarter distribution of Math attendance, Math GPA and Math standardized test scores for the different groups.

[Figure 3]

Panels A through C show that the effects of information and salience percolate to the whole distribution of students, but are especially visible in Panel C for students around the median of the pure control distribution, whose test scores are more pronouncedly shifted to the right. For attendance and standardized test scores, Kolmogorov-Smirnov tests significantly reject the hypothesis that salience and pure control distributions are the same. Across all outcomes, the test fails to reject differences between information and salience distributions at conventional significance levels.

Exploring parents' and students' endline survey data, we find that treated parents ask their children systematically more about schools, incentivize studying to a greater extent, and have higher aspirations about their children's making it to college. Children in treated households report engaging in academic and reading activities to a greater extent. We summarize those results in Appendix D, in Tables D.1 through D.4. Results inform the theory of change of the program, whereby communication positively affects parents' behavior and aspirations, then students' behavior, and finally students' attendance, grades, test scores and promotion rates.

We also consider heterogeneous treatment effects by gender within the theory of change's framework: boys experienced larger treatment effects from the interventions, with higher impacts on attendance, GPA, promotion rates and standardized test scores (Table D.5). Consistently, male students parents' behavior and aspirations are affected to a significantly greater extent, as well as boys' behavior (Tables D.6 through D.8). [22]

[22] For Appendix D, all regressions were estimated using a smaller sub-sample, which excludes ob-

Are gender differences driven by differential parental responses to treatments, or by baseline differences in performance across boys and girls that generate ceiling effects for the latter? To answer that question, we match boys and girls by their baseline characteristics, and re-estimate the regressions above controlling for that propensity score matching.

[FORTHCOMING]

Last, Table C.5 in the Appendix shows results for re-weighting observations by the inverse probability of opting-in the program. Treatment effects are very similar to those showed in table 3 – if anything, slightly larger –, and all conclusions from the main analyses remain unchanged.

4.3 Can more informative messages do better?

Is information really unnecessary, or did the experiment convey too coarse information to produce additional effects? This subsection considers more demanding counterfactuals for salience effects. First, to test whether finer-grain information matters, for a sub-sample of the information treatment group we communicate children's metrics in relative terms to the median behavior of their peers.

Table 4 shows results using the same specification in equation 1, but adding an indicator variable for the relative information treatment. In this table, all ratios and cross-coefficients tests refer to differences between salience and relative information point estimates.

[Table 4]

Similar to Rogers and Feller (2016), relative information effect sizes are larger for some outcomes – notably, for standardized test scores, even though point estimates are actually lower for promotion rates. Nevertheless, it is still the case that awareness messages amount to at least 68% of the effects of information.

While we cannot rule out that even finer-grain information might promote larger effect sizes, our information intervention provides an appropriate counterfactual as it

servations with missing values for any of the outcomes within the theory of change. Table C.2 shows balance tests for this sub-sample.

resembles the typical structure of school-parents communication campaigns in developing countries (e.g.: Berlinski et al. (2016)), which also finds a 0.09 effect size of an SMS information program on students' standardized test scores).

Second, we estimate heterogeneous treatment effects by the share of weeks in which teachers' tried to communicate extreme messages – filling in no stars, what is equivalent to missing most classes or assignments over the previous three weeks –, since those would make it more likely that parents would update their beliefs.

[FORTHCOMING]

4.4 Is there enough statistical power to detect differences?

[FORTHCOMING]

4.5 Are effects driven by interactions between the two treatments?

Does a combination of spillovers across parents, peer effects and teacher effects – all coming from the information treatment – affect those receiving salience messages? Since the main counterfactual we rely on is *not* within-class control group students, this is a relevant concern. If that is the case, treatment effects should be lower within the sub-sample of schools in which there was no information treatment.

To test this hypothesis, we investigate wether salience effects are smaller in sub-sample D, for which only salience messages – and no information – were delivered. We estimate the following model:

$$\begin{aligned} Y_{sci} = \alpha + \beta_1 Salience_{sci} + \beta_2 Info_{sci} + \beta_3 Control_{s=treated,ci} \\ + \beta_4 Salience_{sci} \times \varphi_{s \in D} + \varphi_{s \in D} + \sum_{k=1}^{K} \gamma_k X_{scik} + \theta_s + \varepsilon_{sci} \end{aligned} \tag{2}$$

where $\varphi_{s \in D} = 1$ if the school belongs to sub-sample D (50% salience, 50% control), and 0 otherwise.

If it is the case that the effect of salience is lower on the absence of information, we would expect $\beta_4 < 0$. Table 5 presents the results.

[Table 5]

It is not the case that salience effects are lower on the absence of information; conversely, its effect are even larger within those schools. Once we correct for the fact that, in sub-sample D, the frequency of teachers who weekly filled-in the platform was slightly higher than that of other sub-samples, salience effects are no longer statistically larger within sub-sample D, but it is still the case that they are nowhere lower on the absence of information. [23]

4.6 Are effects driven by differential teacher behavior?

Table 6 shows results for the within-class control group as a counterfactual. Spillover effects from communication are substantial: within-classroom control students experience almost as large effects on attendance and GPA, and statistically identical effects on standardized test scores and promotion rates.

[Table 6]

Since we have to rely on the pure control group as a counterfactual –, an important concern is whether our results are driven by differences in teacher behavior. There are no other differences across the treatment and pure control groups: (i) sub-samples are balanced across a rage of observable characteristics, (ii) students in pure control schools were enrolled through the same process as those in treatment schools, and (iii) principals of all schools, even in the pure control group, are allowed to use the platform to send monthly communication to parents about school events. Despite all commonalities, requiring teachers to weekly fill-in a platform with information about their students may have made them feel they were being monitored, and changed their behavior. For inference about the mechanisms behind communicating with parents to be generalizable beyond our setting, it is crucial to understand the extent to which impacts would remain when parents are nudged directly, on the absence of a platform for teachers.

[23] Differences in frequency are very relevant since, even in the salience group, parents only receive text messages on weeks in which teachers fill-in the platform. To test if higher SMS frequency drives higher salience effects in sub-sample D, we build a new sample in which we equalize the frequency teachers filled-in the platform across sub-samples. To maximize the number of observations we keep in the analysis, we do so by dropping all observations from 7 sub-sample D classrooms for which teachers had filled-in the platform all the 18 weeks, and from 27 classrooms from the sub-sample C (25% salience, 25% info, 25% relative info, 25% control) where average frequency was lower. In this new sample, the average number of times the teacher filled-in the platform is equal across all sub-samples. We then replicate our main results as well as the above analyses for interactions between treatments. Results, shown in Tables H.1 and H.2, are very similar to those of the main text.

To answer that question, for a sub-sample of those enrolled we deliver a nudge program instead, reaching parents directly, without informational requirements or the need to involve teachers at all. Such program shares weekly suggestions of activities for parents to do with their children, over SMS. The program is based on READY4K! (York et al., 2017); see Section 7 for more details.

The main challenge of using that sub-sample is that its students were not statistically identical at baseline to those of our main sample. The reason is that the Education Secretariat required us to work in a different region whenever the communication platform was not made available to principals, and students were relatively low-performing at baseline in this region. Even though we can control for a wide array of students' and parents' characteristics, one may still worry that students of different profiles could have evolved differentially over time due to unobservable factors that cannot be controlled for.

To deal with this concern, we take advantage of the fact that our program was ran only during the second half of the school year, comparing the evolution of the different sub-samples, before and after the program was introduced. The differences-in-differences strategy estimates the causal effects of the nudge program as long as student outcomes in different sub-samples would have evolved identically on the absence of the program. While the identification assumption cannot be tested, we can test whether the different sub-samples were evolving differentially within the first half of the school year, even before the onset of the program. Results are as follows.

We estimate the following model:

$$Y_{scit} = \alpha + \theta_t Post_t + \theta_j Engagement_{sci} + \beta Engagement_{sci} \times Post_t + \varepsilon_{scit} \qquad (3)$$

where Y_{scit} denotes the outcome of interest for student i in classroom c at school s on quarter t; $Post_t = 1$ if $t \geq 3$, and 0 otherwise. Pure control schools stand for the reference category (omitted indicator variable); and $\epsilon_{i,c,t}$ stands for robust standard errors.

Figure 4 displays the quarterly evolution of math attendance and GPA for the pure control group and the engagement treatment. Visibly, students in sub-sample E had significantly worse performance at baseline.

[Figure 4]

Outcomes were moving in parallel for the two groups before the intervention (during the first two quarters); during the last two quarters, however, outcomes for the engagement treatment start trending upward, reversing a declining trend for attendance within pure control schools and fully catching up in grades already by the third quarter.

Figure 5 shows quarterly coefficients for the differences-in-differences estimate of model 3, using the first quarter as the reference period.

[Figure 5]

Panels A and B showcase no statistically significant difference across groups before the onset of the program. For attendance, this difference becomes significant and increases to a 2.3 p.p. and 2.4 p.p. respectively on the second and third quarters. For GPA, the difference becomes significant on the third quarter (of the order of 0.14 s.d.), and marginally insignificant on the last quarter even though engagement's effect size (0.09 s.d.) is the same we find for the main sample (although less precisely estimated from a much smaller sample).

Last, Table 7 compares the nudge program to the salience-only sample (D), contrasting experimental results for the former with differences-in-differences estimates for the latter. Since both samples were receiving one message a week and no information, the only difference between them are potential teacher effects.[24]

[Table 7]

Comparing point estimates across Panels A and B, we can rule out that platform-induced teacher behavior explains more than 1/3 of treatment effects. All in all, results suggest that the bulk of our findings do not stem from teacher effects in treatment schools and can be generalized beyond this setting.

4.7 Are effects short-lived?

A final concern is whether the effects of salience messages are short-lived. Effects could die out over time if parents infer poor performance from salience messages, but gradually realize they were misled. If that is the case, then salience messages should make parents more pessimistic about their children's school performance. Moreover, under the biased

[24] Another difference is that in pure control schools principals could send up to two monthly communications to parents about school events. If those were relevant for treatment effects, we would overestimate teacher effects from the comparison.

inference hypothesis, the ratio between the effects of salience and information messages should close to 1 for low-performing students (since both treatments would affects parents' beliefs the same way), but greater than 1 for high-performing students (since awareness messages would tend to increase monitoring relative to reassuring information messages).

To test whether that is the case, we first look at heterogeneous treatment effects by parents' baseline accuracy. Table 8 presents the results.

[Table 8]

We do not find evidence that salience messages make parents systematically more pessimistic than information messages. Coefficients of both salience and information treatments on parent's beliefs are the same across all pessimistic, accurate and optimistic parents.[25] Moreover, salience makes parents who were pessimistic at baseline significantly more accurate (hence, less pessimistic) at endline in what comes to their children's Math GPA.

Next, Table 9 presents results for heterogeneous treatment effects by students' baseline performance, splitting the sample between below- and above-median students, according to first quarter's Math GPA. We rely on baseline performance rather than teachers' inputs to the platform because student performance after the onset of the program is endogenous to treatment status.

[Table 9]

The ratio between salience and information treatment effects is higher for above-median students in what comes to promotion rates and standardized test scores, in line with the prediction from the biased inference hypothesis, but the opposite is true in what comes to attendance and GPA.[26] Altogether, results do not support the idea that the salience treatment works by making parents systematically more pessimistic about their children's performance.

Last, we look at the dynamics of treatment effects on attendance and GPA, taking advantage of the fact that we have access to quarterly data for administrative outcomes. Figure 6 present the results.

[25]The negative treatment effects on accuracy about attendance are linked to the mismatch between the time span at which we conveyed information about attendance ("over the last 3 weeks") and that for which we could verify attendance at endline (over the last quarter); see Section 6.

[26]Table E.4 on the Appendix shows similar results for heterogeneous effect by students' baseline performance, but considering students' baseline GPA instead of attendance.

[Figure 6]

We find that the gap between salience and the pure control group increases over time both with respect to attendance and GPA. At least within the 6-month length of our study, not only it is not the case that the effects of salience messages die out; they even increase over time.

5 Inattention

Our results point to salience of monitoring benefits as the main driver of the effects of communicating with parents. A natural story for why that may be the case is limited attention, particularly relevant in face of the evidence on how poverty captures attention (Mani et al., 2013), and on how, given limited attention, individuals may fail to learn from dimensions they do not notice (Hanna et al., 2014).

To document whether our findings are consistent with the attention mechanism, we estimate heterogeneous treatment effects by parents' attention and willingness to receive information at baseline. If inattention is the key mechanism behind the effects of salience, we expect treatments effects to be larger amongst inattentive parents. If effects are driven by attention rather than information, we expect treatment effects even among parents with low willingness to receive information at baseline.

We deem parents "inattentive" if they manifest slow reaction times in our baseline phone survey.[27] Specifically, we split the sample according to below- or above-median response times to the first question of the baseline phone survey, which asked parents about their child's attendance over the last 3 weeks (see Appendix B).

Table 10 presents heterogeneous treatment effects by splitting the sample according to this indicator variable. The lower sample size reflects the fact that we can only use parents who answer our baseline phone survey in this table.

[Table 10]

Treatment effects are substantially higher for inattentive parents. Amongst parents with above-median response times, salience effects on GPA and information effects on standardized test scores are almost 0.2 standard deviation, significantly higher than

[27] We follow Lichand and Mani (2017). In the cognitive psychology literature, reaction times are often used as a measure of cognitive performance; see Mani et al. (2013).

those of parents with below-median response times, for whom the treatments have no significant effects on standardized test scores.

As an alternative measure of inattention, we define an indicator variable equal to 1 if a parent requests questions to be repeated at any point throughout the baseline phone survey, and 0 otherwise.

[FORTHCOMING]

Next, willingness to receive information is also measured at the baseline survey. Parents were asked at baseline about their interest in receiving information about their child's school attendance, given the following options: no interest, some interest, or great interest (see Appendix B). We define low willingness to receive information as an indicator variable equal to 1 if a parent expressed no or some interest in receiving information about school attendance, and 0 otherwise.

Tables 11 presents heterogeneous treatment effects by splitting the sample according to this indicator variable. The lower sample size reflects the fact that we can only use parents who answer our baseline phone survey in this table.

[Table 11]

First, the willingness to receive information indicator (WTR) indeed seems to capture parents demand for information: while low-WTR parents do not update beliefs about children's attendance in response to text messages, those with high-WTR do.[28] Second, both salience and information treatments have positive and statistically significant effects even for low-WTR parents. Third, and most strikingly, the ratio of salience to information effects is actually systematically higher for parents with high WTR, which is consistent with attention being the primary mechanism behind the effects of communication. The reason is that, in line with Chassang et al. (2012), parents with higher demand for information should be those who exert higher effort to acquire it within the setting of the randomized control trial. Salience effects are magnified among those parents to a greater extent than information effects, highlighting the complementary nature between attention and decentralized information acquisition by parents.

[28] The negative treatment effects on accuracy about attendance are linked to the mismatch between the time span at which we conveyed information about attendance ("over the last 3 weeks") and that for which we could verify attendance at endline (over the last quarter); see Section 6.

6 Beliefs

The richness of our data allows us to say more about mechanisms. Parents in our sample have mixed beliefs: in what comes to GPA, the sample is about equally distributed across optimistic, pessimistic and accurate parents. This provides an opportunity to test whether beliefs are indeed the mediating mechanism for the effects of communicating with parents, as Bergman (2017) claims.

To test whether this is the case, we start by analyzing treatment effects on parents' beliefs. If beliefs are a key mediating factor for our results, then communication should make parents more accurate, and effects should be concentrated on optimistic parents, who presumably under-monitor their children within the moral hazard framework.

For eliciting beliefs, parents were asked to provide their best estimate of how many times their child had missed school during three weeks prior to the baseline phone survey – to match the frequency at which we report attendance –, choosing the bracket that most closely matched their prior beliefs (no absence; 1 to 2 absences; 3 to 5 absences; or more than 5 absences; see Appendix B). Accuracy is computed by approximating absences over the first-quarter to expected absences over a three-week period. Parents with guesses in the right bracket were considered accurate, those with guesses in a higher (lower) bracket were considered pessimistic (optimistic).[29]

We start by analyzing whether communication made parents more accurate about children's school behavior. Table 8 presents heterogeneous treatment effects on endline accuracy by splitting the sample according to parents' baseline accuracy with respect to their child's Math GPA.

[Table 8]

Results in Panel A suggest communication made parents *less accurate* with respect to Math attendance, across all baseline accuracy categories; significantly so for the effects of information on pessimistic parents. Such negative effects on accuracy are probably linked to the mismatch between the time span at which we conveyed information about attendance ("over the last 3 weeks") and that for which we could verify attendance at endline (over the last quarter). Conversely, when it comes to Math GPA – for which we never shared information over text messages –, Panel B shows that communication seems to *increase* accuracy amongst parents who were not accurate at baseline; significantly so for the effects of salience on pessimistic parents, further

[29] Results are robust to different definitions of accuracy.

corroboration for the attention mechanism. Since we have shown both treatments to positively affect student outcomes to the same extent, the fact that they affect parent's beliefs in opposite ways is suggestive that the latter is not a key mediating factor for treatment effects.

Still, it could be the case that accuracy at endline is a misleading variable, due to the time-span reporting issues: parents may have ultimately become more accurate with respect to both recent attendance and GPA. For this reason, we estimate heterogeneous treatment effects with respect to *belief changes*. If beliefs are central to treatment effects, then treatment effects should be proportional to the extent of belief updating.

[FORTHCOMING]

Moving forward, we analyze whether effects are driven by optimistic parents at baseline. Table 12 presents heterogeneous treatment effects on students' outcomes by splitting the sample according to parents' baseline accuracy with respect to their child's Math GPA.

[Table 12]

All coefficients are less precisely estimated due to both smaller sample size due to high non-response rates in phone surveys and splitting the sample according to baseline accuracy. Results show that children of both pessimistic and optimistic parents experience positive effects of communication; in fact, even accurate parents experience positive and significant impacts on promotion rates.

Last, if effects were fundamentally driven by changes in monitoring effort in response to updated beliefs, the only pattern consistent with this framework would be a an *increase* in monitoring amongst optimistic parents accompanied by a *decrease* in monitoring amongst pessimist parents, as the former presumably under-monitored students at baseline, whereas the latter presumably over-monitored. Table 13 presents heterogeneous treatment effects on parent's behavior by splitting the sample according to parents' baseline accuracy with respect to their child's Math GPA.

[Table 13]

Once again, coefficients are less precisely estimated due to both smaller sample size due to high non-response rates in phone surveys and splitting the sample according to baseline accuracy. Results show positive effects of communication on parental involvement in academic activities, on incentivizing school activities and on talking to the child. Both pessimistic and optimistic parents report higher engagement, and even accurate parents change behavior.

Altogether, our findings suggest that parent's beliefs do not play a central role in the behavioral change leading to better school performance. Rather, parents' *engagement* seems to be the key mediating factor.

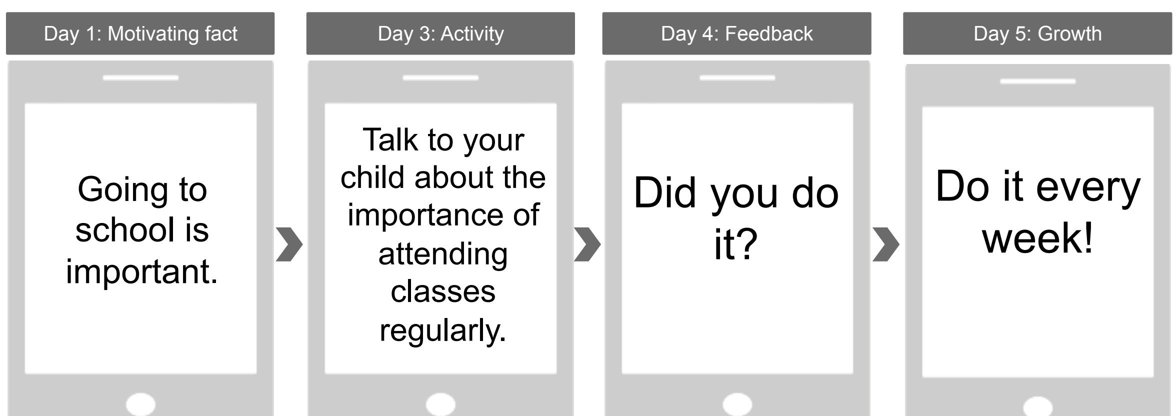
7 Optimal SMS nudging

Based on our findings, a nudge program targeted at capturing parents' attention could potentially have larger effects on educational achievement. The reason is that nudging allows manipulating several features, such as interactivity or time of delivery, which informational interventions cannot – partly because they are constrained by the frequency at which student performance is tracked.

While several papers document significant behavioral changes in responses to nudges over text messages (from literacy, as in Jukes et al. (2017), to adherence to health treatment, as in Raifman et al. (2014)), less is know about the optimal design of those SMS campaigns. Raifman et al. (2014) finds that only shorter text messages were effective for increasing adherence to anti-malarial treatment in Ghana, and Karlan et al. (2016) documents that highlighting savings goals and financial incentives was effective for increasing savings in Bolivia, Peru, and the Philippines, while framing savings in the gains or losses domain had no effects. Despite those recent contributions, policy-makers are still faced with many open questions when designing nudge programs. What it the optimal frequency of texting, so as to most effectively capture attention? At what point do saturation effects kick-in? At what time of the day should messages be sent? Should parents get messages always at the same time? Does interactivity make content more effective? The answers to those questions could provide the 'nuts and bolts' for designing nudge program over text messages by government agencies and international organizations.

We tackle those questions throughout this section. Using the same nudge program introduced in subsection 4.6, we cross-randomize different characteristics of the design of a typical SMS campaign. We assess impacts on students' outcomes of the following features: (i) frequency (1, 2 or 3 messages a week), (ii) SMS delivery time (3 p.m. or 7 p.m., which we deem "work hours" or "off-work hours", respectively), (iii) consistency of delivery times (rotating weekly across 2 p.m., 3 p.m., and 4 p.m. to all messages within that week for those contacted during work hours, and between 6 p.m., 7 p.m., and 8 p.m., for those contacted off-work hours), and (iv) interactivity, given by a feedback flow that weekly asks parents whether they engaged with their child in the activity suggested on that week. Our hypotheses are that higher frequency, off-work hours and alternating delivery times, and interactivity should all be associated with larger effect sizes, consistent with the attention mechanism.

The figure below displays a stylized sequence for a parent assigned to 3 messages a week and interactivity. Those assigned to the group without interactivity do not receive the feedback message on day 4 of every week. Those assigned to 2 messages a week do not receive the growth message on day 5 of every week. Last, those assigned to 1 message a week receive only the activity message, on day 3 of every week. Because all parents assigned to a positive number of messages receive a suggestion of activity, we are able to cross-randomize all features. The control group is held fixed for all features, since SMS delivery time, consistency and interactivity can only be varied amongst those receiving messages.



Assignment is randomized at the student level within classroom, comprising a sample of 3,656 students across 61 public schools. Our research design is summarized by the figure below.

Frequency		
Group	**Definition**	**Sample size**
Control	0 message / week	1218
Treatment A	1 message / week	812
Treatment B	2 messages / week	813
Treatment C	3 messages / week	813
SMS Delivery Time		
Group	**Definition**	**Sample size**
Control	N/A	1218
Treatment A	Work hours	1219
Treatment B	Off-work hours	1219
Consistency of Delivery Time		
Group	**Definition**	**Sample size**
Control	N/A	1218
Treatment A	Varying	1219
Treatment B	Constant	1219
Interactivity		
Group	**Definition**	**Sample size**
Control	N/A	1218
Treatment A	Interactive	1219
Treatment B	Passive	1219

The decision to assign $1/3$ of the sample to the control group is informed by power calculations for hypotheses testing across all SMS features. Treatment 1A ($1/3$ of the remaining subject pool) receives 1 message a week, a suggestion of activity for parents to do along with their children (delivered on Wednesdays). Treatment 1B ($1/3$ of the remaining subject pool) receives 2 messages a week, a motivating fact with information about how an activity is linked to children's development (delivered on Mondays) and the same suggestion of activity (delivered on Wednesdays). Treatment 1C ($1/3$ of the remaining subject pool) receives 3 messages a week, the same motivating fact (delivered on Mondays) and suggestion of activity (delivered on Wednesdays), additionally to a growth message that reinforces habit formation (delivered on Fridays). Treatment 2A ($1/2$ of the treated sample) receives messages during work hours (centered on 3 p.m.), while Treatment 2B (1/2 of the treated sample) receives messages off-work hours (centered on 7 p.m.). Treatment 3A (1/2 of the treated sample) receives all messages during the same time of the day (either 3 p.m. or 7 p.m.), while Treatment 3B (1/2 of the treated sample) receives messages at alternating times (at the scheduled time, 1 hour before and 1 hour after, rotating on a 3-week cycle); all messages within a week are delivered at the same time.[30] Last, Treatment 4A (1/2 of the treated sample) receives a feedback message (delivered on Thursdays) asking whether the parent engaged with the child in the activity suggested on the day before – to which parents can reply 'yes' or 'no' –, while Treatment 4B (1/2 of the treated sample) does not receive feedback messages.

Due to concerns with spillovers within classroom – motivated by our findings in subsection 4.6 –, here we also rely on the pure control group as a counterfactual for estimating treatment effects. Once students in the sub-sample receiving the nudge program are not statistically identical to those in the pure control group (see subsection 4.6), we estimate a differences-in-differences model (as in equation 3)) to document the effects of each treatment arm, relying on Seemingly Unrelated Regressions (SUR) to test hypothesis involving estimates of different arms. Tables 14 and 15 show results for attendance and GPA. We consider outcomes for both Math and Portuguese, since communication in the nudge program was not specific to Math.

[Table 14]

[Table 15]

Results are as follows. In what comes to frequency, more messages systematically increase treatment effects: we detect significantly higher effect sizes of more weekly messages on attendance and GPA. While saturation seems to kick-in beyond 2 messages a week for attendance, there is no evidence that the same is true for GPA, for which coefficients for 3 weekly messages are 50% larger for Math and over 100% larger for Portuguese. Strikingly both subject are affected to the same extent, whereas information effects are mostly contained to the subject it targets.[31] In turn, SMS (target) delivery times did not significantly affect attendance or GPA, perhaps because we cannot

[30]Since it is not possible to guarantee or even track actual delivery times, which can be affected by signal availability and by whether the phone is turned off at that time, what we mean is target delivery time, when MGov Brasil's system sends messages to the the end users. Under good signal, messages are typically delivered within minutes.

[31]For information and salience messages, which were targeted at students' behavior in Math classes, spillover effects on Portuguese standardized test scores are only about half their effect sizes on Math; see Table G.2

guarantee (or even track) actual delivery times, possibly affected by signal availability and by whether the phone is turned off at that time. Next, alternating delivery times significantly increased Portuguese attendance, even though they did not significantly affect Math attendance or GPA for either subject. Last, counter to our hypothesis, interactivity worked backwards in this setting: it significantly decreased decreased Portuguese attendance, and GPA for both subjects – coefficients for that group are between 1/2 and 1/3 of those for that without interactivity. One possibility is that parents felt negative reinforcement from those questions when they failed to engage in the suggested activity in any particular week, but understanding when interactivity may harm or help capturing attention is an interesting topic for future research.

Last, we estimate effect sizes of all 24 possible combinations of SMS features on Math attendance and GPA, and plot them against the benchmark of the effect sizes of the information treatment. Figure 7 presents the results, ordered by effect sizes.

[Figure 7]

Panels A and B allow two main conclusions. First, it is not the case that *any* communication positively affect student outcomes: for at least some combos, effect sizes are zero or very close to zero. Second, attention can do much more than information alone: for most combos, effect sizes are larger than those of information (dashed line at 0.02 p.p. for attendance and 0.09 std for GPA), and effect sizes for "optimal combos" are many times larger, up to 0.33 standard deviation when it come to Math GPA – almost 4-fold the effect of information.

7.1 Relation to the literature

This is the first paper to test the hypothesis that behavioral mechanisms may explain why communicating with parents works, decomposing its effects into lower monitoring costs and higher salience of monitoring benefits. Our study builds on different recent experimental evaluations of school communication program, as well as on a growing body of evidence that suggests parents play a crucial role in shaping their children's behavior and school performance (Barnard, 2004; Houtenville and Conway, 2008; Nye et al., 2006).

Differences in parental inputs are viewed as an important cause of intergenerational inequality (Becker and Tomes, 1979), and family socio-economic status is a key factor behind variation in children's educational achievement (Woessmann and Hanushek, 2011). While poor and rich families differ across many dimensions, few seem as easy to address as their differential monitoring of children's school performance.

A growing education literature suggests parents can affect students' educational behaviors and success when they receive proper information. Bergman (2017) finds that sending parents SMS when their child was missing assignments resulted in significant gains in tests scores, GPA, and measures of student engagement. Kraft and Dougherty (2013) show that frequent teacher-to-parent phone calls increased student engagement as measured by homework completion, in-class behavior, and in-class participation during a summer school program. Bergman and Chan (2017) report a decrease in course failures and absenteeism as a result of alerting parents through SMS about their child's missed assignments, grades and class absences. Berlinski et al. (2016) show that students of treated parents in Chile–who received information on absenteeism, grades, and student behavior–had significantly higher math grades, attendance, better behaviors, and a lower probability of failing the grade at the end of the year.

Informational interventions are mainly based on the hypothesis that there is a moral hazard problem between parents and children, with high monitoring costs (Cunha and Heckman, 2007; Heckman and Mosso, 2014). To that effect, Bursztyn and Coffman (2012) show that poor parents in Brazil prefer conditional cash transfers that mandate school attendance – such that parents get notified when students miss over 15% of classes – to unconditional ones. Consistent with the moral hazard mechanism, such preference disappears when schools systematically share information about their children's attendance.

Alternatively, effects could be driven by behavioral biases – as we show in this paper. Behavioral interventions had already been shown to systematically improve students' outcomes. York et al. (2017) find that a SMS program affected the extent to which parents engaged in home literacy activities with their children, as well as parental involvement at school, which translated into student learning gains in some areas of early literacy. Castleman and Page (2015) report positive effects of a texting program for recent high-school graduates, designed to incentivize college enrollment. Rogers and Feller (2016) show preliminary evidence that sending messages to increase parents' salience about good school behavior was effective at increasing attendance. The contribution of this paper is to show that information and nudge programs share a common denominator: their effects are driven by getting parenting to the top of mind amongst inattentive parents.

As we argue in this paper, the distinction matters for two reasons. First, providing timely and accurate information about children's behavior requires integrated systems and customized communication, which can be quite costly, particularly in developing countries, where real-time information systems are usually not available; conversely, simply nudging to raise salience does not require any such systems in place. Second, and most importantly, if salience is the key driver of the effects of communication, the effects of communication could be much larger. In fact, we show that combining different features of SMS communication allows for potentially much larger effect sizes on students' attendance and GPA. Without the need to invest in real-time information systems, nudging can deliver larger effect sizes at lower costs.

In this vein, our study contributes to a rich literature that investigates cost-effective alternatives to improving educational outcomes in developing countries. As Ludger et al. (2015) and others have shown, students in developing countries learn much less than students of the same age, or in the same grade, learn in OECD countries. Researchers and policy-makers in these regions have been searching for evidence on how to increase enrollment and attendance at scale, and on how to simultaneously improve quality of human capital formation (Glewwe and Muralidharan, 2015). While different approaches have been explored– from cash transfers (Baird et al., 2011; Barrera-Osorio et al., 2011; Behrman et al., 2009; Mo et al., 2013; Schultz, 2004) to scholarships (Blimpo, 2014; Friedman et al., 2011; Kremer et al., 2009; Li et al., 2014) to increasing the quantity and quality of teachers (Chin, 2005; Duflo et al., 2015; Urquiola, 2006; Urquiola and Verhoogen, 2009) and school grants (Das et al., 2013; Lucas and Mbiti, 2014; Newman et al., 2002; Pop-Eleches and Urquiola, 2013; Pridmore and Jere, 2011) –, few have managed to improve student outcomes cost-effectively, through easily scalable interventions.

Lastly, our study also contributes to the still scarce literature on behavioral educational interventions. A growing number of studies studies interventions to tackle parents' inertia and affect parents' routine behavior, (Avvisati et al., 2013; Banerji et al., 2013; Benhassine et al., 2015; Harackiewicz et al., 2002; Kraft and Rogers, 2015), including text messages, email reminders, and letters targeted at parents and students (Castleman and Page, 2015; Hoxby et al., 2013; Jensen, 2010). While the field of behavioral economics has been successfully applied to many areas, so far Education has received comparatively less attention (Lavecchia et al., 2014). Given that investments in children's human capital are crucially about inter-temporal decisions – typically plagued by all sorts of behavioral biases –, the is huge potential for behavioral interventions to improve educational investments . Lavecchia et al. (2014) reviews the recent and growing literature of interventions designed to overcome behavioral barriers in education.

8 Concluding Remarks

We find that weekly communication has large impacts on attendance (2.1 percentage points), test scores (0.09 standard deviation) and promotion rates (3.2 percentage points). Sharing information has no or small additional effects: salience improves outcomes by 89%-126% of the effects of information.

Effects are consistent with inattention: they are larger for parents who are most inaccurate at baseline, most inattentive, and positive even for those with lower willingness to receive information. Consistent with the mechanism, effects of communication are larger for higher-frequency and alternating delivery times. Having said that, delivery on or off work hours did not significantly impact outcomes, and interactivity led to puzzling lower impacts.

Different from Bergman (2017), we found that beliefs are not central to behavior change: communication leads to positive effects even when accuracy responds differently to different treatments. Moreover, positive effects extend beyond optimistic parents; even parents accurate at baseline change behavior and see better school performance at endline. Altogether, our findings suggest that parent's beliefs do not play a central role in the behavioral change leading to better school performance. Rather, parents' *engagement* seems to be the key mediating factor.

The optimal nudge program increased test scores by 0.33 standard deviations, almost 4-fold the effect of information alone. For every dollar invested in such program, Brazilian school systems could save about USD 12.50 just due to lower retention rates – every 9th grader facing retention costs R$ 4,000 to the State, and the program decreases retention rates by 3.2 percentage points for a cost of R$ 10 per student per year.

In what comes to welfare, there are two potential caveats to what otherwise seem to be very positive impacts of nudging. First, we do not know what gets "displaced" from parents' attention when parenting becomes top of mind. If higher attention to children's education leads parents to attend less to children's health, for instance, it would not even be clear that investments in children's human capital experience a net increase from a nudge program. Second, parents have to pay the costs of decentralized monitoring costs whenever the State decides to nudge rather than share centralized information, and the sum of those costs may turn out to be higher under decentralized monitoring. Although we cannot completely dismiss those concerns, it is not hard to believe a story in which not much gets displaced by those nudges, since poor parents are at home with children typically only after work, in the evening, when they are quite cognitively depleted. Moreover, the potential increase in monitoring costs going from centralized to decentralized monitoring is presumably small relative to the increase in individual returns to higher parental engagement, especially given the high wage premium of schooling in developing countries like Brazil.

Our results are consistent with "learning through noticing" (Hanna, Mullainathan and Schwartzstein (2014). Interestingly, that paper claims that such mechanism may provide "insights into educational interventions, suggesting they are useful not only for new technologies but also for existing technologies when there are indications that people are insufficiently attentive to key aspects of production. It also suggests ways of improving these interventions: there can be large benefits from moving away from just providing more data to helping individuals understand the relationships in the data they already have." – Hanna, Mullainathan and Schwartzstein (2014, p. 1347). As parenting is perhaps the most ancient human capital production technology, our results provide additional evidence that their claim may be justified.

References

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E. (2013). Getting parents involved: A field experiment in deprived schools. *Review of Economic Studies*, 81(1):57–83.

Baird, S., McIntosh, C., and Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, 126(4):1709–1753.

Banerji, R., Berry, J., and Shotland, M. (2013). The impact of mother literacy and participation programs on child learning: Evidence from a randomized evaluation in india. *Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).*

Barnard, W. M. (2004). Parent involvement in elementary school and educational attainment. *Children and youth services review*, 26(1):39–62.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., and Perez-Calle, F. (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in colombia. *American Economic Journal: Applied Economics*, 3(2):167–195.

Becker, G. S. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, 87(6):1153–1189.

Behrman, J. R., Parker, S. W., and Todd, P. E. (2009). Schooling impacts of conditional cash transfers on young children: Evidence from mexico. *Economic development and cultural change*, 57(3):439–477.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. (2015). Turning a shove into a nudge? A "labeled cash transfer" for education. *American Economic Journal: Economic Policy*, 7(3):86–125.

Bergman, P. (2017). Parent-child information frictions and human capital investment: Evidence from a field experiment. *Mimeo.*

Bergman, P. and Chan, E. W. (2017). Leveraging technology to engage parents at scale: Evidence from a randomized controlled trial. Technical report, CESifo Group Munich.

Berlinski, S., Busso, M., Dinkelman, T., and Martinez, C. (2016). Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in chile. *Unpublished Manuscript*.

Blimpo, M. P. (2014). Team incentives for education in developing countries: A randomized field experiment in benin. *American Economic Journal: Applied Economics*, 6(4):90–109.

Bursztyn, L. and Coffman, L. C. (2012). The schooling decision: Family preferences, intergenerational conflict, and moral hazard in the brazilian favelas. *Journal of Political Economy*, 120(3):359–397.

Castleman, B. L. and Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115:144–160.

Chassang, S., Padro-i Miquel, G., and Snowberg, E. (2012). Selective trials: a principal-agent approach to randomized control trials. *American Economic Review*, 102(4):1279–1309.

Chin, A. (2005). Can redistributing teachers across schools raise educational attainment? Evidence from Operation Blackboard in India. *Journal of development Economics*, 78(2):384–405.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. Technical report, National Bureau of Economic Research.

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

Duflo, E., Dupas, P., and Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools. *Journal of Public Economics*, 123:92–110.

Friedman, W., Kremer, M., Miguel, E., and Thornton, R. (2011). Education as liberation? Technical report, National Bureau of Economic Research.

Glewwe, P. and Muralidharan, K. (2015). Improving school education outcomes in developing countries: evidence, knowledge gaps, and policy implications. *University of Oxford, Research on Improving Systems of Education (RISE).*

Hanna, R., Mullainathan, S., and Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, 129(3):1311–1353.

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., and Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, 6(1):689–733.

Houtenville, A. J. and Conway, K. S. (2008). Parental effort, school resources, and student achievement. *Journal of Human resources*, 43(2):437–453.

Hoxby, C., Turner, S., et al. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014).

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Jukes, M., Turner, E., Dubeck, M., Halliday, K., Inyega, H., Wolf, S., Zuilkowski, S., and Brooker, S. (2017). Improving literacy instruction in kenya through teacher professional development and text messages support: A cluster randomized trial. *Journal of Research on Educational Effectiveness*, 10,:449–481.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, 62:3393–3411.

Kraft, M. A. and Dougherty, S. M. (2013). The effect of teacher–family communication on student engagement: Evidence from a randomized field experiment. *Journal of Research on Educational Effectiveness*, 6(3):199–222.

Kraft, M. A. and Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, 47:49–63.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, 91(3):437–456.

Lavecchia, A. M., Liu, H., and Oreopoulos, P. (2014). Behavioral economics of education: Progress and possibilities. Technical report, National Bureau of Economic Research.

Li, T., Han, L., Zhang, L., and Rozelle, S. (2014). Encouraging classroom peer interactions: Evidence from chinese migrant schools. *Journal of Public Economics*, 111:29–45.

Lichand, G. and Mani, A. (2017). Cognitive droughts. *Working Paper*.

Lucas, A. M. and Mbiti, I. M. (2014). Effects of school quality on student achievement: Discontinuity evidence from kenya. *American Economic Journal: Applied Economics*, 6(3):234–263.

Ludger, W. et al. (2015). *Universal Basic Skills What Countries Stand to Gain: What Countries Stand to Gain*. OECD Publishing.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *science*, 341(6149):976–980.

Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S., and Brinton, C. (2013). School dropouts and conditional cash transfers: Evidence from a randomised controlled trial in rural china's junior high schools. *The Journal of Development Studies*, 49(2):190–207.

Newman, J., Pradhan, M., Rawlings, L. B., Ridder, G., Coa, R., and Evia, J. L. (2002). An impact evaluation of education, health, and water supply investments by the bolivian social investment fund. *The World Bank Economic Review*, 16(2):241–274.

Nye, C., Turner, H., Schwartz, J., and Nye, C. (2006). Approaches to parent involvement for improving the academic performance. *Campbell Systematic Reviews*, 4.

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *The American Economic Review*, 103(4):1289–1324.

Pridmore, P. and Jere, C. (2011). Disrupting patterns of educational inequality and disadvantage in malawi. *Compare: A Journal of Comparative and International Education*, 41(4):513–531.

Raifman, J., Lanthorn, H., Rokicki, S., and Fink, G. (2014). The impact of text message reminders on adherence to antimalarial treatment in northern ghana: A randomized trial. *PLOS ONE*, 9:https://doi.org/10.1371/journal.pone.0109032.

Rogers, T. and Feller, A. (2016). Reducing student absences at scale. *Unpublished paper.*

Schultz, T. P. (2004). School subsidies for the poor: evaluating the mexican progresa poverty program. *Journal of development Economics*, 74(1):199–250.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural bolivia. *The Review of Economics and Statistics*, 88(1):171–177.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *The American Economic Review*, 99(1):179–215.

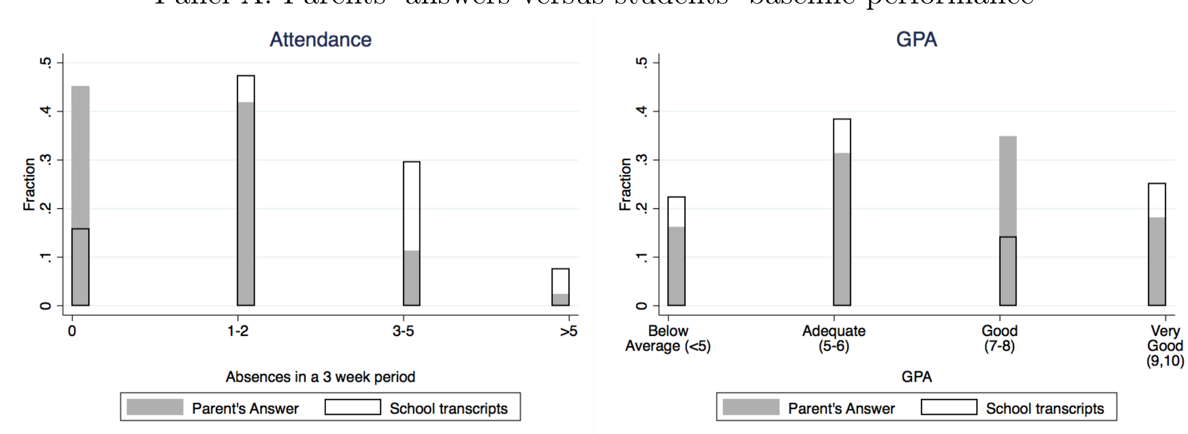
Woessmann, L. and Hanushek, E. A. (2011). The economics of international differences in educational achievement. *Handbook of the Economics of Education*, 3:89–200.

York, B. N., Loeb, S., and Doss, C. (2017). One step at a time: The effects of an early literacy text messaging program for parents of preschoolers. *Mimeo.*

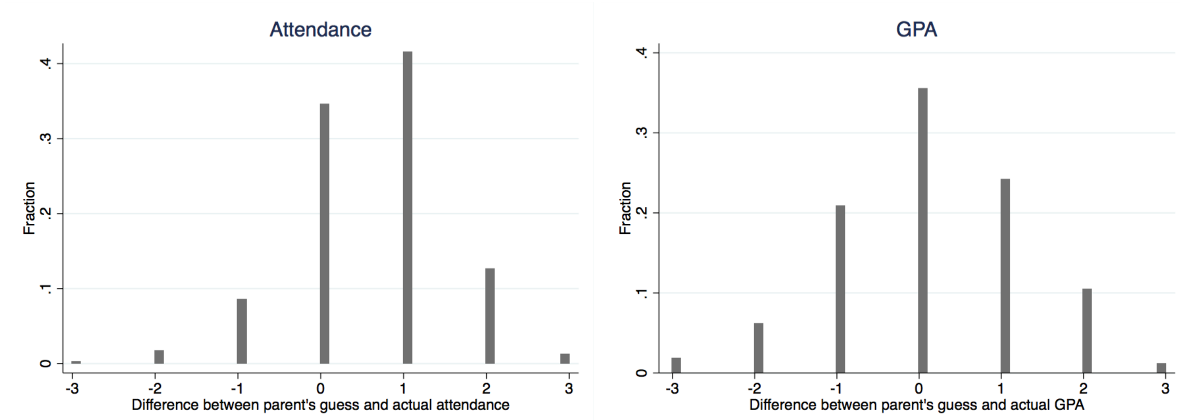
Figures

Figure 1: Parents' accuracy wrt their child's baseline attendance and GPA

Panel A: Parents' answers versus students' baseline performance



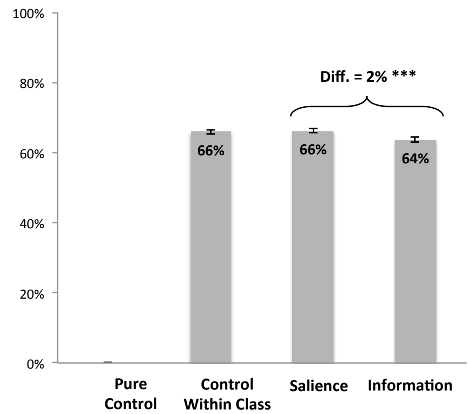
Panel B: Difference parents' answer and baseline performance



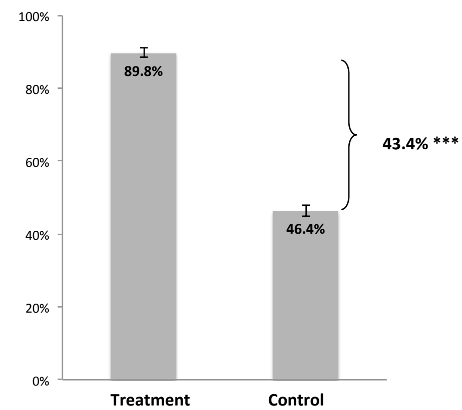
Note: Parents were asked at baseline to give their best estimate on how many times their child misses math classes on a period of three weeks, as well as on their performance in math classes. Data was then crossed with administrative records. Four categories were available for parents' answers on attendance (missed 0; 1-2; 3-5; more than 5). Administrative data register data on attendance on a quarterly basis (period of $\sim$ 9 weeks) and was divided by 3 to validate parents' answers. Four categories were available for parents' answers on performance (below average; adequate; good; very good). The GPA has a 10 point scale, where 5 is the passing grade. Parents' answers below average was determined as a GPA below 5, adequate as 5-6; good as 7-8 and very good as 9-10. Panel A shows parents' answers and school transcripts. Panel B shows the difference between parents' answers and students' performance. Note that the value zero indicates parents were accurate, positive values indicate they were pessimist and negative values indicate they were optimistic.

Figure 2: Manipulation Tests

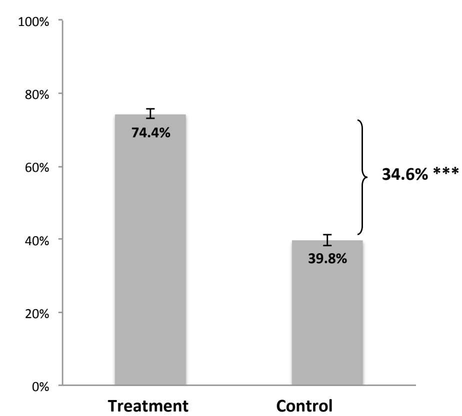
Panel A: Average number of times teachers filled the platform by treatment status during the 18 week period



Panel B: Did parents acknowledge receipt of text messages?



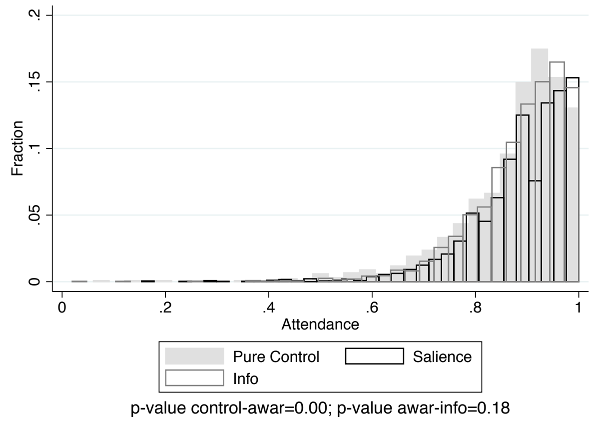
Panel C: Did students know their parents were receiving text messages?



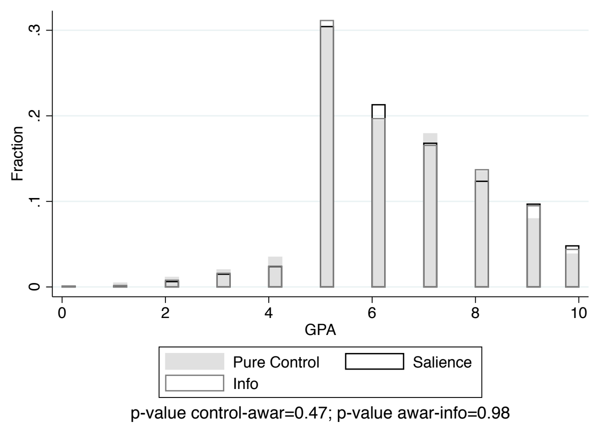
Note: 90% confidence interval. The difference between categories was estimated through a simple regression including fixed effect for strata, and standard errors were clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. In Panel A, data from teachers' platform were used, while Panel B and C used data from parents and students endline survey, respectively.

Figure 3: Distribution Effects

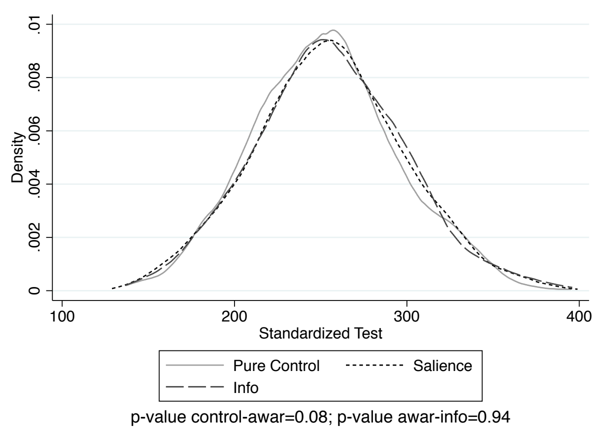
Panel A: Attendance



Panel B: GPA



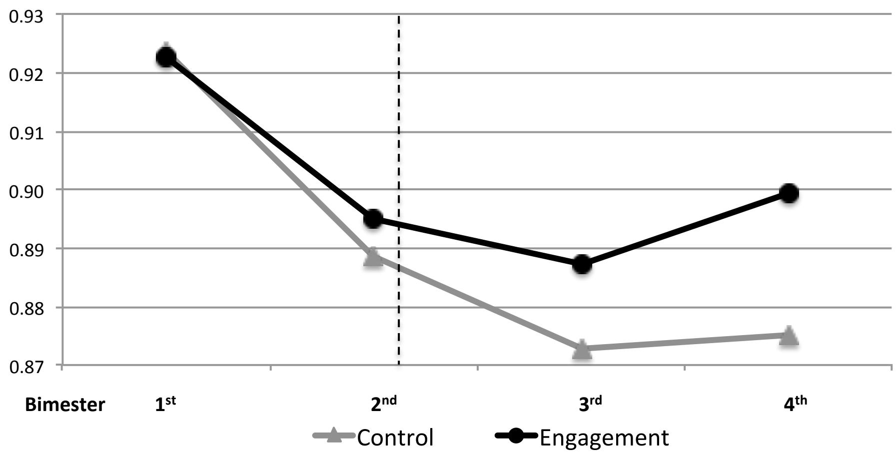
Panel C: Standardized test



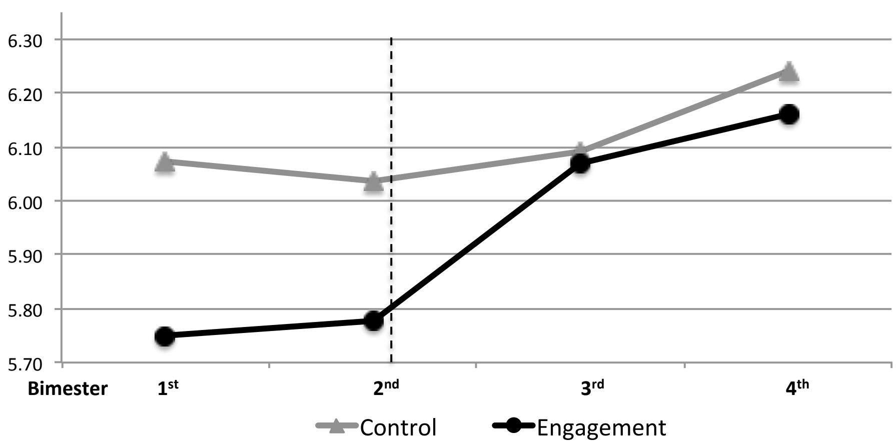
Note: Panels A, B and C show the effect across the distribution of students' attendance, GPA and standardized test for each treatment arm. Data was extracted from administrative records. Attendance is recorded in percentage points (0-1 interval). The GPA has a 10 point scale, where 5 is the passing grade. The standardized test (Saresp) has a 400 scale, where zero is the minimum score. No controls were included. A Kolmogorov-Smirnov equality-of-distributions test was performed to test equality of distribution between the "salience" and "control" groups; and the 'info" and "salience" groups. P-values reports result of the test.

Figure 4: Theory-based nudging program - effect by quarter

Panel A: Attendance



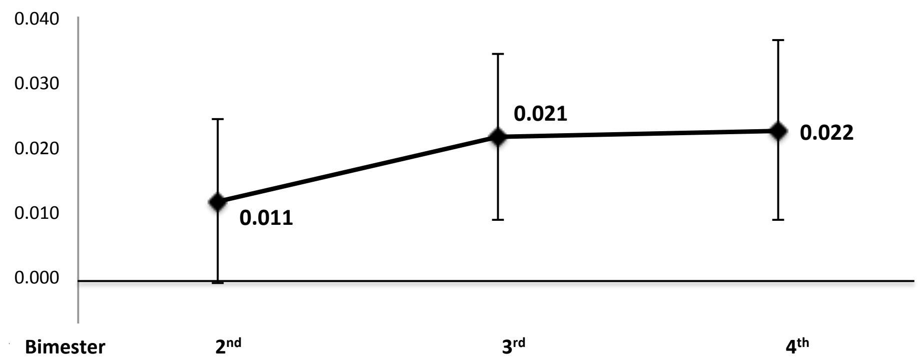
Panel B: GPA



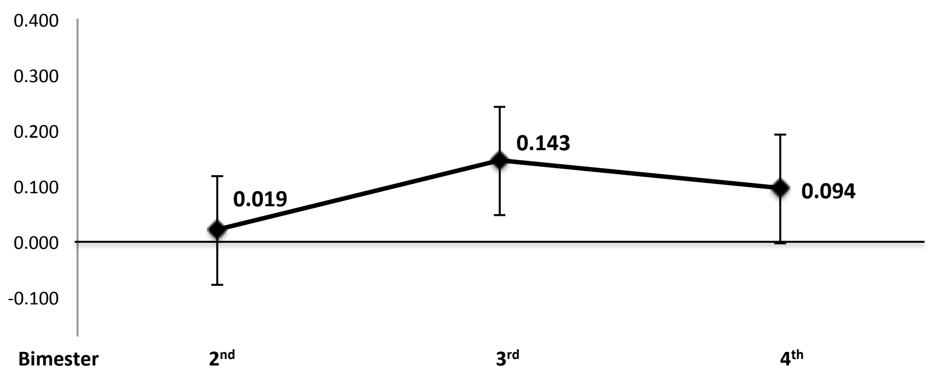
Note: Panels A and B show the raw data for attendance and GPA pre- and post-intervention, for treatment (engagement) and control groups of the theory-based nudging program. Attendance is recorded in percentage points (0-1 interval). The GPA has a 10 point scale, where 5 is the passing grade. The intervention started at the beginning of the third quarter and lasted until the end of the fourth quarter. Attendance and GPA are available for each of the forth quarter, as part of students' transcripts, allowing us to estimate a differences-in-differences model. Promotion rate and standardized test, however, are only available at the end of the school year (post-intervention).

Figure 5: Differences-in-differences coefficient of the theory-based nudging program by quarter

Panel A: Attendance (p.p.)



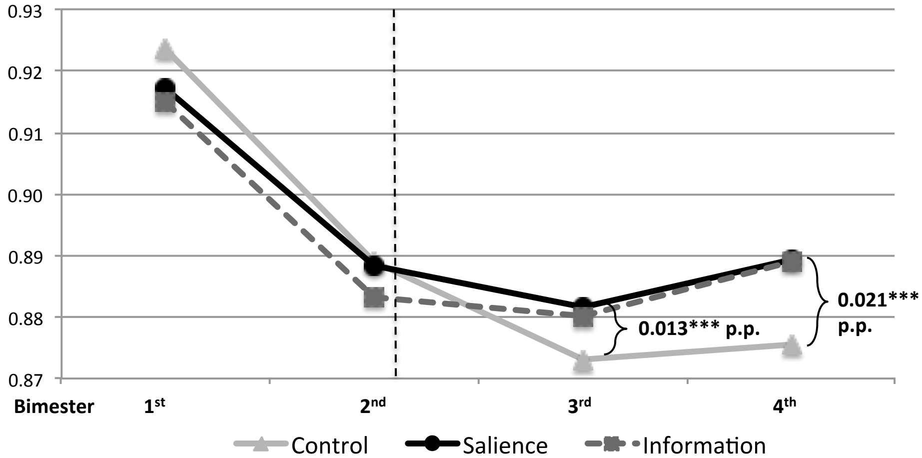
Panel A: GPA (std.)



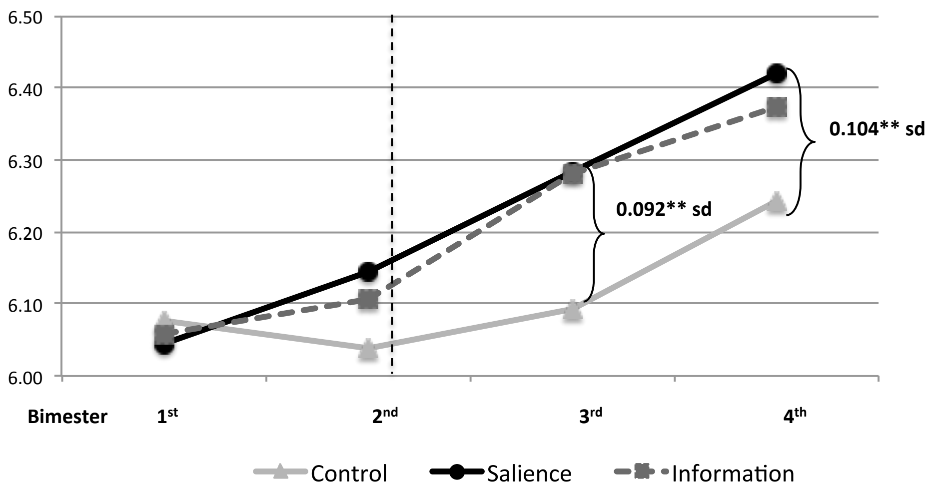
Note: Panels A and B show the differences-in-differences estimates from equation 3 for the theory-based nudging program by quarter, where the first quarter if the reference group. GPA was normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. 90% confidence interval with standard errors clustered at the classroom level are showed. A dummy variable for the control group within class was also included in the model, as well as it's interaction with a time dummy. Attendance and GPA are available for each of the forth quarter, as part of students' transcripts, allowing us to estimate a differences-in-differences model. Promotion rate and standardized test, however, are only available at the end of the school year (post-intervention).

Figure 6: Are effects short-lived? Effect of the intervention over time

Panel A: Effect on attendance over time



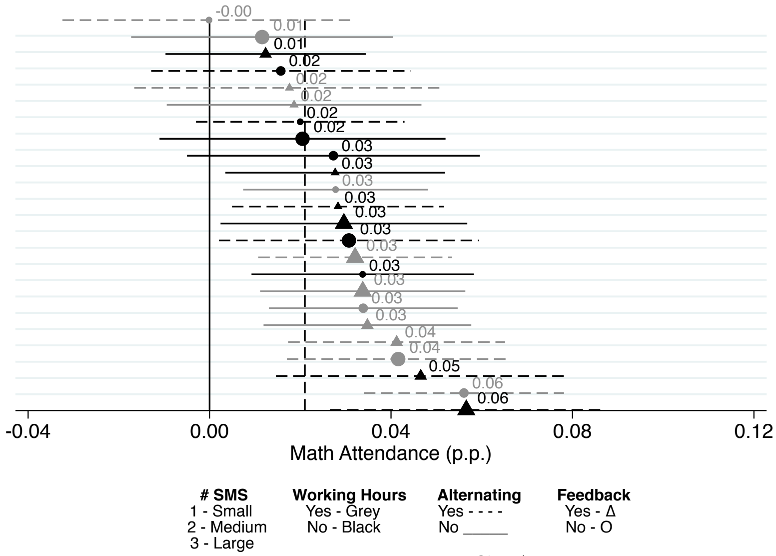
Panel B: Effect on GPA over time



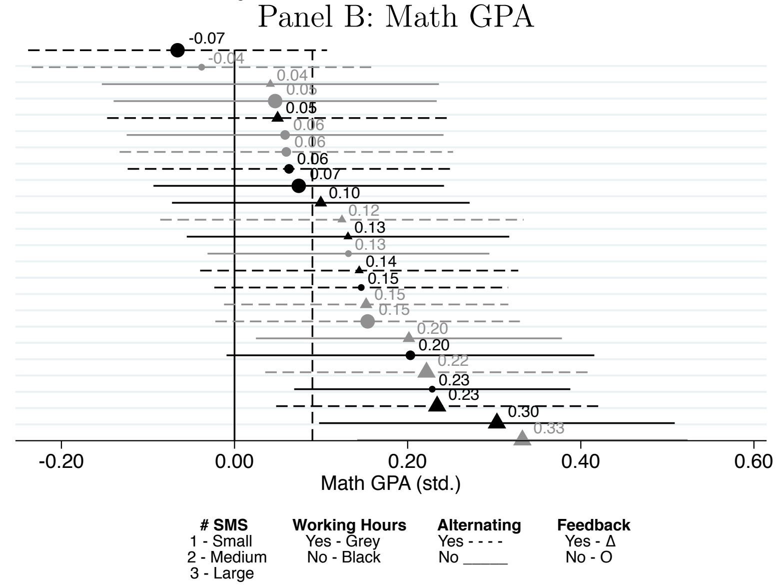
Note: Panels A and B show the raw data for attendance and GPA pre- and post-intervention, for treatment and control groups. Attendance is recorded in percentage points (0-1 interval). The GPA has a 10 point scale, where 5 is the passing grade. The intervention started at the beginning of the third quarter and lasted until the end of the fourth quarter. Attendance and GPA are available for each of the forth quarter, as part of students' transcripts, while promotion rate and standardized test are only available at the end of the school year. The coefficients on the graph show the difference between the salience and pure control group from a model estimated with student controls, strata fixed effect and standard errors clustered at the classroom level, as specified by equation 1. Coefficients for GPA are in standard deviation, where GPA was normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Figure 7: All combinations of features of a nude program targeted at capturing parents' attention

Panel A: Math attendance



Panel B: Math GPA



Note: The nudging program cross-randomizes different feature of the design of a typical SMS campaign. The program assess the impacts of alternative campaign parameters: (i) frequency (0, 1, 2 or 3 times a week), (ii) time of the day (afternoon or evening), (iii) consistency (constant or varying time of delivery), and (iv) interactivity (in the form of a feedback flow that asks whether parents complied with the suggested activity). The combination of each treatment generates 24 cells. Panels A and B show the differences-in-differences estimates of each combination cell on students' attendance and GPA (e.g. effect of receiving 3 SMS per week, during the afternoon, alternating time and with feedback). Each horizontal line of the graph represents one cell. 90% confidence interval with standard errors clustered at the classroom level are showed. The size of the markers indicates the number of SMS received (1, 2 or 3); the color of the marker and error bar indicates if the message was sent during work hours (grey) or evening (black); the error bar line style indicates if the time was alternated (dashed line) or not (continuing line); and the shape of the marker indicates if feedback was sent (yes for triangle and no for circle). Attendance is showed in percentage points (0-1 interval), and GPA is showed in standard deviation, where GPA was normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively.

Tables

Table 1: Descriptive statistics and balance

	Means				Diff=0 p-value	Sample Size
	Pure Control	Control Within Class	Salience	Information		
Panel A: Student characteristics						
Female	0.48	0.50	0.51	0.51	0.14	15589
Age	14.71	14.72	14.71	14.75	**0.03**	15595
Brown	0.34	0.35	0.34	0.35	0.48	15592
Black	0.06	0.05	0.06	0.06	0.45	15592
Portuguese GPA (max 10)	6.18	6.19	6.13	6.13	0.36	15437
Math GPA (max 10)	5.94	5.99	5.92	5.90	0.25	15453
Portuguese attendance	0.91	0.92	0.92	0.91	0.68	15480
Math attendance	0.91	0.91	0.91	0.91	0.30	15440
Panel B: Adult responsible for student						
Mother	0.78	0.76	0.76	0.76	0.28	15597
Age	40.43	40.25	40.34	40.42	0.86	15461
Brown	0.34	0.34	0.34	0.34	0.65	15593
Black	0.07	0.06	0.07	0.07	0.80	15593
Middle school incomplete	0.32	0.30	0.31	0.31	0.66	15591
Middle school complete	0.30	0.26	0.28	0.27	0.17	15591
High School	0.31	0.33	0.30	0.31	0.13	15591
Earns less than 1 MW (1MW $\sim$ \$250)	0.17	0.18	0.17	0.18	0.63	15593
Earns between 1 - 3 MW	0.42	0.45	0.45	0.46	0.41	15593

Note: Means net of randomization strata fixed effects. P-values calculated using randomization strata fixed effects and standard errors clustered at the classroom level. Data on students' gender, age, GPA and attendance was collected from administrative records, and data on students' race and on the adult responsible for student was collected from the baseline survey took by parents who opted-in to the program.

Table 2: Selection in opt-in

	Mean		Diff.	Sample Size
	No	Yes		
Female	0.45	0.50	0.05***	23372
			[0.01]	
Age	14.92	14.73	-0.19***	23398
			[0.01]	
Portuguese GPA (max 10)	5.39	6.16	0.77***	22687
			[0.03]	
Math GPA (max 10)	5.09	5.94	0.84***	22691
			[0.03]	
Portuguese attendance	0.88	0.91	0.04***	22850
			[0.00]	
Math attendance	0.87	0.91	0.04***	22753
			[0.00]	
Cash transfer beneficiary	0.19	0.16	-0.03***	23029
			[0.01]	

Note: Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Because parents who did not opt-in to the program didn't answer the baseline survey, we only have limited information on them, coming from administrative records (students' gender, age, GPA, attendance and if the family receives cash transfer). We run a simple regression, where each of the characteristics in the horizontal line served as dependent variable, and a dummy indicating if parents opted-in served as the independent variable.

Table 3: School transcripts and standardized tests

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.021***	0.090***	0.032***	0.095**
	[0.006]	[0.032]	[0.012]	[0.047]
Information	0.021***	0.071**	0.026**	0.107**
	[0.006]	[0.032]	[0.012]	[0.047]
Control Mean	0.875	0.000	0.938	-0.000
P-value diff. [Info] -[Salience]	0.896	0.221	0.219	0.596
% Salience	0.99	1.26	1.20	0.89
[IC 90%]	[0.8;1.2]	[0.8;1.7]	[0.9;1.5]	[0.6;1.2]
Sample Size	12577	12577	12577	12577
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 4: Salience vs. relative information

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.021***	0.090***	0.032***	0.095**
	[0.006]	[0.032]	[0.012]	[0.047]
Individual Info	0.021***	0.069**	0.029**	0.097**
	[0.006]	[0.032]	[0.012]	[0.047]
Relative Info	0.022***	0.078*	0.017	0.141**
	[0.007]	[0.041]	[0.014]	[0.058]
Control Mean	0.875	0.000	0.938	-0.000
P-value diff. [Rel. Info] -[Salience]	0.770	0.690	0.086	0.252
% Salience	0.94	1.16	1.90	0.68
[IC 90%]	[0.6;1.2]	[0.4;1.9]	[0.0;3.8]	[0.3;1.1]
Sample Size	12577	12577	12577	12577
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 5: Interactions with information?

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.017***	0.070**	0.027**	0.101**
	[0.006]	[0.033]	[0.012]	[0.048]
Information	0.021***	0.070**	0.026**	0.108**
	[0.006]	[0.032]	[0.012]	[0.047]
Salience Only	0.001	0.049*	0.004	0.015
	[0.004]	[0.029]	[0.009]	[0.042]
Sample Size	12577	12577	12577	12577
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 6: School transcripts and test score - no pure control

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.005**	0.030*	0.003	0.000
	[0.003]	[0.017]	[0.006]	[0.025]
Information	0.005*	0.023	-0.000	0.016
	[0.003]	[0.019]	[0.005]	[0.028]
Control Mean	0.887	0.000	0.966	-0.000
P-value diff. [Info] -[Salience]	0.898	0.713	0.561	0.581
% Salience	1.06	1.32	-51.69	0.02
[IC 90%]	[0.2;1.9]	[-0.4;3.1]	[-7470.2;7366.9]	[-2.5;2.5]
Sample Size	11217	11217	11217	11217
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. Controls in the treated schools are the reference group. The pure control group was excluded from this analysis. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 7: A parallel salience intervention

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Panel A				
Salience	0.033***	0.138***	0.045***	0.118*
	[0.007]	[0.040]	[0.011]	[0.060]
Sample Size	3180	3180	3180	3180
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Panel B				
Engagement	0.020**	0.096		
	[0.009]	[0.060]		
Sample Size	7338	7338		

Notes: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. In Panel A, treatment effect was estimated from equation 1 for the subsample D (50% salience + 50% control) and standard error are clustered at the classroom level. Panel B shows differences-in-differences estimates from equation 3 for the parallel salience intervention, where the first quarter is the reference group and the fourth quarter is the final period. Only the group of parents who received one text message per week were included in the analysis of Panel B, and standard error are clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 8: Heterogeneity by parents' baseline beliefs wrt their child's GPA - parents' endline accuracy

	Pessimistic Parents (30.7%)		Accurate parents (36.9%)		Optimistic parents (32.4%)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Accuracy Math Attendance	Accuracy Math GPA	Accuracy Math Attendance	Accuracy Math GPA	Accuracy Math Attendance	Accuracy Math GPA
Salience	-0.04	0.14*	-0.03	0.02	-0.05	0.08
	[0.07]	[0.08]	[0.07]	[0.07]	[0.06]	[0.06]
Information	-0.12*	0.10	-0.04	-0.00	-0.06	0.04
	[0.07]	[0.08]	[0.07]	[0.07]	[0.06]	[0.06]
Control Mean	0.30	0.25	0.29	0.33	0.21	0.21
P-value diff. [Info] -[Salience]	0.13	0.50	0.72	0.61	1.00	0.53
% Salience	0.31	1.42	0.58	-6.65	1.00	1.80
[IC 90%]	[-0.5;1.1]	[0.1;2.7]	[-1.3;2.5]	[-252.3;239.0]	[-0.3;2.3]	[-1.6;5.2]
Sample Size	480	480	576	576	506	506
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Parents were asked at baseline to give their best estimate of their child performance in math classes. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on performance (below average; adequate; good; very good). Administrative data register data on attendance and GPA on a quarterly basis (period of $\sim$ 9 weeks). The GPA has a 10 point scale, where 5 is the passing grade. Parents' answers below average was determined as a GPA below 5, adequate as 5-6; good as 7-8 and very good as 9-10. Parents were also asked at endline to give their best estimate of how many times their child missed school and what was their final math GPA in the past quarter. Five categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; 6-8; more than 8) and parents answers for GPA were absolute values from 1-10. Data was then crossed with administrative records and a dummy variable were created, where parents who estimated right received value 1 and those who estimated wrong received value 0.

Table 9: Heterogeneity by students' baseline attendance

	≤ Median (54.6%)				>Median (45.4%)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)
Salience	0.13***	0.18***	0.05	0.10	0.13**	0.08	0.14*	0.08
	[0.05]	[0.06]	[0.08]	[0.08]	[0.06]	[0.05]	[0.08]	[0.07]
Information	0.12**	0.13**	0.09	0.19**	0.14**	0.10*	0.11	0.06
	[0.05]	[0.06]	[0.08]	[0.08]	[0.06]	[0.05]	[0.08]	[0.07]
Control Mean	-0.23	0.01	-0.12	-0.00	0.25	-0.01	0.13	0.00
P-value diff. [Info] -[Salience]	0.61	0.17	0.41	0.07	0.87	0.64	0.56	0.71
% Salience	1.15	1.38	0.53	0.53	0.96	0.83	1.26	1.31
[IC 90%]	[0.6;1.7]	[0.8;2.0]	[-0.5;1.5]	[0.1;1.0]	[0.5;1.4]	[0.3;1.4]	[0.4;2.2]	[-0.3;3.0]
Sample Size	2300	2073	2300	2073	1872	2295	1872	2295
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: A GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Students with baseline attendance below or equal to the class median were determined as low-performing, and students with baseline attendance above the median were determined as high-performing for the purposes of this analysis.

Table 10: Heterogeneity by parents' attention: time to answer the first question of baseline survey

	Low Attention (>Median 52.5%)				High Attention (≤ Median 47.5%)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)
Salience	0.04***	0.18***	0.05**	0.10	0.02**	0.08	0.04*	0.08
	[0.01]	[0.06]	[0.02]	[0.08]	[0.01]	[0.05]	[0.02]	[0.07]
Information	0.03***	0.13**	0.04*	0.19**	0.03***	0.10*	0.05***	0.06
	[0.01]	[0.06]	[0.02]	[0.08]	[0.01]	[0.05]	[0.02]	[0.07]
Control Mean	0.86	0.01	0.93	-0.00	0.86	-0.01	0.93	0.00
P-value diff. [Info] -[Salience]	0.26	0.17	0.27	0.07	0.55	0.64	0.26	0.71
% Salience	1.18	1.38	1.30	0.53	0.89	0.83	0.78	1.31
[IC 90%]	[0.9;1.5]	[0.8;2.0]	[0.7;1.9]	[0.1;1.0]	[0.6;1.2]	[0.3;1.4]	[0.4;1.1]	[-0.3;3.0]
Sample Size	2073	2073	2073	2073	2295	2295	2295	2295
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: A GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Response times are used by cognitive psychologists as a measure of cognitive control, see Mani et al. (2013). Parents with above-median average response times are treated as inattentive (low attention), while parents with below-median average response (or equal) are treated as attentive (high attention). Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$.

Table 11: Heterogeneity by parents' willingness to receive information (WTR)

	School Transcripts and Test Scores				Parents' Beliefs	
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Accuracy Math Attendance (p.p.)	(6) Accuracy Math GPA (p.p.)
	Low willingness to receive information (WTR) (63.3%)					
Salience	0.03***	0.12**	0.03*	0.08	0.02	0.10**
	[0.01]	[0.05]	[0.02]	[0.07]	[0.04]	[0.04]
Information	0.03***	0.09*	0.04**	0.16**	-0.03	0.02
	[0.01]	[0.05]	[0.02]	[0.07]	[0.04]	[0.04]
Control Mean	0.86	-0.06	0.93	-0.05	0.21	0.23
P-value diff. [Info] -[Salience]	0.57	0.42	0.56	0.10	0.13	0.04
% Salience	1.10	1.31	0.85	0.55	-0.83	4.14
[IC 90%]	[0.8;1.4]	[0.5;2.1]	[0.4;1.3]	[0.1;1.0]	[-5.1;3.4]	[-6.2;14.5]
Sample Size	2578	2578	2578	2578	1071	1071
	High willingness to receive information (WTR) (36.7%)					
Salience	0.04***	0.18***	0.07***	0.14	-0.15**	0.02
	[0.01]	[0.07]	[0.02]	[0.10]	[0.07]	[0.08]
Information	0.04***	0.15**	0.07***	0.07	-0.16**	0.04
	[0.01]	[0.07]	[0.02]	[0.10]	[0.07]	[0.08]
Control Mean	0.86	0.04	0.91	0.07	0.36	0.33
P-value diff. [Info] -[Salience]	0.89	0.46	0.70	0.24	0.67	0.75
% Salience	1.02	1.23	1.06	1.96	0.89	0.59
[IC 90%]	[0.8;1.3]	[0.6;1.8]	[0.8;1.3]	[-0.9;4.8]	[0.5;1.3]	[-1.5;2.7]
Sample Size	1317	1317	1317	1317	620	620
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Parents were asked at baseline about their interest in receiving information about their child's attendance and they had three options: i. no interest, ii. some interest, iii. a lot of interest. Parents who answered i. or ii. were defined as having a low WTR and parents who answered iii. were defined as having a high WTR. Parents were asked at endline to give their best estimate of how many times their child missed school and what was their child final math GPA in the past quarter. Data was then crossed with administrative records and a dummy variable was created, where parents who estimated right received value 1 and those who estimated wrong received value 0.

Table 12: Heterogeneity by parents' baseline beliefs wrt their child's GPA - transcripts and test scores

	Pessimistic Parents (29.4%)				Accurate parents (36.2%)				Optimistic parents (34.5%)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)	(9) Math Attendance (p.p.)	(10) Math GPA (std.)	(11) Promotion Rate (p.p.)	(12) Math Standardized Test (std.)
Salience	0.05***	0.09	0.02*	0.14	0.01	0.08	0.03	0.06	0.03**	0.14**	0.05*	0.03
	[0.01]	[0.07]	[0.01]	[0.10]	[0.01]	[0.07]	[0.02]	[0.10]	[0.01]	[0.06]	[0.03]	[0.11]
Information	0.05***	0.09	0.02	0.14	0.01	0.10	0.04*	0.14	0.02*	0.10	0.05*	0.01
	[0.01]	[0.07]	[0.01]	[0.10]	[0.01]	[0.07]	[0.02]	[0.10]	[0.01]	[0.07]	[0.03]	[0.11]
Control Mean	0.85	0.50	0.97	0.21	0.88	0.01	0.93	0.01	0.84	-0.43	0.91	-0.18
P-value diff. [Info] -[Salience]	0.62	0.91	0.08	0.93	0.99	0.73	0.44	0.19	0.42	0.37	0.87	0.73
% Salience	1.07	1.08	1.60	0.96	1.00	0.82	0.73	0.40	1.26	1.43	1.06	5.05
[IC 90%]	[0.8;1.3]	[-0.1;2.3]	[0.4;2.8]	[0.2;1.7]	[0.2;1.8]	[0.0;1.6]	[0.2;1.3]	[-0.4;1.2]	[0.6;1.9]	[0.4;2.5]	[0.5;1.7]	[-132.5;142.6]
Sample Size	1026	1026	1026	1026	1263	1263	1263	1263	1203	1203	1203	1203
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: A GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Parents were asked at baseline to give their best estimate of their child performance in math classes. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on performance (below average; adequate; good; very good). Administrative data register data on attendance and GPA on a quarterly basis (period of ∼ 9 weeks). The GPA has a 10 point scale, where 5 is the passing grade. Parents' answers below average was determined as a GPA below 5, adequate as 5-6; good as 7-8 and very good as 9-10.

Table 13: Heterogeneity by parents' baseline beliefs wrt their child's GPA - parents' behavior

	Pessimistic Parents (29.4%)			Accurate parents (36.2%)			Optimistic parents (34.5%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Academic Activities	Incentives	Talk	Academic	Incentives Academic	Talk	Academic Academic	Incentives	Talk
Salience	0.22*	0.06	0.10	-0.01	0.07	-0.04	-0.03	0.07	0.21*
	[0.13]	[0.13]	[0.12]	[0.11]	[0.10]	[0.10]	[0.11]	[0.12]	[0.11]
Information	0.20	0.03	0.05	0.19*	0.10	0.10	-0.10	0.14	0.15
	[0.12]	[0.13]	[0.12]	[0.11]	[0.10]	[0.10]	[0.12]	[0.12]	[0.12]
Control Mean	-0.17	-0.05	0.00	0.01	-0.02	0.05	0.12	-0.04	-0.01
P-value diff. [Info] -[Salience]	0.84	0.74	0.54	0.01	0.68	0.07	0.39	0.45	0.40
% Salience	1.08	2.00	2.02	-0.04	0.69	-0.40	0.26	0.53	1.44
[IC 90%]	[0.4;1.8]	[-8.3;12.3]	[-3.6;7.6]	[-1.0;1.0]	[-0.4;1.8]	[-2.6;1.8]	[-1.2;1.7]	[-0.5;1.5]	[0.2;2.7]
Sample Size	865	863	860	1050	1048	1045	995	994	988
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: outcome variables were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Parents were asked at baseline to give their best estimate of their child performance in math classes. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on performance (below average; adequate; good; very good). Administrative data register data on attendance and GPA on a quarterly basis (period of $\sim$ 9 weeks). The GPA has a 10 point scale, where 5 is the passing grade. Parents' answers below average was determined as a GPA below 5, adequate as 5-6; good as 7-8 and very good as 9-10. Students were asked at the endline survey about their parent's behavior, where they had to state how often their parents engage in certain activities (never, almost never, sometimes, almost always). Out of the 12 questions, factor analysis was used to create 3 variables of parents behavior: academic activities (help with homework, help to organize school material, participate in school-parent meetings, talk to the teachers); incentives (incentivize to not miss school, to not be late, to study and to read); talk (ask about homework, ask about grades, ask about day in school and classes).

Table 14: Heterogeneous effects by features of SMS communication - attendance

	(1) Math Attendance (p.p.)	(2) Portuguese Attendance (p.p.)
Frequency		
1 SMS per week	0.020**	0.014*
	[0.008]	[0.008]
2 SMS per week	0.034***	0.028***
	[0.008]	[0.008]
3 SMS per week	0.032***	0.029***
	[0.008]	[0.008]
P-value diff.	0.08	0.02
SMS Delivery Time		
Work hours	0.030***	0.022***
	[0.008]	[0.008]
Off-work hours	0.028***	0.025***
	[0.008]	[0.008]
P-value diff.	0.72	0.54
Consistency of delivery time		
Varying	0.031***	0.028***
	[0.008]	[0.008]
Constant	0.027***	0.020**
	[0.008]	[0.008]
P-value diff.	0.42	0.09
Interactivity		
Interactivity	0.027***	0.019**
	[0.008]	[0.008]
Passive	0.031***	0.029***
	[0.008]	[0.008]
P-value diff.	0.37	0.05
Sample Size	10308	10308
Randomization strata FE	No	No

Note: The nudging program cross-randomizes different feature of the design of a typical SMS campaign. The program assess the impacts of alternative campaign parameters: (i) frequency (0, 1, 2 or 3 times a week), (ii) time of the day (afternoon or evening), (iii) consistency (constant or varying time of delivery), and (iv) interactivity (in the form of a feedback flow that asks whether parents complied with the suggested activity). The table shows the differences-in-differences estimates from equation 3 for the theory-based nudging program, where the first quarter if the reference group, and the fourth quarter is the end period. Standard errors are clustered at the classroom level. A dummy variable for the control group within class was also included in the model, as well as it's interaction with a time dummy. Attendance and GPA are available for each of the forth quarter, as part of students' transcripts, allowing us to estimate a differences-in-differences model. Promotion rate and standardized test, however, are only available at the end of the school year (post-intervention). Attendance is measured in percentage points (0-1 interval). Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table 15: Heterogeneous effects by features of SMS communication - GPA

	(1) Math GPA (std.)	(2) Portuguese GPA (std.)
Frequency		
1 SMS per week	0.095	0.042
	[0.060]	[0.067]
2 SMS per week	0.118**	0.067
	[0.060]	[0.068]
3 SMS per week	0.176***	0.147**
	[0.062]	[0.068]
P-value diff.	0.08	0.02
SMS Delivery Time		
Work hours	0.150**	0.074
	[0.058]	[0.067]
Off-work hours	0.110*	0.097
	[0.056]	[0.064]
P-value diff.	0.34	0.50
Consistency of delivery time		
Varying	0.139**	0.074
	[0.056]	[0.065]
Constant	0.121**	0.096
	[0.059]	[0.068]
P-value diff.	0.68	0.59
Interactivity		
Interactivity	0.090	0.047
	[0.056]	[0.065]
Passive	0.170***	0.123*
	[0.057]	[0.067]
P-value diff.	0.04	0.04
Sample Size	10308	10308
Randomization strata FE	No	No

Note: The nudging program cross-randomizes different feature of the design of a typical SMS campaign. The program assess the impacts of alternative campaign parameters: (i) frequency (0, 1, 2 or 3 times a week), (ii) time of the day (afternoon or evening), (iii) consistency (constant or varying time of delivery), and (iv) interactivity (in the form of a feedback flow that asks whether parents complied with the suggested activity). The table shows the differences-in-differences estimates from equation 3 for the theory-based nudging program, where the first quarter if the reference group, and the fourth quarter is the end period. Standard errors are clustered at the classroom level. A dummy variable for the control group within class was also included in the model, as well as it's interaction with a time dummy. Attendance and GPA are available for each of the forth quarter, as part of students' transcripts, allowing us to estimate a differences-in-differences model. Promotion rate and standardized test, however, are only available at the end of the school year (post-intervention). GPA is showed in standard deviation, where GPA was normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

A Appendix – SMS Text Messages

[FORTHCOMING]

B Appendix – Survey Instruments

[FORTHCOMING]

C Appendix – Balance and attrition tests

In this section, we present balance and attrition tests. Table C.1 shows descriptive statistics and balance test for the main sample used in the analysis (e.g. Tables 3, 4, 5). Table C.2 presents descriptive statistics and balance test for the theory of change sample. Next, Tables C.3 and C.4 contain a selective attrition analysis for completing the surveys by treatment status and by baseline characteristics, respectively. Because parents who opted into the program had different characteristics from those who did not opt in (as we showed in Table 2), in Table C.5 we show results for school transcripts and test scores re-weighting observations by the inverse probability of opting into the program. Finally, Table C.6 describes statistics and balance for the theory-based nudging program for the parents receiving one message per week, which is the sample sample used to run the differences-in-differences analysis described in section 4.

Table C.1: Descriptive statistics and balance - school transcripts and test score sample

	Means				Diff=0 p-value	Sample Size
	Pure Control	Control Within Class	Salience	Info		
Student characteristics						
Female	0.47	0.50	0.51	0.52	0.03	12577
Age	14.69	14.67	14.67	14.71	0.03	12577
Brown	0.36	0.35	0.34	0.35	0.14	12577
Black	0.06	0.05	0.06	0.06	0.79	12577
Portuguese GPA (max 10)	6.39	6.31	6.27	6.28	0.69	12577
Math GPA (max 10)	6.10	6.11	6.05	6.06	0.57	12577
Portuguese attendance	0.92	0.92	0.92	0.92	0.50	12577
Math attendance	0.92	0.92	0.92	0.91	0.39	12577
Adult responsible for student						
Mother	0.77	0.75	0.76	0.76	0.45	12577
Age	40.39	40.28	40.34	40.57	0.68	12577
Brown	0.36	0.34	0.34	0.34	0.15	12577
Black	0.07	0.06	0.07	0.07	0.71	12577
Middle school incomplete	0.32	0.30	0.30	0.31	0.32	12577
Middle school complete	0.28	0.26	0.28	0.28	0.48	12577
High School	0.34	0.33	0.32	0.31	0.19	12577
Earns less than 1 MW (1MW $\sim$ \$250)	0.17	0.17	0.17	0.17	0.80	12577
Earns between 1 - 3 MW	0.44	0.46	0.46	0.47	0.80	12577

Note: Means net of randomization strata fixed effects. P-values calculated using randomization strata fixed effects and standard errors clustered at the classroom level. Data on students' gender, age, GPA and attendance was collected from administrative records, and data on students' race and on the adult responsible for student was collected from the baseline survey took by parents who opted-in to the program.

Table C.2: Descriptive statistics and balance - theory of change sample

	Means				Diff=0 p-value	Sample Size
	Pure Control	Control Within Class	Salience	Info		
Student characteristics						
Female	0.50	0.50	0.52	0.52	0.18	9539
Age	14.65	14.65	14.66	14.68	0.24	9539
Brown	0.36	0.35	0.33	0.34	0.33	9539
Black	0.05	0.05	0.05	0.05	0.68	9539
Portuguese GPA (max 10)	6.51	6.45	6.39	6.39	0.51	9539
Math GPA (max 10)	6.21	6.22	6.20	6.17	0.87	9539
Portuguese attendance	0.93	0.93	0.93	0.93	0.30	9539
Math attendance	0.93	0.92	0.92	0.92	0.45	9539
Adult responsible for student						
Mother	0.78	0.75	0.76	0.76	0.43	9539
Age	40.62	40.39	40.34	40.74	0.64	9539
Brown	0.35	0.34	0.34	0.33	0.27	9539
Black	0.07	0.06	0.07	0.07	0.67	9539
Middle school incomplete	0.31	0.29	0.29	0.28	0.44	9539
Middle school complete	0.28	0.26	0.27	0.28	0.37	9539
High School	0.33	0.34	0.32	0.33	0.42	9539
Earns less than 1 MW (1MW $\sim$ \$250)	0.16	0.16	0.16	0.16	0.86	9539
Earns between 1 - 3 MW	0.44	0.47	0.46	0.47	0.92	9539

Note: Means net of randomization strata fixed effects. P-values calculated using randomization strata fixed effects and standard errors clustered at the classroom level. Data on students' gender, age, GPA and attendance was collected from administrative records, and data on students' race and on the adult responsible for student was collected from the baseline survey took by parents who opted-in to the program.

Table C.3: Selective attrition - survey completion

	(1) Baseline Survey - Parents	(2) Endline Survey - Parents	(3) Endline Survey - Students
Salience	-0.016	0.022	0.016
	[0.020]	[0.024]	[0.016]
Information	-0.008	0.039	0.013
	[0.021]	[0.024]	[0.016]
Control Within Class	-0.006	0.045*	0.020
	[0.020]	[0.023]	[0.016]
P-value Salience=Info=Control Within	0.828	0.412	0.694
Sample Size	4862	4653	15597
Randomization strata FE	Yes	Yes	Yes

Note: pure control is the omitted group. Parental survey was considered completed if at least 11 questions were answered, and student survey was considered completed if at least 75% of the questions were answered. We run a simple regression where a dummy indicating if parents completed the survey served as the outcome variable and treatment status served as independent variables. Randomization stratum fixed effects were also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table C.4: Marginal probability of completing the survey

	(1) Baseline Survey - Parents	(2) Endline Survey - Parents	(3) Endline Survey - Students
Student characteristics			
Female	0.006	-0.010	0.015
	[0.012]	[0.013]	[0.007]
Age	-0.017*	-0.027*	-0.055*
	[0.009]	[0.009]	[0.006]
Brown or Black	-0.041***	-0.012***	-0.025***
	[0.012]	[0.013]	[0.007]
Math GPA (max 10)	0.012***	0.016***	0.027***
	[0.003]	[0.003]	[0.002]
Math attendance	0.147**	0.213**	0.774**
	[0.067]	[0.070]	[0.045]
Adult responsible for student			
Mother	0.007	0.057	-0.006
	[0.015]	[0.017]	[0.008]
Age	-0.003***	-0.002***	0.001***
	[0.001]	[0.001]	[0.000]
Brown or Black	-0.052***	-0.010***	-0.012***
	[0.013]	[0.013]	[0.007]
Low Education (middle school incomplete)	-0.070***	-0.059***	-0.042***
	[0.014]	[0.015]	[0.008]
Cash transfer beneficiary	-0.032**	-0.039**	-0.029**
	[0.016]	[0.018]	[0.010]

Note: Parental survey was considered completed if at least 11 questions were answered, and student survey was considered completed if at least 75% of the questions were answered. We run a simple regression, where each of the characteristics in the horizontal line served as independent variable, and a dummy indicating if parents completed the survey served as dependent variable. A different regression was estimated for each characteristic. Randomization stratum fixed effects were also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$.

Table C.5: School transcripts and standardized tests - weighting by the probability of opting-in the program

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.022***	0.100***	0.038***	0.096**
	[0.006]	[0.032]	[0.013]	[0.046]
Information	0.022***	0.077**	0.031**	0.105**
	[0.007]	[0.032]	[0.013]	[0.046]
Control Mean	0.875	0.000	0.938	-0.000
P-value diff. [Info] -[Salience]	0.854	0.141	0.162	0.680
% Salience	0.98	1.31	1.24	0.91
[IC 90%]	[0.8;1.2]	[0.9;1.8]	[0.9;1.6]	[0.6;1.3]
Sample Size	12550	12550	12550	12550
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Inverse probability weighting was used to weight estimates by the probability of opting-in the program based on observables.

Table C.6: A parallel salience intervention: balance

	Means			Diff=0 p-value	Sample Size
	Pure Control	Control Within Class	Engagement		
Panel A: Student characteristics					
Female	0.47	0.51	0.50	0.23	3058
Age	14.68	14.66	14.69	0.68	3058
Brown	0.36	0.32	0.31	**0.05**	3058
Black	0.06	0.05	0.05	0.53	3058
Portuguese GPA (max 10)	6.37	5.99	5.99	**0.00**	3019
Math GPA (max 10)	6.07	5.79	5.75	**0.00**	3021
Portuguese attendance	0.93	0.92	0.93	0.91	3037
Math attendance	0.92	0.92	0.92	0.88	2975
Panel B: Adult responsible for student					
Mother	0.78	0.76	0.74	0.14	3058
Age	40.38	40.77	40.47	0.51	3008
Black	0.07	0.07	0.07	0.88	3058
Middle school incomplete	0.31	0.28	0.26	**0.06**	3058
Middle school complete	0.30	0.25	0.25	**0.03**	3058
High School	0.33	0.32	0.33	0.85	3058
Earns less than 1 MW (1MW $\sim$ \$250)	0.16	0.15	0.13	**0.10**	3058
Earns between 1 - 3 MW	0.43	0.46	0.47	0.11	3058

Note: P-values computed from robust standard. Engagement treatment includes only parents who received one text message per week. Data on students' gender, age, GPA and attendance was collected from administrative records, and data on students' race and on the adult responsible for student was collected from the baseline survey took by parents who opted-in to the program.

D Appendix – Theory of change

This section presents tables for the theory of change analysis, as well as the heterogeneous effects for boys and girls, both described in section 4. We also explain in more details the variables used in the analysis. The theory of change analysis uses data from students endline survey, where students answered questions about their parent's behavior and aspirations, as well as their own behavior. A common sample of 9539 students was used to investigate results on parent's behavior and aspirations, student's behavior, and school transcripts and test score.

At the endline survey, students were asked to state how often their parents engage in certain activities (never, almost never, sometimes, almost always). Out of the 12 questions, factor analysis was performed to create 3 variables of parental behavior: *academic activities* (help with homework, help to organize school material, participate in school-parent meetings, talk to the teachers); *incentives* (incentivize to not miss school, to not be late, to study and to read); *talk* (ask about homework, ask about grades, ask about day in school and classes). Students were also asked if their parents believed they would go to college and a dummy variable for *parent's aspirations* was created, which assumes value one if parents do believe the student will go to college and zero otherwise.

Finally, students were requested to answer how many hours per day (0, 15 minutes, 30 minutes, 1 hours, 2 hours, more than 2 hours) they spend in each of the following activities: i. studying at home on weekdays; ii. studying at home on weekends; iii. studying at home the day before a test; iv. reading a book; v. reading the newspaper; vi. reading magazines; vii. watching TV; viii. navigating on the internet or social media; and ix. helping with housework. We used factor analysis to create three variables of student's behavior: *academic activities* (items i, ii and iii); *reading activities* (items iv., v and vi.) and *other activities* (items vii, viii and ix).

All the variables were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. Results were estimated according to equation 1.

Table D.1 shows results for school transcripts and test score; Tables D.2 and D.3 present results for parent's behavior and aspirations, respectively; and Table D.4 describes results for student's behavior. Next, Tables D.5, D.6, D.7, and D.8 show heterogeneous results for boys and gilrs, following the same order: school transcripts and test score, parent's behavior and aspirations, and student's behavior.

Table D.1: School transcripts and test score

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.016***	0.072**	0.030**	0.075
	[0.006]	[0.034]	[0.012]	[0.053]
Information	0.017***	0.058*	0.026**	0.091*
	[0.006]	[0.034]	[0.012]	[0.053]
Control Mean	0.889	0.000	0.945	0.000
P-value diff. [Info] -[Salience]	0.634	0.420	0.477	0.510
% Salience	0.94	1.24	1.12	0.82
[IC 90%]	[0.7;1.1]	[0.6;1.8]	[0.8;1.4]	[0.4;1.3]
Sample Size	9539	9539	9539	9539
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$.

Table D.2: Parents' behavior

	(1) Academic activities	(2) Incentives	(3) Talk
Salience	0.064	0.096**	0.122***
	[0.050]	[0.041]	[0.043]
Information	0.092*	0.075*	0.147***
	[0.051]	[0.042]	[0.044]
Control Mean	-0.000	0.000	-0.000
P-value diff. [Info] -[Salience]	0.263	0.382	0.374
% Salience	0.69	1.29	0.83
[IC 90%]	[0.2;1.2]	[0.6;2.0]	[0.5;1.1]
Sample Size	9539	9539	9539
Randomization strata FE	Yes	Yes	Yes
Student controls	Yes	Yes	Yes

Note: Variables were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. At the endline survey, students were asked to state how often their parents engage in certain activities (never, almost never, sometimes, almost always). Out of the 12 questions, factor analysis was performed to create 3 variables of parental behavior: *academic activities* (help with homework, help to organize school material, participate in school-parent meetings, talk to the teachers); *incentives* (incentivize to not miss school, to not be late, to study and to read); *talk* (ask about homework, ask about grades, ask about day in school and classes).

Table D.3: Parents' aspirations

	(1) Parents' Aspirations College
Salience	0.095***
	[0.036]
Information	0.092**
	[0.036]
Control Mean	-0.000
P-value diff. [Info] -[Salience]	0.891
% Salience	1.04
[IC 90%]	[0.6;1.5]
Sample Size	9539
Randomization strata FE	Yes
Student controls	Yes

Note: The dependent variable was normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. At the endline survey, students were asked if their parents believed they would go to college and a dummy variable for *parent's aspirations* was created, which assumes value one if parents do believe the student will go to college and zero otherwise.

Table D.4: Students' behavior

	(1) Academic activities	(2) Reading activities	(3) Other activities
Salience	0.123**	0.113*	-0.110**
	[0.050]	[0.060]	[0.052]
Information	0.151***	0.116*	-0.108**
	[0.051]	[0.065]	[0.054]
Control Mean	0.000	-0.000	0.000
P-value diff. [Info] -[Salience]	0.344	0.946	0.933
% Salience	0.81	0.98	1.02
[IC 90%]	[0.5;1.1]	[0.5;1.4]	[0.6;1.4]
Sample Size	9539	9539	9539
Randomization strata FE	Yes	Yes	Yes
Student controls	Yes	Yes	Yes

Note: Variables were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. At the endline survey, students were requested to answer how many hours per day (0, 15 minutes, 30 minutes, 1 hours, 2 hours, more than 2 hours) they spend in each of the following activities: i. studying at home on weekdays; ii. studying at home on weekends; iii. studying at home the day before a test; iv. reading a book; v. reading the newspaper; vi. reading magazines; vii. watching TV; viii. navigating on the internet or social media; and ix. helping with housework. Factor analysis was performed to create three variables of student's behavior: *academic activities* (items i, ii and iii); *reading activities* (items iv., v and vi.) and *other activities* (items vii, viii and ix).

Table D.5: School transcripts and test score - boys and girls

	Boys				Girls				Diff. (Girls)-(Boys)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)	Math Attendance (p.p.)	Math GPA (std.)	Promotion Rate (p.p.)	Math Standardized Test (std.)
Salience	0.02***	0.13***	0.04**	0.10*	0.01*	0.02	0.01	0.04	-0.01	-0.12**	-0.03*	-0.06
	[0.01]	[0.04]	[0.02]	[0.06]	[0.01]	[0.04]	[0.01]	[0.06]	[0.01]	[0.05]	[0.02]	[0.06]
Information	0.02***	0.12***	0.04**	0.13**	0.01**	0.00	0.01	0.05	-0.01	-0.12**	-0.03*	-0.07
	[0.01]	[0.04]	[0.02]	[0.06]	[0.01]	[0.04]	[0.01]	[0.06]	[0.01]	[0.05]	[0.02]	[0.07]
Control Mean	0.88	-0.22	0.92	-0.02	0.89	0.23	0.97	0.02				
P-value diff. [Info] -[Salience]	0.68	0.65	0.86	0.55	0.32	0.55	0.47	0.71				
% Salience	1.06	1.10	1.03	0.82	0.79	6.28	1.30	0.79				
[IC 90%]	[0.8;1.3]	[0.7;1.5]	[0.7;1.4]	[0.4;1.3]	[0.5;1.1]	[-135.5;148.1]	[0.4;2.2]	[-0.2;1.7]				
Sample Size	4654	4654	4654	4654	4885	4885	4885	4885				
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes				
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes				

Note: GPA and standardized test were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

Table D.6: Parents' behavior - boys and girls

	Boys			Girls			Diff. (Girls)-(Boys)		
	(1) Academic activities	(2) Incentives	(3) Talk	(4) Academic activities	(5) Incentives	(6) Talk	Academic activities	Incentives	Talk
Salience	0.13**	0.07	0.14***	0.00	0.11*	0.11*	-0.12*	0.04	-0.03
	[0.06]	[0.06]	[0.05]	[0.06]	[0.06]	[0.06]	[0.07]	[0.08]	[0.07]
Information	0.13**	0.05	0.17***	0.05	0.09	0.12**	-0.08	0.03	-0.04
	[0.06]	[0.06]	[0.05]	[0.07]	[0.06]	[0.06]	[0.08]	[0.08]	[0.07]
Control Mean	-0.02	-0.02	0.00	0.02	0.02	-0.00			
P-value diff. [Info] -[Salience]	0.86	0.66	0.43	0.21	0.48	0.63			
% Salience	0.95	1.32	0.82	0.06	1.26	0.86			
[IC 90%]	[0.5;1.4]	[-0.2;2.8]	[0.5;1.2]	[-2.0;2.1]	[0.5;2.0]	[0.4;1.3]			
Sample Size	4654	4654	4654	4885	4885	4885			
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes			
Student controls	Yes	Yes	Yes	Yes	Yes	Yes			

Note: Variables were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. At the endline survey, students were asked to state how often their parents engage in certain activities (never, almost never, sometimes, almost always). Out of the 12 questions, factor analysis was performed to create 3 variables of parental behavior: *academic activities* (help with homework, help to organize school material, participate in school-parent meetings, talk to the teachers); *incentives* (incentivize to not miss school, to not be late, to study and to read); *talk* (ask about homework, ask about grades, ask about day in school and classes).

Table D.7: Parents' aspirations - boys and girls

	Boys	Girls	Diff. (Girls)-(Boys)
	(1) Parents' Aspirations College	(2) Parents' Aspirations College	Parents' Aspirations College
Salience	0.12**	0.08	-0.04
	[0.06]	[0.05]	[0.08]
Information	0.10*	0.09*	-0.02
	[0.06]	[0.05]	[0.08]
Control Mean	-0.09	0.09	
P-value diff. [Info] -[Salience]	0.76	0.79	
% Salience	1.12	0.91	
[IC 90%]	[0.4;1.8]	[0.4;1.4]	
Sample Size	4654	4885	
Randomization strata FE	Yes	Yes	
Student controls	Yes	Yes	

Note:The dependent variable was normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. At the endline survey, students were asked if their parents believed they would go to college and a dummy variable for *parent's aspirations* was created, which assumes value one if parents do believe the student will go to college and zero otherwise.

Table D.8: Students' behavior - boys and girls

	Boys			Girls			Diff. (Girls)-(Boys)		
	(1) Academic activities	(2) Reading activities	(3) Other activities	(4) Academic activities	(5) Reading activities	(6) Other activities	Academic activities	Reading activities	Other activities
Salience	0.19***	0.17**	-0.09	0.06	0.06	-0.13**	-0.13*	-0.11	-0.04
	[0.06]	[0.07]	[0.06]	[0.07]	[0.07]	[0.07]	[0.07]	[0.08]	[0.08]
Information	0.18***	0.15**	-0.13*	0.12*	0.08	-0.09	-0.06	-0.07	0.04
	[0.05]	[0.07]	[0.07]	[0.07]	[0.08]	[0.07]	[0.07]	[0.08]	[0.08]
Control Mean	-0.14	-0.07	-0.18	0.14	0.08	0.18			
P-value diff. [Info] -[Salience]	0.81	0.73	0.38	0.13	0.65	0.26			
% Salience	1.06	1.11	0.74	0.51	0.77	1.44			
[IC 90%]	[0.7;1.4]	[0.6;1.7]	[0.3;1.2]	[-0.1;1.1]	[-0.0;1.5]	[0.5;2.4]			
Sample Size	4654	4654	4654	4885	4885	4885			
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes			
Student controls	Yes	Yes	Yes	Yes	Yes	Yes			

Note: Dependent variables were normalized relative to the distribution of the comparison group (pure control), such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. At the endline survey, students were requested to answer how many hours per day (0, 15 minutes, 30 minutes, 1 hours, 2 hours, more than 2 hours) they spend in each of the following activities: i. studying at home on weekdays; ii. studying at home on weekends; iii. studying at home the day before a test; iv. reading a book; v. reading the newspaper; vi. reading magazines; vii. watching TV; viii. navigating on the internet or social media; and ix. helping with housework. Factor analysis was performed to create three variables of student's behavior: *academic activities* (items i, ii and iii); *reading activities* (items iv., v and vi.) and *other activities* (items vii, viii and ix).

E Appendix – Mechanisms

In this section, we present extra tables on the mechanisms to complement the analysis of section 6. In section 6, Tables 8, 12, and 13 describe heterogeneity analysis by parents' baseline beliefs with respect to their child's GPA. Results are showed for parent's endline accuracy, students' transcripts and test scores, and parents' behavior. In this section, we replicate these results for parents baseline beliefs with respect to their child's attendance, instead of GPA, as showed by Tables E.1, E.2, and Table E.3. Moreover, Table 9 of section 6 shows heterogeneous analysis by students' baseline attendance, and in this section we show a similar analysis, but for students' baseline GPA instead of attendance (Table E.4). Finally, Table E.5 replicates the heterogeneous analysis by parents' endline accuracy showed in section 6 (Table **??**) for parents' accuracy on students' attendance, instead of GPA.

Table E.1: Heterogeneity by parents' baseline beliefs wrt their child's attendance - parents' endline accuracy

	Pessimistic Parents (10.2%)		Accurate parents (35.9%)		Optimistic parents (53.8%)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Accuracy Math Attendance	Accuracy Math GPA	Accuracy Math Attendance	Accuracy Math GPA	Accuracy Math Attendance	Accuracy Math GPA
Salience	-0.08	-0.12	-0.02	0.09	-0.03	0.08
	[0.12]	[0.13]	[0.07]	[0.06]	[0.05]	[0.05]
Information	-0.08	-0.14	-0.07	0.05	-0.04	0.02
	[0.12]	[0.14]	[0.06]	[0.06]	[0.05]	[0.05]
Control Mean	0.29	0.38	0.30	0.24	0.22	0.24
P-value diff. [Info] -[Salience]	1.00	0.88	0.35	0.42	0.67	0.20
% Salience	1.00	0.89	0.29	1.86	0.64	3.25
[IC 90%]	[-0.8;2.8]	[-0.3;2.0]	[-1.1;1.7]	[-1.0;4.7]	[-0.6;1.9]	[-5.9;12.4]
Sample Size	171	171	600	600	898	898
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Parents were asked at baseline to give their best estimate of how many times their child misses school in a period of three weeks. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; more than 5). Administrative data register data on attendance on a quarterly basis (period of $\sim$ 9 weeks). Administrative data was divided by 3 to validate parents' answers. Parents were also asked at endline to give their best estimate of how many times their child missed school and what was their final math GPA in the past quarter. Five categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; 6-8; more than 8) and parents answers for GPA were absolute values from 1-10. Data was then crossed with administrative records and a dummy variable were created, where parents who estimated right received value 1 and those who estimated wrong received value 0.

Table E.2: Heterogeneity by parents' baseline beliefs wrt their child's attendance - transcripts and test score

	Pessimistic Parents (10.4%)				Accurate parents (35.3%)				Optimistic parents (54.2%)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)	(9) Math Attendance (p.p.)	(10) Math GPA (std.)	(11) Promotion Rate (p.p.)	(12) Math Standardized Test (std.)
Salience	0.02	0.21*	0.03	0.11	0.02**	0.06	0.06**	0.11	0.04***	0.14**	0.03	0.08
	[0.02]	[0.11]	[0.03]	[0.15]	[0.01]	[0.07]	[0.02]	[0.10]	[0.01]	[0.06]	[0.02]	[0.08]
Information	0.03	0.16	0.02	0.16	0.02*	0.02	0.06***	0.12	0.03***	0.16***	0.03	0.08
	[0.02]	[0.12]	[0.03]	[0.15]	[0.01]	[0.07]	[0.02]	[0.10]	[0.01]	[0.06]	[0.02]	[0.07]
Control Mean	0.89	-0.01	0.96	0.00	0.87	0.07	0.92	0.02	0.84	-0.05	0.94	-0.01
P-value diff. [Info] -[Salience]	0.40	0.64	0.64	0.63	0.82	0.40	0.54	0.78	0.27	0.70	0.80	0.96
% Salience	0.66	1.25	1.37	0.67	1.07	2.62	0.88	0.86	1.17	0.90	1.10	0.97
[IC 90%]	[-0.0;1.3]	[0.2;2.3]	[-0.3;3.0]	[-0.4;1.7]	[0.6;1.6]	[-7.3;12.5]	[0.6;1.2]	[0.1;1.6]	[0.9;1.5]	[0.5;1.3]	[0.4;1.8]	[-0.1;2.0]
Sample Size	399	399	399	399	1350	1350	1350	1350	2073	2073	2073	2073
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. Parents were asked at baseline to give their best estimate of how many times their child misses school in a period of three weeks. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; more than 5). Administrative data register data on attendance on a quarterly basis (period of $\sim$ 9 weeks). Administrative data was divided by 3 to validate parents' answers.

Table E.3: Heterogeneity by parents' baseline beliefs wrt their child's attendance - parents' behavior

	Pessimistic Parents (10.4%)			Accurate parents (35.3%)			Optimistic parents (54.2%)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Academic Activities	Incentives	Talk	Academic	Incentives Academic	Talk	Academic Academic	Incentives	Talk
Salience	0.16	0.10	0.12	0.06	-0.00	0.17	0.05	0.03	0.03
	[0.20]	[0.21]	[0.19]	[0.11]	[0.11]	[0.11]	[0.10]	[0.10]	[0.09]
Information	0.27	0.03	0.08	0.15	0.18*	0.19*	0.06	-0.00	0.06
	[0.21]	[0.22]	[0.20]	[0.11]	[0.10]	[0.11]	[0.10]	[0.10]	[0.09]
Control Mean	-0.12	-0.08	0.00	0.02	-0.01	-0.00	-0.00	-0.04	0.04
P-value diff. [Info] -[Salience]	0.47	0.68	0.79	0.25	0.03	0.78	0.90	0.65	0.65
% Salience	0.62	2.93	1.51	0.44	-0.01	0.88	0.87	-74.33	0.54
[IC 90%]	[-0.2;1.4]	[-22.2;28.1]	[-3.1;6.1]	[-0.4;1.3]	[-1.0;1.0]	[0.2;1.6]	[-0.8;2.5]	[-28496.5;28347.9]	[-1.1;2.2]
Sample Size	329	324	329	1137	1140	1139	1700	1699	1687
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Parents were asked at baseline to give their best estimate of how many times their child misses school in a period of three weeks. Data was then crossed with administrative records and parents who estimated exactly right were determined as accurate, those who estimated below were determined optimistic and those who estimated above were determined pessimist. Four categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; more than 5). Administrative data register data on attendance on a quarterly basis (period of $\sim$ 9 weeks). Administrative data was divided by 3 to validate parents' answers. Students were asked at the endline survey about their parent's behavior, where they had to state how often their parents engage in certain activities (never, almost never, sometimes, almost always). Out of the 12 questions, factor analysis was used to create 3 variables of parents behavior: academic activities (help with homework, help to organize school material, participate in school-parent meetings, talk to the teachers); incentives (incentivize to not miss school, to not be late, to study and to read); talk (ask about homework, ask about grades, ask about day in school and classes).

Table E.4: Heterogeneity by student's baseline GPA

	≤ Median (54.7%)				>Median (45.3%)			
	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)	(5) Math Attendance (p.p.)	(6) Math GPA (std.)	(7) Promotion Rate (p.p.)	(8) Math Standardized Test (std.)
Salience	0.18***	0.18***	0.08	0.10	0.07	0.08	0.11	0.08
	[0.05]	[0.06]	[0.08]	[0.08]	[0.05]	[0.05]	[0.08]	[0.07]
Information	0.16***	0.13**	0.07	0.19**	0.09	0.10*	0.15*	0.06
	[0.05]	[0.06]	[0.08]	[0.08]	[0.05]	[0.05]	[0.08]	[0.07]
Control Mean	-0.57	0.01	-0.36	-0.00	0.65	-0.01	0.42	0.00
P-value diff. [Info] -[Salience]	0.55	0.17	0.84	0.07	0.70	0.64	0.50	0.71
% Salience	1.12	1.38	1.13	0.53	0.84	0.83	0.76	1.31
[IC 90%]	[0.8;1.5]	[0.8;2.0]	[-0.0;2.3]	[0.1;1.0]	[0.2;1.5]	[0.3;1.4]	[0.2;1.3]	[-0.3;3.0]
Sample Size	2223	2073	2223	2073	1949	2295	1949	2295
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: A GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Students with baseline GPA below or equal to the class median were determined as low-performing, and students with baseline GPA above the median were determined as high-performing for the purposes of this analysis.

Table E.5: Heterogeneity by parents' baseline accuracy wrt attendance

	School Transcripts and Test Scores				Parents' Beliefs	
	(1)	(2)	(3)	(4)	(5)	(6)
	Math Attendance (p.p.)	Math GPA (std.)	Promotion Rate (p.p.)	Math Standardized Test (std.)	Accuracy Math Attendance (p.p.)	Accuracy Math GPA (p.p.)
	Less accurate parents (64.1%)					
Salience	0.03***	0.14***	0.03*	0.06	-0.03	0.06
	[0.01]	[0.05]	[0.02]	[0.08]	[0.04]	[0.04]
Information	0.03***	0.15***	0.03	0.08	-0.05	0.02
	[0.01]	[0.05]	[0.02]	[0.07]	[0.04]	[0.04]
Control Mean	0.85	-0.05	0.94	-0.01	0.23	0.26
P-value diff. [Info] -[Salience]	0.38	0.96	0.59	0.76	0.68	0.32
% Salience	1.14	0.99	1.20	0.83	0.70	3.13
[IC 90%]	[0.9;1.4]	[0.6;1.4]	[0.5;1.9]	[-0.1;1.7]	[-0.4;1.8]	[-7.6;13.9]
Sample Size	2472	2472	2472	2472	1069	1069
	More accurate parents (35.9%)					
Salience	0.02**	0.06	0.06**	0.11	-0.02	0.09
	[0.01]	[0.07]	[0.02]	[0.09]	[0.07]	[0.06]
Information	0.02*	0.02	0.06***	0.12	-0.07	0.05
	[0.01]	[0.07]	[0.02]	[0.10]	[0.06]	[0.06]
Control Mean	0.87	0.06	0.92	0.01	0.30	0.24
P-value diff. [Info] -[Salience]	0.83	0.40	0.54	0.78	0.34	0.44
% Salience	1.07	2.63	0.88	0.86	0.28	1.78
[IC 90%]	[0.6;1.6]	[-7.4;12.6]	[0.6;1.2]	[0.1;1.6]	[-1.1;1.6]	[-0.9;4.4]
Sample Size	1350	1350	1350	1350	600	600
Randomization strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Parents were asked at baseline to give their best estimate of how many times their child misses school in a period of three weeks. Data was then crossed with administrative records and parents who estimated exactly right were determined as more accurate and those who estimated wrong were determined as less accurate. Parents were also asked at endline to give their best estimate of how many times their child missed school and what was their final math GPA in the past quarter. Data was then crossed with administrative records and a dummy variable was created, where parents who estimated right received value 1 and those who estimated wrong received value 0.

F Appendix – Results platform scores

As described in section 3, a web-platform was created specifically to this project. Math teachers from treatment schools were oriented to fill in the platform every week with that week's dimension of students' behavior: attendance, lateness or assignment completion, for a duration of 18 weeks. Teachers filled information regarding student behavior on each dimension considering the past three weeks[32]. The system required teacher to fill in information for all students. In this section we investigate the effect of the program on the platform scores.

Each week, teachers evaluated students using a 4 point scale, where 1 was the minimum and 4 was the maximum. For this analysis, we reversed coded scores for lateness, to investigate the effect on punctuality. For each week, we estimated the following model:

$$Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$$

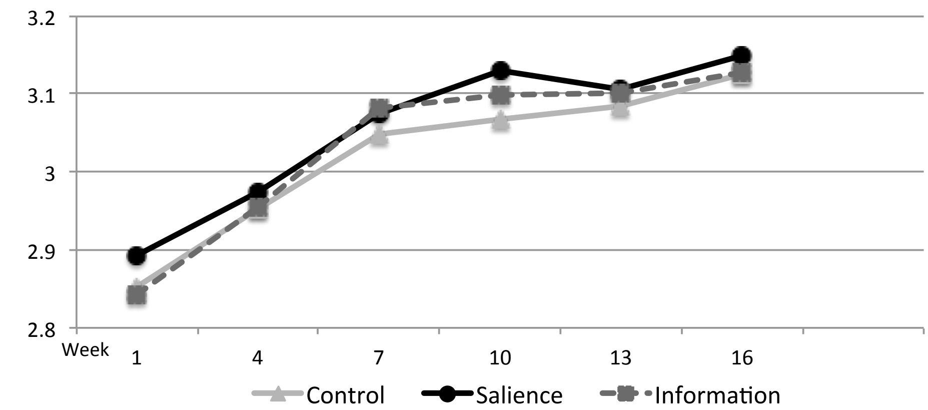
where $Y_{i,c,s}$ denotes the weekly score of each dimension for student i in classroom c of stratum s, the within-class control stand for the reference category (omitted indicator variable), $X_{k,i,c,s}$ is a matrix of student's covariates, θ_s are randomization stratum FE, and $\epsilon_{i,c,s}$ is an error term, clustered at the classroom level. Results are presented in Table F.1, where Panel A show data for attendance, Panel B for punctuality and Panel C for assignment completion. Note that teachers from the pure control schools did not fill the platform and the control group in the graph represents control students in the treated classrooms.

Next, the platform scores of each dimension–attendance, lateness and assignment completion–were averaged and we estimated the same model for the averaged score of each dimension, as showed in Table F.1. The scores were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively.

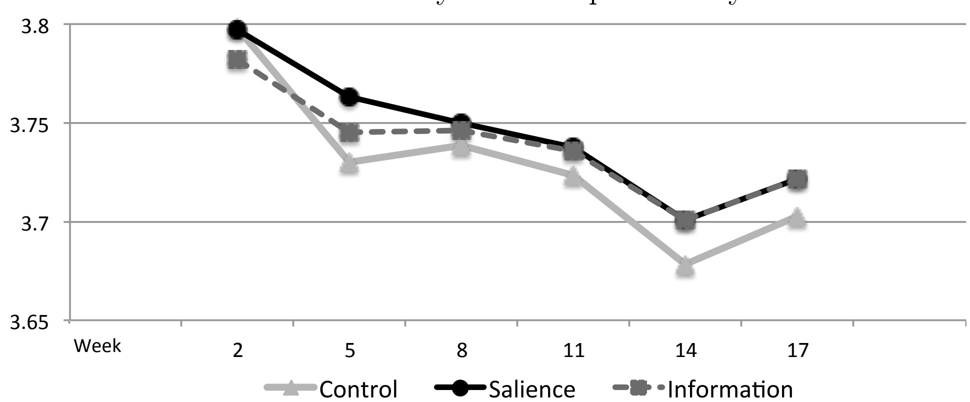
[32]Students have around 6 class of Mathematics per week.

Figure F.1: Weekly effect on platform scores

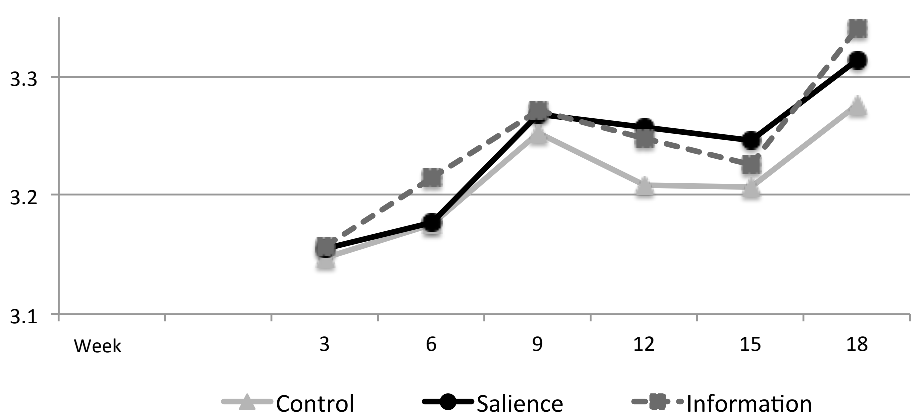
Panel A: Weekly effect on attendance



Panel B: Weekly effect on punctuality



Panel C: Weekly effect on assignment completion



Note: For each outcome and each week, the following equation was estimated: $Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$, where $Y_{i,c,s}$ denotes the weekly score for student i in classroom c of stratum s, the within-class control stand for the reference category (omitted indicator variable), $X_{k,i,c,s}$ is a matrix of student's covariates, θ_s are randomization stratum FE, and $\epsilon_{i,c,s}$ is an error term, clustered at the classroom level. Each week, teachers evaluated students using a 4 point scale, where 1 was the minimum and 4 was the maximum. For this analysis, we reversed coded scores for lateness, to investigate the effect on punctuality.

Table F.1: Results on platform scores - average of all weeks

	(1) Attendance (std.)	(2) On Time (std.)	(3) Assignment Completion (std.)
Salience	0.046**	0.028	0.027
	[0.022]	[0.020]	[0.019]
Information	0.025	0.022	0.044**
	[0.026]	[0.022]	[0.022]
Control Mean	3.043	3.729	3.237
P-value diff. [Info] -[Salience]	0.427	0.822	0.436
% Salience	1.86	1.24	0.61
[IC 90%]	[-1.1;4.8]	[-0.7;3.2]	[-0.1;1.3]
Sample Size	11529	11529	11529
Randomization strata FE	Yes	Yes	Yes
Student controls	Yes	Yes	Yes

Note: The platform scores of each dimension–attendance, lateness and assignment completion–were averaged for each student and then normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. For each score, the following equation was estimated: $Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$, where $Y_{i,c,s}$ denotes the averaged score for student i in classroom c of stratum s, the within-class control stand for the reference category (omitted indicator variable), $X_{k,i,c,s}$ is a matrix of student's covariates, θ_s are randomization stratum FE, $\epsilon_{i,c,s}$ is an error term, clustered at the classroom level, and $\%Salience = \beta_1/\beta_2$.

G Appendix – Spillover

This section presents results on spillover within classroom (peers) and within students (discipline), by comparing the control group of treated classrooms with the pure control group. For each outcome of interest, we estimate the the same model estimated on section 4 (equation 1)[33] but we now show in the table results for the control group of the treated classrooms (and we omit coefficients from the treatment groups).

Table G.1 shows results for the spillover within classroom on students' transcripts and test score, and Table G.2 present results for spillover within student on students' transcripts and test score and parents' endline accuracy.

Table G.1: Spillover within classroom

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Control Within Class	0.018***	0.070**	0.030**	0.085*
	[0.006]	[0.031]	[0.012]	[0.047]
Sample Size	12577	12577	12577	12577
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardize test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. For each outcome of interest, the following model was estimated: $Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \beta_3 Control_{i,c=treated,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$, where $Y_{i,c,s}$ denotes the outcome of interest for student i in classroom c of stratum s; pure control schools stand for the reference category (omitted indicator variable); *Control* assumes value 1 for the control group in treatment schools and 0 otherwise; $X_{k,i,c,s}$ is a matrix of student's covariates; θ_s is a randomization stratum FE and $\epsilon_{i,c,s}$ is an error term, clustered at the classroom level. Only coefficients for the control group is displayed in the table (β_3), coefficients for salience and information were omitted ($\beta_1 and \beta_2$). Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01.

[33] $Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \beta_3 Control_{i,c=treated,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$

Table G.2: Spillover within student

	School transcript and test score			Parent's accuracy	
	(1)	(2)	(3)	(4)	(5)
	Portuguese Attendance (p.p.)	Portuguese GPA (std.)	Portuguese Standardized Test (std.)	Accuracy Portuguese Attendance (p.p.)	Accuracy Portuguese GPA (p.p.)
Salience	0.007	0.066*	0.032	0.009	-0.005
	[0.005]	[0.036]	[0.043]	[0.029]	[0.031]
Information	0.007	0.053	0.047	0.027	0.051*
	[0.005]	[0.036]	[0.043]	[0.029]	[0.031]
Sample Size	12577	12577	12577	3069	3069
Randomization strata FE	Yes	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes

Note: GPA and standardize test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. For each outcome of interest, the following model was estimated: $Y_{i,c,s} = \alpha + \beta_1 Salience_{i,c,s} + \beta_2 Info_{i,c,s} + \beta_3 Control_{i,c=treated,s} + \sum \gamma_k X_{k,i,c,s} + \theta_s + \varepsilon_{i,c,s}$, where $Y_{i,c,s}$ denotes the outcome of interest for student i in classroom c of stratum s; pure control schools stand for the reference category (omitted indicator variable); *Control* assumes value 1 for the control group in treatment schools and 0 otherwise; $X_{k,i,c,s}$ is a matrix of student's covariates; θ_s is a randomization stratum FE and $\epsilon_{i,c,s}$ is an error term, clustered at the classroom level. Only coefficients for the control group is displayed in the table (β_3), coefficients for salience and information were omitted ($\beta_1 and \beta_2$). Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. Parents were asked at endline to give their best estimate of how many times their child missed school and what was their final Portuguese GPA in the past quarter. Five categories were available for parents' answers on attendance (missed 0, 1-2; 3-5; 6-8; more than 8) and parents answers for GPA were absolute values from 1-10. Data was then crossed with administrative records and a dummy variable were created, where parents who estimated right received value 1 and those who estimated wrong received value 0.

H Appendix – Robustness: equalizing the number of times teacher filled the platform by subsample

As showed in Figure H.1, the number of times the teacher filled the platform over the 18 weeks was not equal across the different subsamples. To test if this difference might be somehow affecting the results, we analyze a separate sample, where we equalize the number of times teachers fill the platform by subsample. We do so by eliminating 7 classrooms from the salience only sample, where teachers had filled the platform all the 18 weeks; and 27 classrooms from the subsample containing all treatments (25% salience, 25% ind. info; 25% relative info, 25% control) where teacher participation was low (teachers filled 3 times or less the platform). In this new sample, the average number of times the teacher fill the platform is equal for all subsamples. We then replicate our main results on school transcripts and test score (showed in Table 3) as well as the analyses testing if there is interaction between salience and information (showed in Table 5). Results are showed in tables H.1 and H.2.

Figure H.1: Average number of times teachers filled the platform by subsample during the 18 week period

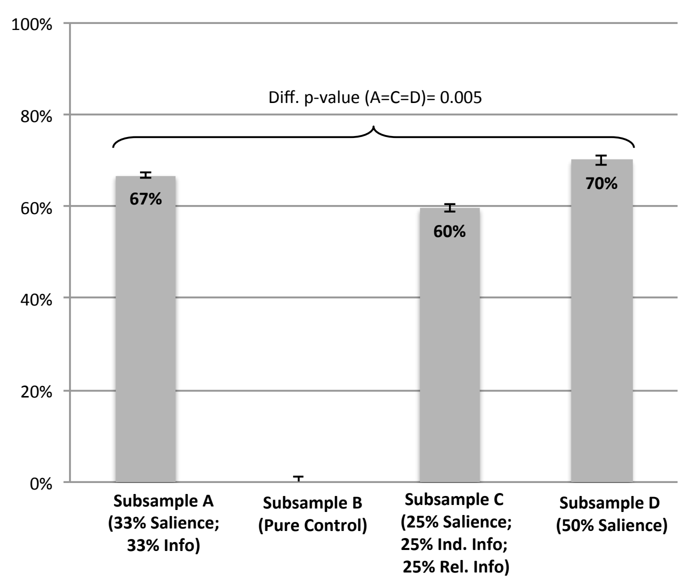


Table H.1: Robustness school transcript and test score - equalizing SMS received by subsample

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.019***	0.085***	0.030***	0.108**
	[0.006]	[0.032]	[0.011]	[0.045]
Information	0.019***	0.070**	0.026**	0.110**
	[0.006]	[0.032]	[0.011]	[0.046]
Control Mean	0.875	0.000	0.938	-0.000
P-value diff. [Info] -[Salience]	0.994	0.368	0.323	0.929
% Salience	1.00	1.20	1.17	0.98
[IC 90%]	[0.8;1.2]	[0.8;1.6]	[0.8;1.5]	[0.6;1.3]
Sample Size	11951	11951	11951	11951
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if p<0.1, ** if p<0.05 and *** if p<0.01. To equalize the number of SMS received, 7 classrooms from the salience only sample were excluded, where teachers had filled the platform all the 18 weeks; and 27 classrooms from the subsample containing all treatments (25% salience, 25% ind. info; 25% relative info, 25% control) where teacher participation were low (teachers filled 3 times or less the platform) where also excluded.

Table H.2: Interactions with information? Equalizing SMS received by subsample

	(1) Math Attendance (p.p.)	(2) Math GPA (std.)	(3) Promotion Rate (p.p.)	(4) Math Standardized Test (std.)
Salience	0.016**	0.068**	0.027**	0.110**
	[0.006]	[0.033]	[0.011]	[0.047]
Information	0.019***	0.070**	0.026**	0.110**
	[0.006]	[0.032]	[0.011]	[0.046]
Salience Only	0.002	0.030	0.002	-0.004
	[0.004]	[0.029]	[0.009]	[0.044]
Sample Size	11951	11951	11951	11951
Randomization strata FE	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes

Note: GPA and standardized test were normalized relative to the distribution of the comparison group, such that the mean and standard deviation of the comparison group is zero and one, respectively. A dummy variable for the control group within class was also included. Standard error clustered at the classroom level. Significance levels are denoted by * if $p<0.1$, ** if $p<0.05$ and *** if $p<0.01$. To equalize the number of SMS received, 7 classrooms from the salience only sample were excluded, where teachers had filled the platform all the 18 weeks; and 27 classrooms from the subsample containing all treatments (25% salience, 25% ind. info; 25% relative info, 25% control) where teacher participation were low (teachers filled 3 times or less the platform) where also excluded.