ACTIVATING PARENTAL MANAGERIAL CAPITAL: HOW ROLE CLARITY IMPROVES EDUCATION OUTCOMES [1]

Mahounan Yedomiffi

Dartmouth College

Version 2026-03-25

Please click here for the most recent version!

Abstract

Why do some households fail to translate educational investments into learning? This paper introduces parental managerial capital—the ability to organize children's time, monitor effort, and sustain routines that support learning—as a productive input in human-capital accumulation. In Benin, I first document a large intention–implementation gap: most parents say that monitoring attendance, enforcing lesson review, and tracking progress are their responsibility, yet many do not implement the corresponding routines. I then test whether relaxing managerial-capacity constraints improves outcomes using a randomized phone-based coaching intervention that provides a simple, literacy-independent managerial technology. The intervention substantially narrows the intention–implementation gap and raises achievement: GPA increases by 0.11 standard deviations, and grade completion rises by 12 percentage points among students initially at risk of repetition. Although the program provides no grades or performance records, treated parents become more accurate about their child's academic standing, suggesting that stronger household management generates information endogenously through monitoring and school contact. The findings identify household managerial capital as an important and an overlooked source of educational inequality.

JEL Codes: I21, I24, I25, O15, C93

Keywords: Parental managerial capital; household production; randomized field experiment; secondary education; Sub-Saharan Africa (Benin).

1. I thank Richard Akresh, Dan Bernhardt, Alex Eble, Eric Edmonds, Dean Karlan, Paul Novosad, Adam Osman, Germán Reyes, Rebecca Thornton, Chris Udry, and Leonard Wantchekon for their comments and suggestions. I also thank participants at the Harvard PIER Seminar, UIUC AMRL, NEUDC, MWIEDC and BREAD. I thank Togbedji Gansey, and Iranlowo Abakpin Lawin for their research assistance. I gratefully acknowledge financial support from Rebecca Thornton, Enthony Violaris, UIUC graduate college dissertation travel grant, UIUC Nelle M Signor Fellowship, UIUC Boltz Fellowship, UIUC Applied Micro Small grant, and UIUC Ferber & Sudman Award for survey research. This study receives authorization from the Ministry of Secondary Education under Authorization Number 145/MESTFP/DC/SGM/DESG/SA and UIUC IRB approval number 23204. All errors are mine. Contact Information: mahounan.yedomiffi@dartmouth.edu

1 Introduction

Why do some households fail to translate educational investments into learning? This paper introduces parental managerial capital—the ability to organize children's time, monitor effort, and sustain routines that support learning. In all schooling systems, learning depends in part on how households organize children's time and effort outside the classroom. In the setting I study, most parents report that ensuring attendance, monitoring progress, and enforcing lesson review are their responsibility, yet far fewer report implementing the corresponding practices. This intention–implementation wedge suggests that the puzzle is not primarily whether parents value their children's education, but why broadly endorsed responsibilities often remain unimplemented.

This paper proposes a simple interpretation of the puzzle: households differ in managerial capital in much the same way that firms do. Just as management practices drive productivity differences across firms with similar technologies (Bloom and Van Reenen 2007; Bloom and Van Reenen 2010), gaps in household managerial capacity—the ability to structure time, monitor effort, and guide behavior—can generate persistent differences in human-capital accumulation.[2] The key friction I emphasize is not a lack of aspirations, but an implementation failure. Parents may endorse responsibility yet still face substantial start-up and follow-through costs—planning, attention, and coordination costs—of turning that responsibility into sustained routines. As a result, households can operate below their human-capital production frontier even when monetary inputs and institutional access are held fixed.

Education provides a useful setting to study this household managerial-capital mech-

2. There is a well-documented gap in school performance between children from low- and high-income households (Filmer and Pritchett 1999; Glewwe and Kremer 2006).
This interpretation is consistent with evidence that maternal education affects how children's time is organized at home: Andrabi, Das, and Khwaja 2012 show that children of more educated mothers spend substantially more time on educational activities outside school.

anism because household routines are observable and weak institutional feedback makes low managerial capacity costly in ways that are measurable in surveys and administrative records. First, using baseline data, I document a sizable intention–implementation gap in managerial practices (the role–action wedge). The wedge is systematically larger among illiterate parents and for higher-cost practices that require coordination and follow-through, consistent with the view that start-up and implementation costs—rather than low aspirations—limit household investment. I then use experimental variation to estimate the returns to narrowing this wedge.

Second, to identify the returns to household managerial capital, I implemented a randomized field experiment across 40 public secondary schools in southern Benin. In 20 intervention schools, students were randomly assigned within class to treatment or control, yielding a sample of 2,094 households with 8^{th}-grade students. The remaining 20 schools served as pure controls to assess spillovers. The intervention provided an interactive managerial "technology" delivered through phone-based coaching. Over four months, parents in the treatment group received, on average, ten structured calls from trained facilitators. The first call reframed the household role from primarily financing schooling to managing the child's time and effort. Subsequent calls combined practical guidance with follow-up and troubleshooting: parents discussed what they tried since the previous call, raised constraints from their household and school context, and the facilitator helped adapt a set of low-burden practices—establishing study routines, temporarily reallocating chores around exams to protect study time, and monitoring attendance through occasional school visits.[3]

3. The action menu was refined through qualitative work with school officials, parents, and students, and reflects practices that more educated parents and teachers in the same system consistently described as feasible and high leverage. I also draw on my own experiences growing up in rural Ivory Coast, where students who stayed on track tended to be those whose time and effort were structured by an adult, while many others drifted despite aspirations in contexts where parents often viewed their role as largely limited to paying fees.

A useful benchmark is the literature on parental information and attention interventions, which shows that report cards, test scores, reminders, and other salient signals can affect educational decisions by changing what parents know. A natural maintained assumption in that framework is that once parents are informed, they can translate that information into action. My design instead holds fixed direct information provision and varies parents' capacity to act: the intervention provides a literacy-independent managerial technology for monitoring effort, structuring time, and sustaining follow-through. This distinction matters empirically because it allows me to distinguish between two views of parental information: one in which information is an exogenous input into household choices, and another in which information is itself generated endogenously through stronger household management.

The intervention shifts the household inputs it targets. Relative to the control group, treated households adopt managerial routines at higher rates—especially higher-cost routines and school coordination—narrowing the intention–implementation wedge. Treated students are more likely to be relieved of chores near exams, based on both parent and student reports. These changes based on self-reported outcomes are corroborated by objective outcomes: treated students are less likely to miss school, and parents' beliefs about academic standing become more accurate even though the program provides no performance information. Together, these patterns indicate that coaching reduces the gap between endorsed responsibility and implementation and induces endogenous information acquisition through increased monitoring and school contact. These behavioral changes translate into meaningful academic gains both in terms of improved GPA and grade completion. Effects are especially pronounced among students at risk of grade repetition. One year later, dropout rates fall, with the largest reductions among girls and among children of illiterate parents. These improvements occur without requiring tutoring or additional educational spending; they instead reflect changes in how children's time is structured and how effort is organized at home and through selective coordination with schools. In this sense, the intervention raises the productivity of existing educational inputs by shifting the time–effort margin.

Several additional results support the managerial-capital interpretation. I find no evidence that treatment operates through changes in financial expenditures on schooling: treated households do not reduce tuition payments, delay payments, or become less likely to pay fees in full. Instead, the main adjustment margin is within-household time reallocation, with treated children more likely to be excused from chores around exam periods while parents and siblings temporarily absorb those tasks. I also find no evidence of negative spillovers within classrooms. Finally, the persistence of higher-cost management practices and the longer-run dropout effects suggest changes that go beyond short-run reactivity to program contact.

This paper contributes to our understanding of how households implement responsibilities they already endorse when implementation is costly. Parents may agree that monitoring effort, structuring time, and supporting learning are part of their role, yet lack the managerial capacity required to translate these intentions into sustained routines. I conceptualize this capacity as household managerial capital—a productive input in human-capital accumulation that governs how effectively families organize time, monitor effort, and sustain accountability. This perspective parallels the management literature in firms, where differences in management practices explain productivity gaps across organizations facing similar technologies and resources (Bloom and Van Reenen 2007; Bloom and Van Reenen 2010). To my knowledge, the paper provides the first evidence using exogenous variation in parental managerial capacity to establish a causal link between household managerial capital and children's educational outcomes. The results suggest that analogous differences in household managerial capacity can generate persistent inequality in children's outcomes even when schooling access and monetary inputs are held fixed.

The paper also contributes to the literature on information frictions in parental decision-making. A large literature shows that report cards, test scores, reminders, and other salient signals can affect educational decisions by changing what parents know.[4] A natural maintained assumption in that framework is that once parents are better informed, they can translate that information into action. My design instead holds fixed direct information provision and varies parents' capacity to act. Treated parents become more accurate about their child's academic standing even though they receive no grades or administrative records, implying that parental information can be generated endogenously through stronger household management. Informational deficits may therefore reflect not only missing external information, but also weak household routines for generating and acting on information.

Finally, the paper contributes to the education literature on how parents respond to information about children's academic performance. A large literature shows that providing parents with information about test scores, attendance, or school quality can affect educational decisions and achievement by updating parental beliefs and attention (e.g., Andrabi, Das, and Khwaja 2017; Bergman 2021; Rogers and Feller 2018). In Dizon-Ross

4. Examples include interventions that provide information about student performance, attendance, or school quality, as well as related designs that shift attention or salience (e.g., Dizon-Ross 2019; Barrera-Osorio et al. 2020; Hastings and Weinstein 2008; Nguyen 2008; Andrabi, Das, and Khwaja 2017; Rogers and Feller 2018; York, Loeb, and Doss 2019; Bergman 2021; Berlinski et al. 2021; De Walque and Valente 2023). Many parental-support interventions also bundle child-specific performance information with information about the education system or recommended behaviors for parents. By contrast, the intervention studied here provides no performance information and focuses on literacy-independent managerial practices that help parents monitor and support learning through feasible routines.

2019, parents who receive more accurate information about their children's performance adjust investments across children, in part by reallocating resources away from lower-performing students. My results point to a different margin. In this setting, parents become more accurate about their child's academic standing even though the intervention provides no performance information, and they respond by increasing support for students most at risk—consistent with the largest first-year gains occurring among children near repetition margins and the largest longer-run dropout reductions occurring among children of illiterate parents. The evidence suggests that when households possess a feasible managerial technology for supporting learning, better information need not lead to disengagement from weaker students; instead it can enable targeted support where needs are greatest.

The remainder of the paper is structured as follows. Section 2 presents the context and experimental design. Section 3 outlines the survey and administrative data collection. Section 4 presents the empirical strategy, and Section 5 reports the empirical results. Section 6 provides the cost-benefit analysis. Section 7 concludes.

2 Context and Experimental Design

2.1 Context and sample

This study takes place in Benin, a French-speaking West African country with a centralized public secondary school system. I focus on 8^{th}-grade students (the second year of lower secondary), a stage at which course loads intensify, time and effort management begins to shape subsequent progression. The study is implemented in Atlantique, a densely populated region in southern Benin with a large concentration of public secondary schools.

From 95 eligible public secondary schools, I randomly selected 40 schools (20 intervention, 20 pure control). These schools averaged four 8$^{\text{th}}$-grade classes and approximately 50 students per class.

The sample consists of 2,094 households (from an initial 2,256), drawn from the 20 intervention schools. Table A.1 reports descriptive statistics. Households have six members on average and parental education is low: mothers have completed 2.29 years of schooling on average and fathers 4.90 years. While 75.9% of households have electricity (including solar) and 75.5% own a radio or TV, only 54.8% have a desk at home where children could study. Nearly all households (98.4%) primarily speak local languages at home. Mobile phone access is near universal (97.7%), making phone-based delivery feasible at scale.

2.2 The need for improved parental managerial practices

In Benin's public secondary schools, large classes and limited individualized feedback compress the scope for in-class remediation, so achievement depends importantly on whether students sustain review and effort outside the classroom. Yet schools provide households with little actionable guidance on how to structure learning at home, and formal performance feedback arrives infrequently. In this environment, the relevant household margin is rarely private tutoring: most families cannot afford it, and many parents lack the literacy needed to provide content assistance directly. Consistent with this, only 5% of students in our sample had a private tutor at baseline. The high-leverage margin may therefore be organizational rather than curricular—protecting study time, establishing routines, monitoring progress, and enforcing attendance—inputs that raise the productivity of time devoted to learning.

A motivating puzzle: responsibility is widely endorsed, yet managerial practices are scarce. Baseline survey data shown in Figure III reveal a sharp disconnect between stated responsibility and concrete implementation. When asked who should ensure that children review lessons, monitor progress, keep notebooks up to date, and avoid missing class, about 90–93% of respondents selected parents as responsible across domains. Yet corresponding practices are much less common. Only 19% of households report having a structured lesson-review schedule at home and only 11% report visiting the school to check attendance; by contrast, 63% report asking children about test scores and 39% report checking the child's notebook.[5] In other words, many parents endorse responsibility as a norm, but relatively few adopt the higher-leverage routines that convert time at home into learning.

A simple representation of the role–action wedge. To discipline interpretation, it is useful to summarize these patterns with two indices.

Let $d \in \{\text{Attendance, Progress, Review, Notebook}\}$ index domains. Define role endorsement

$$R_i \equiv \frac{1}{4}\sum_d r_{id}, \qquad r_{id} \in \{0,1\}, \tag{1}$$

and implementation of managerial practices

$$P_i \equiv \frac{1}{4}\sum_d p_{id}, \qquad p_{id} \in \{0,1\}. \tag{2}$$

5. The implementation measures are designed to capture concrete actions that are feasible in this setting, rather than abstract intentions.

A natural measure of the intention–implementation wedge is the share of domains in which responsibility is endorsed but the corresponding action is not taken:

$$G_i^{\text{unmet}} \equiv \frac{1}{4} \sum_d \mathbf{1}\{r_{id} = 1,\ p_{id} = 0\}. \tag{3}$$

This definition focuses on *unmet responsibility* and therefore avoids negative values that arise under the alternative difference measure $R_i - P_i$ when a household reports taking an action despite not assigning responsibility to parents in that domain.[6]

In this setting, role endorsement is near-universal while implementation varies sharply across domains. In the baseline data, the average role-endorsement index is high ($\bar{R} \approx 0.91$), whereas the average implementation index is much lower ($\bar{P} \approx 0.33$), implying that many households endorse parental responsibility without adopting the corresponding routines. Consistent with this, the average unmet-responsibility wedge, G_i^{unmet}, is large: on average, parents endorse responsibility but do not take the corresponding action in 63 percentage points of domains.

Figure V presents the wedge by baseline characteristics. The wedge is substantially larger for illiterate parents (70 percentage points) than for literate parents (56 percentage points), consistent with higher startup and cognitive costs of translating intentions into routines. By contrast, the wedge is similar by the gender of the focal student (63 percentage points for both girls and boys). It is also somewhat larger among students at the bottom of the baseline performance distribution (65 percentage points) than among

6. For transparency, I separately track the share of domains in which the household reports *action without endorsed responsibility*,

$$G_i^{\text{over}} \equiv \frac{1}{4} \sum_d \mathbf{1}\{r_{id} = 0,\ p_{id} = 1\}. \tag{4}$$

In the baseline data, this "over-implementation" pattern is present but uncommon so the main descriptive patterns are driven by unmet responsibility rather than by action without endorsement. By construction, $R_i - P_i = G_i^{\text{unmet}} - G_i^{\text{over}}$ at the domain-average level.

those in the top quartile (60 percentage points), consistent with weaker baseline structure in households whose children are already struggling.

Conceptual interpretation: an intention–implementation wedge in the home production of learning. Managerial practices matter because they raise effective study time by protecting time from competing demands and by increasing effort and accountability during that time. A parsimonious way to capture this is:

$$S_i = S_i^0 + \phi P_i, \tag{5}$$

where S_i denotes effective study time and S_i^0 captures baseline study conditions. Achievement is increasing in effective study time,

$$Y_i = \alpha + \beta S_i + u_i, \tag{6}$$

and downstream outcomes such as promotion and dropout respond when achievement crosses relevant thresholds.

The baseline role–action patterns suggest that the central friction is not whether households endorse parental responsibility, but whether they can operationalize that responsibility into stable routines. Many mechanisms could in principle generate such a wedge. I focus on two that emerged consistently in the focus groups with teachers, parents, school leaders, and students, and that map directly into the practices we measure and the coaching content.

First, respondents—especially low-literacy parents—often described a narrow model of the home production of learning in which parental inputs are productive only through tutoring or financial expenditures. In this view, routine-setting, monitoring, and time-allocation discipline have low perceived marginal returns, so they are crowded out by competing demands even when responsibility is endorsed. Second, even when parents expressed a desire to "do more," they frequently lacked an implementable managerial technology: concrete templates, scripts, and contingency plans that lower the fixed costs of getting started and make follow-through feasible under illiteracy and time scarcity.

These two forces help rationalize the stark domain-level differences in Figure III. Low-friction actions that can be taken immediately (e.g., asking about test scores) are relatively common, whereas practices that require durable routine formation (e.g., establishing a regular review schedule) or transaction/coordination with the school (e.g., visiting to check attendance) are much rarer.

Implementation costs and the margins targeted by coaching. The intervention targets parental managerial practices that differ in their implementation costs. To discipline interpretation, it is useful to distinguish three categories: (i) low-friction conversational checks that can be implemented immediately, (ii) home-based verification and enforcement that require time and follow-through, and (iii) actions with substantial startup or transaction/coordination costs. Using the four canonical role–action pairs measured at baseline, unmet responsibility rises sharply across these three categories: 0.33 for conversational checks, 0.54 for home verification/enforcement, and 0.78 for high-startup/transaction practices (Figure IV). This monotone gradient suggests that the wedge is primarily an implementation problem rather than a lack of endorsed responsibility, and it provides a cost-based organizing framework for the mechanism and heterogeneity results that follow.

A compact way to represent these frictions is that parents adopt and sustain routines when perceived benefits exceed implementation costs:

$$R_i \cdot \Delta_i(P) \;\geq\; \kappa_i + c(P), \tag{7}$$

where $\Delta_i(P)$ summarizes perceived returns to managerial practices, κ_i captures startup and transaction barriers (e.g., low literacy, limited knowledge of school timelines, difficulty contacting schools, and the cognitive burden of planning), and $c(P)$ denotes ongoing time and enforcement costs of monitoring. In this view, high R_i can coexist with low P_i when either perceived returns are low or implementation barriers are high.

Why this matters for who benefits. This framework implies that interventions that raise $\Delta_i(P)$ and reduce implementation costs should primarily increase managerial practices, with larger effects when (i) implementation barriers are high—for example, in low-literacy households—and (ii) the marginal return to increased effective study time is steep—for example, for students near promotion or dropout margins. More generally, in settings where role endorsement is not near-universal, improving outcomes may require moving both margins (shifting responsibility beliefs and implementation).

From a motivating wedge to a broader set of practices. The role–action measures in Figure III are intentionally parsimonious: they capture a small set of domains for which we measure role endorsement and implementation practices at baseline. The intervention, however, targets a broader bundle of parental managerial practices. In practice, parents can influence learning through multiple actions at home (e.g., protecting study time, monitoring task completion, limiting distractions) and through selective engagement with the school (e.g., checking attendance, clarifying expectations, following up with teachers). Conceptually, the framework extends naturally by letting P_i summarize implementation across a larger set of actions, potentially with action-specific costs. The central idea remains the same: in this setting responsibility is widely endorsed, and the binding constraint is translating that responsibility into sustained, feasible routines.[7]

Intervention overview. To target this intention–implementation wedge, I designed a phone-based coaching intervention that complements already-high role endorsement by (i) sharpening parents' beliefs about the payoff to managing children's time and effort (increasing $\Delta_i(P)$) and (ii) providing a literacy-independent managerial technology—concrete routines, scripts, and troubleshooting support—that lowers startup and transaction barriers to follow-through (reducing κ_i). I describe the experimental design next and then return to implementation details of the intervention.

2.3 Experimental Design

2.3.1 Sampling

In each intervention school, parents of 8$^{\text{th}}$-grade students were invited to attend an in-person baseline meeting held between November 22 and December 17, 2022 (Figure II). To organize logistics and limit wait times, students were ordered alphabetically within class, and school staff contacted parents sequentially following this list. Meeting times were scheduled based on parent availability and the number of enumerators assigned to each school, as the combined consent and baseline survey process required approximately one hour per parent.

7. Formally, letting $a \in \mathcal{A}$ index the full set of recommended practices, define

$$P_i^{\mathcal{A}} \equiv \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} p_{ia}. \tag{8}$$

The intervention is designed to increase $P_i^{\mathcal{A}}$ by raising perceived returns and reducing implementation costs, with especially large scope for practices that involve high startup or transaction costs.

Outreach proceeded down the alphabetical list until the target number of participants was reached. When a parent was unavailable or reported a scheduling conflict, staff moved to the next parent on the list. Once the target sample size was attained, schools did not continue outreach. As a result, approximately 55% of parents attended the baseline meeting and constitute the study sample. Among the remaining 45%, most were not contacted, and a small share were contacted but could not attend due to scheduling conflicts.

Because outreach followed a predetermined alphabetical ordering and did not condition on academic performance, parental engagement, or observable student characteristics, selection into the study sample reflects logistical constraints rather than screening on outcomes. Section 5.7 assesses external validity by comparing students whose parents did not attend the baseline meeting to control-group students. Baseline test scores and year-end grade completion rates are similar across these groups, and spillover patterns are also comparable, suggesting that the enrolled sample is broadly representative of the eligible population.

The final sample includes 2,094 students across 20 schools, with 1,038 in the treatment group and 1,056 in the control group, as summarized in Figure I.

2.3.2 Treatment: Content and Implementation

The intervention was designed to activate a neglected input in the education production function: parental managerial practices. The intervention therefore targeted two related constraints: (i) low perceived returns to routine-setting and monitoring when parents cannot tutor, and (ii) high implementation costs—fixed costs of getting started.

Delivery and core design feature: interactive coaching. The treatment consisted of coaching sessions delivered through phone calls over four months (January 11 to May 20, 2023; Figure II). Sessions followed structured scripts and were conducted by facilitators trained to combine standardized guidance with household-specific problem-solving. A central design feature is that the calls were interactive rather than one-way information. Facilitators served as a sounding board: parents described constraints they faced, selected actions to implement, and then revisited those choices in subsequent calls. This feedback loop is intended to (i) raise perceived returns to household organizational inputs by making the link between specific routines and learning outcomes explicit, and (ii) reduce implementation costs by helping parents adapt the same practice to their time constraints, literacy, and household structure.

Where the action menu comes from. The set of recommended practices was not chosen ad hoc. It was built from the in-depth focus groups and refined iteratively with teachers and school staff. Three criteria guided selection. First, practices had to be literacy-independent: implementable without subject knowledge. Second, they had to be high leverage in the local production process—i.e., plausibly increase effective study time by improving time allocation discipline, monitoring, and accountability. Third, practices had to be institutionally compatible: consistent with school routines and teacher expectations, so that households could implement them without requiring changes in school inputs. Consistent with these criteria, literate parents disproportionately described using these same practices in daily management, and teachers and administrators repeatedly emphasized them as margins on which households could influence effort and preparation.

Step 1: Clarify the "why" and select an achievable goal (raising perceived returns). The initial call reframed parents' role in a system with weak feedback loops. Facilitators clarified the responsibilities of government, teachers, and administrators and emphasized constraints students face in large public classrooms: limited individualized feedback and infrequent performance updates. The central message was that home study is not elective in this environment: absent routines and oversight, many students will not consistently study on their own, and poor performance can persist unnoticed until it is difficult to recover. Even when parents cannot teach content, they can raise performance by increasing the productivity of study time through routine formation, monitoring, and accountability. Facilitators then helped parents translate this logic into a small, concrete commitment for the coming days (e.g., one monitoring action and one time-allocation action), with the goal of building a feasible routine rather than maximizing intensity immediately.

Step 2: Build accountability in a way that fits the household (reducing startup and monitoring costs). Early calls introduced a menu of literacy-independent strategies anchored in common school practices: establishing a regular study time; asking about test dates and test scores; checking notebooks for missing work; supervising homework completion; and adopting simple rules that discipline time use (e.g., "review before leisure"). A central component was managing competing demands on time, especially household chores. Facilitators discussed how parents could protect study time by occasionally suspending or delaying a chore the child typically performs during high-stakes periods (e.g., before exams). This margin is directly measured in the surveys: we ask whether parents ever asked the child not to perform a usual chore in order to study, and, conditional on doing so, who substituted into chores. In the analysis, I summarize substitution into two economically relevant categories: additional chores performed by an adult (parents) versus by other children (siblings).

Subsequent calls focused on translating intentions into implementable routines. Facilitators helped parents choose actions that matched their time constraints and household structure and offered practical ways to sustain follow-through. If a parent was not home in the evening, the facilitator helped identify a substitute monitor (another adult or an older sibling) or suggested shifting review to an earlier window. If chores crowded out study time, facilitators discussed concrete alternatives (e.g., shifting tasks to weekends, reallocating across siblings, or compressing chores during non-exam periods). The objective was not strict compliance with a template, but an accountability mechanism the household could maintain.

Step 3: Recognize setbacks and troubleshoot (iterative adaptation). Each call included a brief review of what the parent tried since the previous call, what barriers arose, and what alternative strategy might work better. Common issues included difficulty reaching school officials, uncertainty about when to approach teachers, fatigue after late work hours, and child resistance. Facilitators responded with concrete substitutes (e.g., visiting on days when officials are more likely to be present, using class representatives as intermediaries, shifting the timing of review, simplifying the routine, or switching from supervision to notebook verification when time was scarce). This iterative structure is designed to prevent setbacks from being interpreted as failure and to help households converge to a stable routine suited to their constraints.

Action menu organized by implementation costs. To connect intervention content to the conceptual framework, it is useful to group practices by their primary implementation costs in this context:

1. **Low-friction conversational checks (low startup, low monitoring):** initiating brief conversations about school and asking about test scores and upcoming test dates.

2. **Home verification and enforcement (higher monitoring and follow-through):** supervising homework completion, checking during lesson review to ensure the child is studying, and checking notebooks for missing work or blank pages.

3. **High startup or high transaction/coordination actions (fixed costs and/or school contact):** establishing a regular study schedule and household rules around study time; reallocating chores around exams to protect study time; and, when feasible, visiting the school to verify attendance and speaking with teachers or school staff to obtain updates.

These practices are intended to increase effective study time—the product of time and effort—by reducing time leakage (e.g., from chores or distractions) and increasing accountability during study time.

Implementation fidelity. On average, parents received 10 successful calls (88% success rate). Facilitators recorded brief call notes to ensure continuity in follow-up and to structure household-specific troubleshooting across calls.

3 Data and Outcomes

The analysis combines parent and student surveys with school administrative records. Survey data were collected in three rounds from parents (baseline, endline, and follow-up) and in one round from students (endline). Administrative data provide repeated measures of test scores, grade point averages, attendance/absenteeism, tuition payments, grade completion, and dropout a year after the intervention ended.

3.1 Survey data

Baseline parent survey. Baseline parent surveys were administered during in-person enrollment meetings between November 22 and December 17, 2022 (Figure II). We surveyed 2,256 parents/guardians; the analysis sample includes 2,094 students after excluding cases in which respondents withdrew consent during baseline procedures and before treatment assignment. The baseline questionnaire measures household background and the key inputs and information sets in the home production of learning: household managerial practices, knowledge of the school system, and beliefs and expectations about the child's performance.

Endline parent and student surveys. Endline data were collected June 1–30, 2023. We re-administered the core modules and added measures of implementation constraints and household time allocation, including chore reallocation around study and exams. We interviewed 1,921 parents and 1,902 students. To limit concerns about differential reporting by treated parents, the student survey included parallel measures of household managerial practices and chore allocation.

Follow-up parent survey. One year after the intervention ended, we conducted a short follow-up survey with parents to measure persistence of managerial practices.

3.2 Administrative data and outcome construction

We digitized school administrative records for each enrolled student.[8] These records provide test scores by subject, semester GPAs, and year-end GPAs. Administrative collection occurred in five rounds: baseline test scores; first-semester test scores; first-semester GPA; second-semester test scores; and second-semester and year-end GPA. These repeated measures allow the analysis of learning dynamics over the school year.

Administrative records also include student absenteeism and tuition payments. One year after the intervention, we returned to schools to measure subsequent enrollment and dropout.

Outcomes. To test the mechanism, I first measure the intention–implementation gap in parental managerial inputs (the role–action wedge) and how the intervention changes it. Survey outcomes capture both role endorsement and the adoption of management practices, as well as parents' knowledge about academic standing and household time allocation, including chore reallocation around exams. I then examine impacts on student learning and school progression—subject-level test scores, GPA, and grade completion—and follow students one year later to measure enrollment and dropout. Because the intervention targets parents management practices around their child's education, it can affect performance broadly across the curriculum. I therefore estimate impacts by subject (Biology, Physics, Mathematics, Language, and History–Geography) and report pooled indices for STEM (Biology, Physics, Mathematics) and Humanities (Language, History–Geography). Secondary outcomes include absenteeism and tuition payment measures, which help assess alternative channels and potential crowd-out.

8. Administrative records are ledger books maintained by schools, with one page per student and one book per class. We photographed all pages and digitized them. A separate team conducted verification checks on a random sample of entries.

4 Empirical Strategy

4.1 Identification

The empirical analysis proceeds in two steps, reflecting the paper's mechanism and outcome structure. I first estimate the effect of the intervention on household managerial inputs—parents' organizational practices around schooling, including monitoring and accountability at home, coordination with the school, knowledge of performance, and time-allocation discipline (including chore reallocation during high-stakes periods). I then examine whether these changes in household inputs translate into improvements in student outcomes.

The randomized experimental design provides a strong identification strategy and allows me to attribute any differences in outcome indicators between the treatment and control groups to the managerial practices provided by the coaching program. The effect I measure here is the intention to treat effect, and I measure it by estimating equation 9.

$$Y_{it} = \alpha + \beta_0 Y_{i0} + \beta_1 Treat_i + \beta_2 X_i + \epsilon_{it} \tag{9}$$

In equation 9, Y_{it} measures the outcome of interest for individual i at time t, Y_{i0} is the baseline value of the outcome of interest for individual i whenever feasible, $Treat_i$ is a dummy variable equal one if the individual i is treated and 0 otherwise. I include class-specific parental literacy strata fixed effect. I utilize the double-post LASSO procedure outlined in Belloni, Chernozhukov, and Hansen 2014 to select baseline covariates X_i. Standard errors are clustered at the individual level. My coefficient of interest is β_1. β_1 measures the effect of being assigned to the treatment group on Y. I also perform heterogeneity analyses using a fully interacted version of equation 9.

I examine heterogeneity along three dimensions: student baseline performance, and two pre-specified characteristics—parental literacy and student gender.

4.2 Outcome families and index construction

Several mechanism outcomes are measured through batteries of related items. Because the intervention targets a specific set of household organizational inputs—rather than a single behavior—I report effects on summary indices that aggregate items within a family. This approach disciplines interpretation and avoids overemphasizing any one action.

For indices, given these are binary outcomes, I use equally weighted averages so coefficients can be interpreted as percentage-point changes in the share of practices adopted within a family. I group practices following the same implementation cost taxonomy introduced above (conversational checks; home verification/enforcement; high startup/transaction) Individual action-level estimates are reported in the Appendix.

5 Empirical Results

5.1 Baseline Balance

Using baseline survey and secondary data, I compare the treated and control groups on a set of characteristics and baseline outcomes. Table I shows the results. In the table, I present the control group means, the treated group means, the difference in means between treated and control, and the p-value of the equality test between treatment and control. I can reject the null hypothesis that the treatment arm is equal to the control for only two variables out of 29. I also run a regression where I regress all 29 variables on treatment status, including strata fixed effects and cluster standard errors at the individual level. The p-value of global significance is 0.32, showing that the model is not globally significant. These results confirm that the randomization successfully created two groups that are similar in their characteristics; hence, any difference can be attributed to the coaching intervention. The subsequent subsections present the results of the impact evaluation of my intervention.

5.2 Indicators of Improved Parental Managerial Practices

This section provides evidence that the intervention reduces the intention-implementation gap, and provides evidence of a broader improvement in parental managerial practices. It also provides evidence of changes in objective indicators—such as student absenteeism and parental knowledge of academic progress—that corroborate the reported behavioral changes.

5.2.1 Reducing the intention–implementation gap

We begin by testing whether the intervention reduces the intention–implementation wedge documented at baseline. For each canonical domain d (Attendance, Progress, Review, Notebook), define $g_{id} = \mathbf{1}\{r_{id} = 1,\ p_{id} = 0\}$, which equals one when the respondent assigns primary responsibility to parents in domain d at baseline but the corresponding action is not reported. The household-level wedge is $G_i^{\text{unmet}} = \frac{1}{4}\sum_d g_{id}$. Since role endorsement is measured at baseline and is near-universal, changes in G_i^{unmet} primarily reflect changes in implementation rather than shifts in stated responsibility.

Panel A of Table II shows that the intervention reduces this wedge. The overall unmet-responsibility index falls by 0.08 relative to a control mean of 0.46, a reduction of about 17%. The wedge declines across all three implementation-cost categories: by

0.09 for low-friction conversational checks (control mean 0.20), by 0.11 for home verification/enforcement behaviors (control mean 0.49), and by 0.06 for high-startup/high-transaction actions (control mean 0.57). While reductions are present in each category, the cost-based decomposition is informative for interpretation: the baseline wedge is largest in higher-cost domains, which are precisely the practices the intervention is designed to make operationally feasible. Consistent with this view, point estimates of gap reductions tend to be larger in higher-barrier groups and for higher-cost practices, but these differences are generally not statistically distinguishable across groups; the main exceptions are a larger reduction for illiterate households in high-startup/high-transaction actions (significant at 10%) (Table A.3).

The next subsection examines implementation directly. In addition to the canonical actions used to construct the wedge, the coaching calls emphasized a broader set of managerial practices. I therefore study treatment effects on implementation indices defined over the full action set, and report action-level estimates in the Appendix.

5.2.2 Expansion of parental managerial practices

Panels B and C of Table II report effects on implementation indices. Panel B uses the four canonical actions that correspond one-to-one with the role-endorsement questions. Implementation increases in each category, with an overall rise of 0.09 relative to a control mean of 0.50. Panel C extends the analysis to the full set of eleven actions emphasized during coaching. The overall implementation index rises by 0.09 off a control mean of 0.54. Effects are positive across all three cost categories: 0.11 for conversational checks, 0.11 for home verification/enforcement, and 0.06 for high startup/transaction actions. The smaller level effect on high-cost actions is consistent with the presence of fixed and coordination costs that make these practices harder to adopt over a four-month horizon, even when coaching reduces those frictions.

Table A.5 and A.6 report action-level estimates and documents that the increases are not driven by a single item within any category. Table A.3 show that high-cost implementation gains are concentrated among households with higher baseline implementation barriers (e.g., low-literacy parents), while low-friction actions rise broadly across groups. These patterns mirror the baseline role–action wedge and are consistent with coaching operating through both belief updating about returns to managerial practices and reductions in the fixed and cognitive costs of sustained follow-through.

In Table A.7, I report treatment effects using outcomes measured independently in a student survey administered separately from the parent survey. Facilitators never interacted with students during implementation, so these measures provide an external check on parent self-reports. The estimated effects on student-reported parental practices are similar in magnitude—and if anything slightly larger in most cases—than the corresponding parent-reported estimates. To the extent that students were not directly exposed to the intervention, this pattern supports the interpretation that treated parents changed their managerial practices rather than merely changing how they reported them.

Chore reallocation and time protection. A core margin targeted by the coaching was time-allocation discipline: protecting study time by temporarily relaxing chore demands during high-stakes periods (e.g., before tests). Table III shows that treated households are more likely to report having excused the child from a usual chore in order to study: by 11 percentage points in parent reports and 14 percentage points in student reports. Relative to the control mean of 0.35, these correspond to increases of 31% and 40%, respectively. The student reports provide a useful external check on parent self-reports—especially for chore allocation—because the student survey was conducted independently, facilitators never interacted with students, and parents did not know in advance that students would be surveyed or that these items would be asked. These effects indicate that the intervention shifted not only monitoring and routine formation, but also intra-household allocation of the child's time toward academic effort. Consistent with within-household substitution, treated households report increased replacement of the child's chores by other household members (Table III).

Taken together with Table II, the evidence is consistent with the intervention increasing household organizational inputs in the home production of learning. Treated households report greater adoption of managerial practices (monitoring, routine formation, and school coordination) and a reallocation of time away from chores toward study.

5.2.3 Objective Evidence of Household Managerial Inputs: Absenteeism and Parental Information

The preceding section documents changes in self-reported household managerial practices. Because survey responses can be affected by differential reporting, I next examine two outcomes that provide more verifiable evidence of household organizational inputs: student absenteeism and parents' information about academic performance. These outcomes map directly to the conceptual framework. Absenteeism responds to time-allocation discipline and monitoring (e.g., ensuring the child leaves for school, enforcing routines, and addressing missed days). Parental information about performance can improve endogenously when households increase monitoring intensity and coordination with the school, even without receiving formal performance updates.

Absenteeism. Table IV shows that the intervention reduces absenteeism. Students in treated households are 6.8 percentage points less likely to miss school during the study period, an 11% decline relative to the control mean of 0.61. This effect is consistent with the intervention shifting routine formation and accountability around attendance, including occasional school contact. Heterogeneity patterns show no statistically significant differences across sub-groups.

Parental information about academic standing. The intervention did not provide parents with grades or administrative performance records. Any change in parental knowledge must therefore arise from households' own monitoring and information acquisition—through conversations with the child or contact with teachers and school staff. To measure parental information, I compare parents' reports of first-semester performance to administrative records. Because many parents in this setting (especially those with low literacy) do not track exact GPAs and instead reason about performance using salient thresholds (passing versus failing), the survey collected both a threshold-based measure and an exact-GPA report.

Table IV reports results for both measures. Among parents who report a GPA, the absolute deviation between the parent's reported GPA and the administrative GPA declines by 0.09 points (about 15% relative to the control mean of 0.61). Parents are also 6 percentage points more likely to correctly classify whether the child is above the passing threshold (10/20), a 7.5% increase over the control mean of 0.80. Response rates to these items do not differ by treatment status, mitigating concerns that differential item nonresponse drives the results. Because the intervention provided no performance information, the improvement in parental accuracy must reflect endogenous information acquisition through increased monitoring and school contact, consistent with the view that managerial practices generate information rather than merely act on it.

Taken together, these patterns provide objective evidence that the intervention increased household organizational inputs in ways that are observable outside self-reports: it reduced absences and improved parents' information about academic standing without directly supplying performance data. The next section examines whether these changes translate into improvements in achievement and progression outcomes.

5.2.4 Persistence of managerial practices a year after coaching ends

A natural concern is that endline changes reflect short-run reactivity to being contacted rather than durable changes in household decision rules. To assess persistence, I conducted a short follow-up survey a year after coaching ended, re-measuring a subset of the targeted practices.[9] Average effects attenuate in the full sample, but persistence is concentrated among households with higher baseline implementation constraints. Among illiterate parents, the intervention increases the follow-up practice index by 0.05 and raises high-startup/transaction practices by 0.08. Among literate parents, effects are close to zero. Differences between illiterate and literate households are statistically significant for high-cost practices and for the overall index (see Table VIII). At the item level, persistent gains among illiterate parents include increased school contact (visiting the school and checking attendance) and continued home monitoring (e.g., homework supervision and test-date checks) as shown in Table IX. This pattern is consistent with the intervention reducing startup and coordination costs in households where such costs are binding, rather than generating purely contemporaneous compliance. Section 5.5 shows that longer-run progression outcomes move in the same direction.

9. The subset includes: initiating conversations with the child; asking about test scores and upcoming test dates; supervising homework completion; checking on the child during lesson review; visiting the school; and making unannounced visits to verify attendance.

5.3 The Intervention Improves Student Performance

I evaluate impacts on two primary academic outcomes: grade completion and year-end GPA. Grade completion is an indicator equal to one if the student advances from 8^{th} to 9^{th} grade and zero otherwise. Advancement is determined by the year-end GPA, with a national threshold of 10 out of 20. Students below this threshold repeat the grade, making grade completion a high-stakes outcome and a natural measure of academic progress.

Year-end GPA is a continuous summary of achievement. It is computed as a weighted average of first-semester GPA (one-third) and second-semester GPA (two-thirds), following Benin's official grading rule. Each semester GPA aggregates subject-level performance, itself a weighted combination of standardized tests (two-thirds) and teacher-administered quizzes (one-third). This structure reflects both the central role of standardized assessments and teacher-level variation in continuous assessment.

To unpack where performance changes occur, I examine subject-level outcomes in two ways. First, I report pooled indices for STEM (Mathematics, Physics, Biology) and Humanities (Language, History–Geography). Second, I disaggregate by subject. All GPA and test-score outcomes are standardized within class using the control-group mean and standard deviation, which facilitates comparison across outcomes.

Table V, Column 1 (Figure VI) shows that the intervention increases grade completion by 4 percentage points, a 6% gain relative to the control mean of 0.68. Column 2 (Figure VII) shows a 0.11 standard deviation increase in year-end GPA. The GPA effect is concentrated in the second semester (0.12 SD), with smaller and statistically insignificant gains in the first semester (0.07 SD). This is consistent with the implementation schedule: calls began one week before the final test of the first semester, leaving limited scope to affect earlier assessments.

Columns 3 and 4 indicate that GPA gains are concentrated in STEM subjects. Table A.10 (Figure VIII) further disaggregates effects: the largest improvements occur in Physics (0.11 SD), followed by Biology (0.09 SD) and History–Geography (0.05 SD). Physics gains appear in both semesters, while other improvements emerge primarily in the second semester. These estimates are robust to alternative specifications and are not driven by selective test participation.[10]

These results are consistent with a mechanism in which coaching increases the household organizational inputs that raise the productivity of study time—routine formation, monitoring, and time-allocation discipline—rather than directly increasing subject-specific instruction. Randomization occurs within classrooms, holding fixed teachers, curricula, peers, and schedules, so the estimates isolate changes plausibly operating through the home production of learning rather than school-level inputs. Section 5.6 examines within-class spillovers.

5.3.1 The Intervention Benefits Academically Weaker Students the Most

A key prediction of the framework is that managerial inputs have the highest marginal returns when baseline study behavior is far from the threshold for promotion. I therefore examine heterogeneity by baseline performance, focusing on students most at risk of repetition.

I divide students into three performance groups based on scores in the first standardized test of the academic year: a bottom quartile (scoring below 10 out of 20 in at least four out of six subjects), a middle half, and a top quartile.[11] Table VI, Panel A shows large gains for the bottom quartile: grade completion increases by 12 percentage points

10. If the intervention increased test-taking among students who would otherwise skip assessments, estimated effects could be mechanically attenuated. I find no significant treatment effect on test participation, alleviating this concern.

11. Sixty percent of students in the bottom quartile have illiterate parents.

(a 40% increase relative to a control mean of 0.30) and year-end GPA rises by 0.16 SD. Effects for the middle and top groups are small and statistically indistinguishable from zero. For grade completion, I reject equality of treatment effects between the bottom quartile and each higher-performance group.

These findings match the threshold nature of grade progression: modest improvements in effective study time and accountability can generate large changes in promotion among students close to the passing cutoff, while leaving higher-performing students relatively unaffected.[12]

5.3.2 Heterogeneity by Parental Literacy and Student Gender

I next examine heterogeneity along two pre-specified dimensions: parental literacy and student gender. These dimensions proxy for household implementation constraints and for gendered differences in time demands.

Table A.11 shows that grade completion increases by 6 percentage points for students with illiterate parents and by 2 percentage points for students with literate parents; year-end GPA effects are 0.13 SD and 0.10 SD, respectively. Differences across literacy groups are not statistically significant, but the direction is consistent with the intervention's design: it targets literacy-independent organizational inputs rather than academic tutoring.

Turning to gender, Table A.12 reports effects separately for girls and boys. Grade completion increases by 5 percentage points for both groups. GPA gains are larger for girls (0.19 SD) than for boys (0.10 SD), though the difference is not statistically significant. This pattern is consistent with the possibility that time-allocation discipline (including

12. Results are robust to alternative baseline-performance partitions (e.g., deciles aggregated into three equal-sized groups).

chore reallocation) is especially consequential for girls; Section 5.2.2 documents corresponding shifts in household time allocation.

Finally, Table A.13 reports results by the interaction of parental literacy and student gender. Point estimates are largest for girls with illiterate parents (10 percentage points for grade completion), but these differences are not statistically precise and should be interpreted cautiously.

Overall, heterogeneity patterns are consistent with the view that coaching is most valuable where household organizational inputs are initially scarce—particularly among academically weaker students and, suggestively, among girls in low-literacy households.

5.4 Household managerial inputs do not crowd out financial expenditures on schooling

A natural concern is that increasing household organizational inputs may crowd out other uses of parental time—including income-generating activities—thereby tightening budget constraints and reducing educational expenditures. Conversely, the coaching could make schooling more salient and induce higher spending. In either case, the intervention could affect monetary investments such as tuition payments and private tutoring.

Table A.17 reports impacts on three measures of tuition payment behavior: total tuition paid, the probability of full payment, and the timing of full payment. The estimates are small and statistically indistinguishable from zero. The point estimate for total tuition paid is −147 CFA relative to a control mean of 14,590 CFA; the probability of full payment declines by 2 percentage points relative to a control mean of 0.69; and the timing of full payment shifts by 0.53 days. None of these effects is statistically significant and the magnitudes are economically modest.

I also examine expenditures on private tutoring, which represent a salient discretionary educational expenditure in this setting. I find no detectable effect on tutoring expenditures.

Taken together, these results suggest that the intervention increased household managerial practices without materially altering contemporaneous financial spending on schooling. This pattern is consistent with the interpretation that the intervention primarily relaxes informational and implementation constraints—helping households translate endorsed responsibility into routines—rather than requiring large reallocation of resources away from market work. The absence of detectable effects on tuition and tutoring also reduces concerns that improved academic outcomes are mechanically driven by increased spending rather than by changes in household organization.

These findings complement a literature emphasizing financial barriers to schooling by showing that, among enrolled students, low-cost changes in household organization can improve outcomes without increasing educational spending.

5.5 Dropout one year later

A common concern with low-cost behavioral interventions is that short-run effects reflect transitory reactivity rather than durable changes in household decision rules. A related concern in this setting is that improved progression could be achieved mechanically by pushing marginal students into the next grade, potentially increasing subsequent dropout if students are promoted without sustained changes in study inputs.

Table VII examines dropout one year after the intervention. In the full sample, the point estimate is small and statistically indistinguishable from zero (-0.8 percentage points). Effects are concentrated among students from illiterate households: dropout falls by 4.1 percentage points, a 34% reduction relative to the 0.12 control mean (Column 2). In contrast, the point estimate for students from literate households is positive and not statistically significant (Column 3). The difference between the two literacy-specific effects is statistically significant, indicating that the long-run gains are concentrated in households facing higher baseline implementation constraints.

Columns 4 and 5 show a similar pattern by student gender. Dropout declines by 3.6 percentage points for girls relative to a control mean of 0.30, while the effect for boys is small and not statistically significant; the difference in treatment effects between girls and boys is statistically significant. These gender-differentiated dropout effects need not imply stronger persistence in the specific managerial practices measured at follow-up: the follow-up survey was intentionally short and covers only a subset of the practices emphasized during coaching. In particular, it does not re-measure key time-allocation margins such as chore reallocation around exams, which may be especially relevant for girls and could plausibly mediate longer-run persistence in schooling.

Overall, Table VII suggests that the intervention does not generate the adverse dynamic implied by a purely mechanical "promotion push." Instead, the long-run dropout reductions are concentrated in disadvantaged groups, consistent with sustained changes in household organizational inputs, even if some persistent margins are not fully observed in the follow-up measurement.

5.6 No evidence of negative spillovers on untreated students

Because treatment and control students share classrooms—and households are embedded in the same local communities—interference is a natural concern. If the intervention induced any reallocation of scarce school inputs (e.g., teacher attention) toward treated students, untreated classmates could be harmed, mechanically inflating within-class treatment estimates. More generally, within-school spillovers could cause the within-class comparison to differ from the individual-level effect.

I assess this concern by exploiting the design's two school types: (i) intervention schools with individual-level randomization and (ii) "pure control" schools that were not exposed to the experiment. Comparing untreated students in intervention schools to students in pure control schools provides a direct check on whether untreated students are harmed, on net, by proximity to treated peers.

Table A.9 reports three comparisons for grade completion. First, treated students in intervention schools outperform students in pure control schools by 6.2 percentage points (Column 1), consistent with the main within-class estimate. Second, untreated students in intervention schools perform slightly better than students in pure control schools (2 percentage points; Column 2), and the estimate is not statistically distinguishable from zero. This pattern provides no evidence of negative spillovers large enough to account for the main results. Third, pooling treated and untreated students in intervention schools and comparing them to pure control schools yields a 4 percentage point gain (Column 3), similar to the within-class effect. Together, these comparisons suggest no displacement of untreated students; if anything, the point estimates are consistent with modest positive spillovers, which would imply that within-class estimates understate the total effect.

5.7 Robustness

This section reports a set of checks aimed at ruling out the most common threats to interpretation: differential attrition, selective enrollment into the study sample, small strata in literacy-stratified randomization, selective test participation, and sensitivity to the choice of controls. The results are summarized here; corresponding tables are in the Appendix.

Attrition: There are two relevant sources of attrition: missing administrative outcomes due to student transfers/dropout during the school year and non-response in the endline parent survey. Attrition in second-semester administrative records is low (2.24% of the baseline sample), and endline parent survey non-response is 8.26%. Treatment status does not predict attrition (p=0.13). A fully interacted specification that allows attrition to vary with baseline covariates also shows no evidence of differential attrition by treatment (joint p=0.85). To assess the worst-case implications of attrition, I report Lee bounds in Tables A.18–A.20; the conclusions are unchanged.

Sample selection into the study: Because enrollment was organized by administrative procedures rather than opt-in recruitment, selection into the study sample primarily reflects logistics. Within each intervention school, parents were contacted sequentially using an alphabetical ordering of students and outreach stopped once the target number of baseline interviews was reached. To assess external validity, Table A.21 compares outcomes for students whose parents were not reached before the target was met to outcomes for students in the experimental sample. Grade completion and related outcomes are similar across comparisons suggesting that the enrolled sample is not a highly selected subset of the eligible population.

Small strata and randomization inference: Randomization was stratified by parental literacy within class, which can generate small strata in some classrooms. Table A.22 reports randomization-inference p-values, confirming that the main results are unlikely to be artifacts of the particular treatment assignment.[13]

Selective test participation: Administrative test-score outcomes are missing for some students because not all students take all scheduled tests. Table A.23 re-estimates the main learning results in the subsample of students who took all scheduled tests in every subject. Estimates are similar, indicating that differential test participation is unlikely to drive the findings.

Sensitivity to controls: The main specifications use double post-LASSO to select controls from the baseline covariate set.[14] Table A.24 shows that results are similar under a standard ANCOVA specification without additional controls, indicating that the findings are not sensitive to the covariate selection procedure.

Overall, these checks suggest that the estimated effects are not driven by differential attrition, selective sampling, artifacts of small strata, selective test-taking, or the choice of controls.

6 A Highly Cost-Effective Intervention

6.1 Cost-effectiveness and fiscal implications

This section benchmarks the intervention's costs against short-run fiscal and household resource implications. The exercise is not a full welfare analysis: it abstracts from general equilibrium effects and from long-run earnings gains, and instead focuses on transparent, near-term margins that are tightly linked to measured outcomes. I report calculations for the subgroup most affected by the intervention—students at high risk of grade repetition—

13. See, e.g., Rosenbaum and Rosenbaum 2002 for the randomization-inference approach.
14. See Belloni, Chernozhukov, and Hansen 2014.

and treat private cost savings conservatively.

Implementation costs. The per-student implementation cost in the study is \$9.50, inclusive of facilitator time, training and supervision, and phone-related costs. In scaling calculations below, I assume costs scale proportionally with the number of students reached; if there are fixed costs or scale economies, this assumption is conservative for large-scale implementation.

Fiscal externalities from avoided repetition among at-risk students. Grade repetition is fiscally salient because it mechanically increases the number of student-years the public system must finance. I therefore translate the estimated effect on grade completion for at-risk students into a benchmark measure of resources freed within the public system. I define at-risk students as those scoring below 10/20 in at least four of six subjects on the baseline standardized assessment; in this subgroup, grade completion in the control group is 0.30 and the intervention increases grade completion by 12 percentage points.

Consider scaling the program to 1,000 at-risk students. The implied number of repetitions avoided is $1{,}000 \times 0.12 = 120$. The gross fiscal externality depends on the public cost of an additional lower-secondary student-year. UNESCO/World Bank data report government expenditure per secondary student as 11.3% of GDP per capita for Benin (2015). Using GDP per capita in 2015, this corresponds to approximately \$149 per secondary student-year in 2023 after adjusting for inflation. Under this benchmark, the public resources associated with 120 avoided repetitions are about \$17,880 ($=120\times\$149$).

Two caveats are important. First, this calculation uses an average spending benchmark; the short-run marginal fiscal savings may be smaller when inputs are fixed. In that case, the appropriate interpretation is that avoided repetitions free capacity (teacher time, classroom seats) rather than generating one-for-one cash savings. Second, the benchmark year reflects data availability; if one updates the calculation using more recent GDP per capita levels while holding the spending share fixed, the implied cost per student-year is higher, yielding larger fiscal externalities.

A useful way to present the policy implication is a break-even condition. With 120 repetitions avoided, the program cost of \$9,500 implies that the government breaks even whenever the marginal cost of an additional student-year exceeds \$79 (= \$9,500/120). Under the \$149 benchmark, the intervention is fiscally self-financing for the public sector even before accounting for any additional social benefits.

Household resource implications. Avoiding repetition also reduces out-of-pocket costs for households. To keep the calculation conservative, I include only annual school fees (excluding uniforms, materials, transport, and the opportunity cost of delayed labor-market entry). Public-school lower-secondary fees in Benin are 16,180 CFA (approximately \$30). Under this range, avoiding 120 repetitions implies household fee savings of roughly \$3,600. These figures understate total private savings because they omit other direct and indirect schooling costs.

Cost per unit of educational progress. A complementary benchmark reports cost per unit of progress. For at-risk students, the intervention generates 0.12 additional promotions per treated student, implying a cost of roughly \$79 per additional student promoted (= \$9.50/0.12). This statistic is transparent and does not require assumptions about public budgets; it can be compared directly to the fiscal benchmark above and to alternative policies aimed at reducing repetition.

Finally, long-run outcomes provide an additional perspective. One year after the intervention, dropout falls by 4.1 percentage points among students from illiterate households (Table VII). This pattern suggests that the intervention's benefits extend beyond short-run promotion and is consistent with durable changes in household organizational inputs.

6.2 Government revenue externalities: an accounting benchmark

As a complementary benchmark, I compute an accounting measure of the intervention's potential fiscal externality through higher lifetime tax revenues. This exercise is not a full welfare analysis: it abstracts from general equilibrium effects and from the broader social value of learning, and it relies on transparent assumptions about (i) the number of additional students who complete lower secondary due to the intervention, (ii) earnings associated with completion, and (iii) the fraction of earnings that generates tax revenue.

Setup. I scale the calculation to a cohort of 1,000 treated students, so program costs equal \$9,500 (=1,000×\$9.50). I consider two completion effects: 4 percentage points (the full-sample estimate) and 12 percentage points (the estimate for academically marginal students). These effects imply 40 and 120 additional lower-secondary completers, respectively.

I assume that individuals enter the labor market and begin generating tax revenue at age 18 and contribute through age 60. Earnings are held constant in real terms. I consider two earnings benchmarks: the statutory monthly minimum wage and the average monthly earnings of lower-secondary graduates in the EHCVM survey. I discount future revenues at 5%.

Tax capture and interpretation. Mapping earnings into fiscal revenue requires assumptions about effective taxation in an economy with high informality. I therefore parameterize tax capture as the share of earnings that generates income-tax revenue (either because work is formal or because earnings are effectively taxed). I report an upper bound (100% capture) and an intermediate benchmark (15% capture). The tax rate is set to 10%; this can be interpreted as an effective average tax take on earnings (direct and indirect), and results scale linearly with this parameter.

Results. Table X reports the present discounted value of lifetime tax receipts, net benefits, and payback horizons. Under the intermediate tax-capture benchmark (15%), the program yields positive net fiscal benefits for marginal students and pays back within 5–7 years of labor-market entry, depending on the earnings benchmark. For the full sample, the same benchmark implies longer payback horizons (roughly two to four decades), reflecting the smaller completion effect. Because the calculations are mechanical and depend on assumptions about tax capture and earnings, I emphasize the range across scenarios rather than a single point estimate.

Schooling-cost assumption. The benchmark table treats public schooling costs as unchanged, effectively assuming that the completion gains operate primarily through avoiding repetition rather than extending time in school beyond what would otherwise occur. An alternative scenario in which induced completion increases continuation beyond middle school would raise public costs but also increase private earnings and fiscal revenues.

6.3 Cost per unit of academic progress

A complementary benchmark reports achievement gains per dollar spent. Table A.25 reports standardized achievement gains per $100 in 2011 USD following Bhula et al. (2020). Under this harmonization, the intervention delivers 1.57 SD of year-end GPA per $100. I interpret these comparisons as descriptive; the main policy-relevant margin in this setting is progression (grade completion and dropout).

Taken together, the cost benchmarks suggest that the intervention delivers sizable improvements in educational progression at low program cost. In particular, the gains in grade completion and the reduction in repetition imply meaningful short-run fiscal externalities, and the tax-revenue accounting exercise indicates that plausible long-run revenue gains can be large relative to implementation costs under conservative tax-capture assumptions. These calculations are necessarily approximate, but they support the conclusion that light-touch coaching that relaxes household implementation constraints can be a fiscally and administratively feasible complement to more resource-intensive education policies in low-capacity settings.

7 Conclusion

This paper studies a constraint on human capital formation that is salient in low-income settings with low parental education: the underprovision of household organizational inputs that raise the productivity of children's study time. In rural Benin, schools provide limited individualized feedback and infrequent performance updates, while many parents lack the literacy needed for content-based help. In this environment, a high-leverage household margin is managerial rather than curricular—routine formation, monitoring, time-allocation discipline, and selective coordination with the school.

I evaluate a low-cost phone-based coaching intervention designed to relax two frictions that emerged from qualitative work and are visible in the baseline data: parents often undervalue non-curricular managerial practices, and they face high fixed and cognitive costs of translating endorsed responsibility into implementable routines. The intervention delivered on average ten structured calls over four months. Calls were interactive: facilitators helped parents choose feasible actions, anticipate constraints, and adapt routines through iterative troubleshooting.

The intervention increases student achievement and progression. Grade completion rises by 4 percentage points (6% relative to the control mean) and year-end GPA increases by 0.11 standard deviations, with gains concentrated in the second semester. Effects are substantially larger for academically marginal students at baseline: grade completion increases by 12 percentage points and GPA by 0.16 standard deviations. Subject-level estimates indicate that improvements are concentrated in STEM performance.

Mechanism evidence aligns with the conceptual framework. Using role-anchored measures, the intervention reduces the intention–implementation wedge, with larger baseline wedges and larger gains in higher-cost domains. Using a broader action set emphasized during coaching, treated households report increased adoption of managerial practices across low-friction checks, home verification/enforcement, and higher-cost routines and school coordination. Independent of self-reports, the intervention reduces student absenteeism and improves parents' information about academic standing despite providing no grades directly, consistent with increased endogenous monitoring and school contact. The intervention also shifts intra-household time allocation by increasing the likelihood that children are temporarily excused from routine chores to protect study time.

Longer-run outcomes suggest that the intervention's benefits are not purely transitory.

One year after the intervention, dropout falls for disadvantaged students, with effects concentrated among children of illiterate parents and among girls. These patterns are consistent with the persistence evidence showing sustained changes in higher-cost practices among low-literacy households.

Finally, cost benchmarks indicate that the intervention delivers sizable progression gains at low program cost. Because grade repetition is fiscally salient, avoided repetition generates meaningful public resource implications, and an accounting exercise suggests that plausible long-run fiscal externalities through higher tax revenues can be large relative to implementation costs under conservative tax-capture assumptions. These calculations are necessarily approximate, but they underscore that coaching designed to reduce household implementation constraints can be a fiscally and administratively feasible complement to more resource-intensive education policies.

The broader implication is that models and policies focused exclusively on school inputs or household financial constraints miss an important margin: household organization of children's time and effort. In settings where learning depends heavily on study outside school, interventions that provide a low-literacy-appropriate managerial technology—and that help households adapt it to their constraints—can generate meaningful and durable improvements in educational trajectories.

References

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2012. "What did you do all day?: maternal education and child outcomes." *Journal of Human Resources* 47 (4): 873–912.

———. 2017. "Report cards: The impact of providing school and child test scores on educational markets." *American Economic Review* 107 (6): 1535–1563.

Baird, Sarah, Craig McIntosh, and Berk Ozler. 2010. "Cash or condition? Evidence from a cash transfer experiment." *Evidence from a Cash Transfer Experiment (March 1, 2010). World Bank Policy Research Working Paper,* no. 5259.

Barrera-Osorio, Felipe, Kathryn Gonzalez, Francisco Lagos, and David J Deming. 2020. "Providing performance information in education: An experimental evaluation in Colombia." *Journal of Public Economics* 186:104185.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *Review of Economic Studies* 81 (2): 608–650.

Bergman, Peter. 2021. "Parent-child information frictions and human capital investment: Evidence from a field experiment." *Journal of political economy* 129 (1): 286–322.

Berlinski, Samuel, Matias Busso, Taryn Dinkelman, and Claudia Martınez. 2021. *Reducing parent-school information gaps and improving education outcomes: Evidence from high-frequency text messages.* Technical report. National Bureau of Economic Research.

Bhula, Rachel, Michael Mahoney, and Katherine Murphy. 2020. *Conducting Cost-Effectiveness Analysis [Set of Online Resources].* Accessed March 25, 2021. https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The quarterly journal of Economics* 122 (4): 1351–1408.

———. 2010. "Why do management practices differ across firms and countries?" *Journal of economic perspectives* 24 (1): 203–224.

De Walque, Damien, and Christine Valente. 2023. "Incentivizing school attendance in the presence of parent-child information frictions." *American Economic Journal: Economic Policy* 15 (3): 256–285.

Dizon-Ross, Rebecca. 2019. "Parents' beliefs about their children's academic ability: Implications for educational investments." *American Economic Review* 109 (8): 2728–2765.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American economic review* 101 (5): 1739–1774.

Filmer, Deon, and Lant Pritchett. 1999. "The effect of household wealth on educational attainment: evidence from 35 countries." *Population and development review* 25 (1): 85–120.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher incentives." *American Economic Journal: Applied Economics* 2 (3): 205–227.

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

Hastings, Justine S, and Jeffrey M Weinstein. 2008. "Information, school choice, and academic achievement: Evidence from two experiments." *The Quarterly journal of economics* 123 (4): 1373–1414.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to learn." *The Review of Economics and statistics* 91 (3): 437–456.

Nguyen, Trang. 2008. "Information, role models and perceived returns to education: Experimental evidence from Madagascar." *Unpublished manuscript* 6.

Rogers, Todd, and Avi Feller. 2018. "Reducing student absences at scale by targeting parents' misbeliefs." *Nature Human Behaviour* 2 (5): 335–342.

Rosenbaum, Paul R, and Paul R Rosenbaum. 2002. *Overt bias in observational studies.* Springer.

York, Benjamin N, Susanna Loeb, and Christopher Doss. 2019. "One step at a time: The effects of an early literacy text-messaging program for parents of preschoolers." *Journal of Human Resources* 54 (3): 537–566.

8 Tables

Table I: Baseline Summary Statistics

	Control	Treatment	Control-Treatment	
	Mean	Mean	Mean	p-val T=C
Panel A: Respondent Background				
Father or Mother	0.81	0.82	-0.01	0.42
Age	42.38	43.04	-0.65	0.11
Literate	0.49	0.49	0.00	0.89
Understand report cards	0.56	0.56	0.00	0.91
Panel B: Student Information				
Age	14.15	14.07	0.08	0.40
Male	0.55	0.52	0.03	0.26
Attended kindergarten	0.29	0.30	-0.01	0.94
Ever repeated	0.58	0.61	-0.03	0.14
Panel C: Student Performance				
Physics	10.49	10.33	0.16	0.57
Biology	10.06	10.16	-0.10	0.78
Math	8.00	8.35	-0.35	0.02**
Language	10.34	10.44	-0.09	0.35
History-Geography	12.14	12.05	0.08	0.50
Panel D: Parents Knowledge & Actions at School				
Parents know number of subjects	0.34	0.32	0.02	0.23
Parents know number of tests	0.49	0.48	0.02	0.33
Parents know when 1st semester ends	0.33	0.32	0.01	0.79
Parents know passing grade	0.70	0.68	0.02	0.48
Parents received two report cards	0.47	0.46	0.01	0.71
Parents visited school	0.37	0.36	0.00	0.92
Parents checked attendance	0.11	0.10	0.01	0.50
Parents talked to teachers	0.18	0.17	0.01	0.80
Panel E: Parents Actions at Home				
Parents talked to child	0.79	0.79	0.00	0.92
Parents supervised homework	0.49	0.50	-0.01	0.74
Parents asked test scores	0.64	0.61	0.03	0.14
Parents asked test dates	0.46	0.45	0.01	0.81
Parents checked on child during reviews	0.53	0.52	0.01	0.54
Panel F: Parents' Baseline Belief				
Parent is correct about student's passing/failing	0.73	0.71	0.03	0.052*
Joint F-test test p-value				0.22
Students	1056	1038		

Notes: Data source is a baseline survey and administrative data from schools. "Parent Is Correct about Student's Passing/Failing" is a dummy variable that takes 1 if the parent was able to accurately tell whether the child is below or above the passing threshold of 10. The p-values are derived from regressions of each variable on a treatment dummy, with strata fixed effects included and I use robust standard errors. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table II: Effect of the Intervention on Parental Actions

	(1)	(2)	(3)	(4)
	Conversational (Percentage points)	Verificational (Percentage points)	High Cost (Percentage points)	Index (Percentage points)
Panel A: Gap Reduction (4 Actions)				
Treatment	-0.093***	-0.109***	-0.056***	-0.079***
	(0.016)	(0.021)	(0.014)	(0.011)
Control Mean	0.20	0.49	0.57	0.46
Observations	1,921	1,918	1,921	1,921
Panel B: Action Implementation (4 Actions)				
Treatment	0.110***	0.115***	0.058***	0.087***
	(0.017)	(0.022)	(0.014)	(0.011)
Control Mean	0.77	0.46	0.37	0.50
Observations	1,921	1,921	1,921	1,921
Panel C: Action Implementation (Full Set of 10 Actions)				
Treatment	0.113***	0.105***	0.063***	0.091***
	(0.012)	(0.015)	(0.013)	(0.009)
Control Mean	0.76	0.59	0.34	0.54
Observations	1,921	1,921	1,921	1,921

Notes: This table reports the effect of the intervention on parental actions. Columns (1)–(3) correspond to conversational, verificational, and high-cost parental actions. Column (4) reports equally weighted averages, so coefficients can be interpreted as percentage-point changes. Panel A reports effects on the reduction of the action gap for the four canonical actions. Panel B reports implementation effects for the same four canonical actions. Panel C reports implementation effects using the full set of eleven actions. Standard errors are in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table III: Allocation of Household Chores

	(1)	(2)	(3)
	Students Doing Fewer Chores (Percentage points)	Parents Doing More Chores (Percentage points)	Siblings Doing More Chores (Percentage points)
Panel A: Parents Survey			
Treatment	0.11***	0.08***	0.04**
	(0.02)	(0.02)	(0.02)
Control Mean	0.35	0.26	0.14
Observations	1,940	1,940	1,940
Panel B: Students Survey			
Treatment	0.14***	0.07***	0.08***
	(0.02)	(0.02)	(0.02)
Control Mean	0.37	0.21	0.19
Observations	1,902	1,902	1,902

Notes: This table presents the impact of the intervention on household chore allocation. Column (1) is an indicator equal to one if the student was relieved from chores at least once to allow more time to study. Column (2) is an indicator equal to one if parents reported doing more chores themselves, and Column (3) is an indicator equal to one if siblings took on additional chores. Panel A uses parent-reported data, and Panel B uses student-reported data. Standard errors are in parentheses and are clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table IV: Absenteeism and Parental Knowledge (Full Sample)

	(1)	(2)	(3)
	Absent at least once (Indicator)	Parental error about student performance (GPA points)	Parent correct about passing/ failing (Indicator)
Treatment	-0.068***	-0.09**	0.06***
	(0.019)	(0.04)	(0.02)
Control Mean	0.61	0.61	0.80
Observations	2,085	903	1,921

Notes: Column (1) reports the effect of the intervention on student absenteeism, defined as an indicator equal to 1 if the student was absent at least once during the intervention period. Columns (2) and (3) report effects on parental knowledge of the child's academic performance during the first semester of the 2022–2023 school year. Column (2) measures the absolute gap between the parent's reported GPA and the child's actual GPA. Column (3) is an indicator equal to 1 if the parent correctly identified whether the child scored above or below the passing threshold (GPA ≥ 10). All regressions include the baseline value of the outcome when available and incorporate the same set of controls as in the original tables (including double post-LASSO where applicable). Standard errors are clustered at the individual level and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table V: Effect of the Intervention on Student Performance

	(1)	(2)	(3)	(4)
	Grade Completion (Indicator)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Pooled				
Treatment	0.04**	0.11**	0.07**	0.03
	(0.02)	(0.05)	(0.05)	(0.03)
Control Mean	0.68	0.00	0.01	0.00
Observations	2,056	2,056	6,174	6,180
Panel B: 2nd Semester				
Treatment	-	0.12***	0.06**	0.05*
	-	(0.05)	(0.03)	(0.03)
Control Mean	-	0.00	0.00	0.00
Observations	-	2,056	4,091	4,094
Panel C: 1st Semester				
Treatment	-	0.07	0.09**	0.00
	-	(0.05)	(0.04)	(0.03)
Control Mean	-	0.01	0.01	0.01
Observations	-	2,094	2,083	2,086

Notes: This table reports the effect of the intervention on student outcomes. Grade Completion (Col. 1) is an indicator equal to one if the student passed grade 8, defined as a year-end GPA of at least 10 out of 20. GPA (Col. 2) measures standardized year-end and semester GPA. STEM (Col. 3) and Humanities (Col. 4) are standardized averages of subject scores: STEM combines Physics, Biology, and Math; Humanities combines Language and History-Geography. Panel A pools both semesters. Panel B covers the second semester, and Panel C covers the first semester. All regressions except Col. (1) control for the baseline outcome. Standard errors are in parentheses and are clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table VI: Effect on Different Quartiles of Baseline Performance

	(1)	(2)	(3)	(4)
	Grade Completion (Indicator)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Bottom Quartile				
Treatment	0.12**	0.16*	0.09	0.11**
	(0.04)	(0.08)	(0.06)	(0.06)
Control Mean	0.30	-0.74	-0.80	-0.85
Observations	540	540	1,608	1,614
Panel B: 2nd and 3rd Quartile				
Treatment	0.02	0.03	0.07**	0.01
	(0.03)	(0.05)	(0.04)	(0.04)
Control Mean	0.76	-0.02	0.01	0.04
Observations	1,102	1,102	3,317	3,317
Panel C: Top Quartile				
Treatment	0.01	0.12	0.08	0.04
	(0.02)	(0.11)	(0.07)	(0.07)
Control Mean	0.98	1.04	1.10	1.06
Observations	414	414	1,249	1,249
P-value Bottom = Middle	0.027	0.214	0.766	0.181
P-value Bottom = Top	0.004	0.675	0.768	0.378

Notes: This table shows treatment effects by baseline performance quartile. Baseline performance is defined using scores from the first standardized test of the academic year. Panel A reports effects for students in the bottom quartile, Panel B for the middle 50 percent, and Panel C for the top quartile. Grade Completion (Col. 1) is an indicator for passing grade 8, defined as a year-end GPA of at least 10 out of 20. GPA (Col. 2) is standardized year-end GPA. STEM (Col. 3) averages standardized scores in Physics, Biology, and Math. Humanities (Col. 4) averages standardized scores in Language and History-Geography. Regressions in Cols. (2)–(4) control for the baseline outcome. Standard errors are in parentheses and are clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table VII: Dropout Rate One Year after the Intervention

	(1)	(2)	(3)	(4)	(5)
	Full Sample (Percentage points)	Illiterate Parents (Percentage points)	Literate Parents (Percentage points)	Female Students (Percentage points)	Male Students (Percentage points)
Treatment	-0.008	-0.041**	0.027	-0.036**	0.019
	(0.012)	(0.018)	(0.016)	(0.018)	(0.018)
Control Mean	0.19	0.12	0.24	0.30	0.09
Observations	2,068	1,058	1,010	967	1,101
p-value		0.006		0.028	

Notes: This table reports the effect of the intervention on student dropout one year after the intervention. Dropout is an indicator equal to one if the student left school and zero otherwise. Column (1) shows results for the full sample. Columns (2) and (3) report results by parental literacy, and Columns (4) and (5) by student gender. All regressions use a double post-LASSO procedure to select controls from the full set of baseline variables in Table 1. Standard errors are reported in parentheses and are clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table VIII: Effect on Parental Action Categories by Literacy (Follow-up)

	(1)	(2)	(3)	(4)
	Conversational (Percentage points)	Verificational (Percentage points)	High Cost (Percentage points)	Index (Percentage points)
Panel A: Illiterate Parents				
Treatment	0.04	0.04*	0.08***	0.05***
	(0.021)	(0.025)	(0.027)	(0.018)
Control Mean	0.76	0.56	0.42	0.61
Observations	908	908	908	908
Panel B: Literate Parents				
Treatment	0.008	-0.007	-0.022	-0.005
	(0.017)	(0.024)	(0.026)	(0.016)
Control Mean	0.87	0.73	0.52	0.73
Observations	875	875	875	875
p-value	0.35	0.14	0.006	0.021

Notes: This table reports the effect of the intervention on aggregated categories of parental actions separately for illiterate and literate parents. The action categories are constructed using the list of specific actions reported in Table IX measured during the follow-up survey a year after the intervention ended. Conversational actions correspond to parent–child communication activities; verificational actions capture monitoring of academic information; and high-cost actions include activities that require greater time or effort, such as school visits or direct supervision of homework. Column (4) reports an average across all actions. The P-value row reports the p-value for the equality of treatment effects between literate and illiterate parents. Standard errors are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table IX: Effect of the Intervention on Specific Parental Actions (Follow-up)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Talked to Child (Percentage points)	Asked Test Scores (Percentage points)	Asked Test Dates (Percentage points)	Supervised Homework (Percentage points)	Checked on the Child During Lessons' Review (Percentage points)	Visited School (Percentage points)	Checked Attendance (Percentage points)
Panel A: Illiterate Parents							
Treatment	0.008	0.035	0.053*	0.059**	0.028	0.103***	0.060**
	(0.022)	(0.028)	(0.030)	(0.030)	(0.030)	(0.032)	(0.030)
Control Mean	0.87	0.75	0.66	0.48	0.90	0.50	0.34
Observations	908	908	908	908	908	908	908
Panel B: Literate Parents							
Treatment	0.034*	-0.028	0.019	-0.019	0.004	0.002	-0.046
	(0.020)	(0.020)	(0.025)	(0.029)	(0.027)	(0.031)	(0.030)
Control Mean	0.89	0.92	0.80	0.70	0.88	0.64	0.41
Observations	875	875	875	875	875	874	874
p-value	0.386	0.048	0.343	0.058	0.104	0.023	0.015

Notes: This table reports the effect of the intervention on specific parental actions separately for illiterate and literate parents. Each column corresponds to a different parental action measured during the follow-up survey a year after the intervention ended. The P-value row reports the p-value for the equality of treatment effects between literate and illiterate parents. Standard errors are reported in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

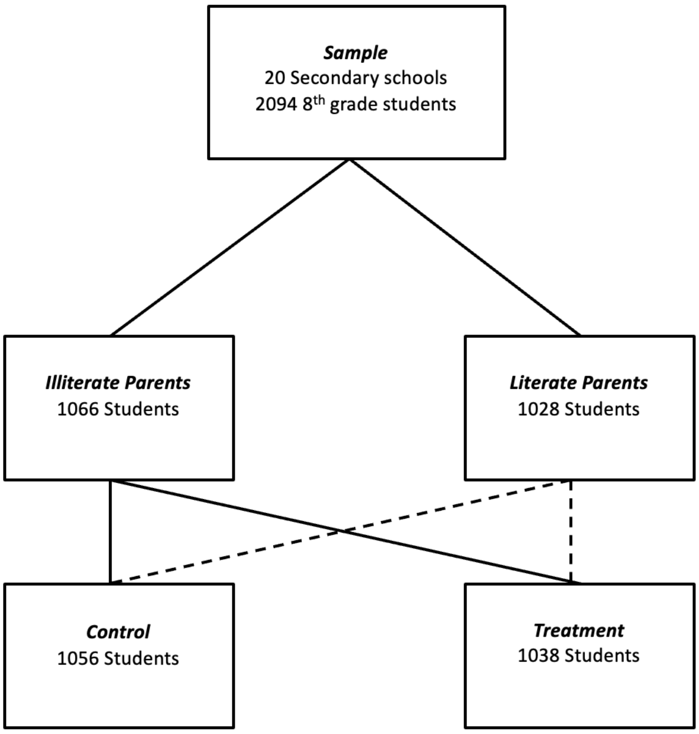
Table X: Cost Benefit Analysis Accounting for Tax Revenues

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: All (4 percentage points)				**Panel B: Marginal Students (12 percentage points)**			
	100% Formal Sector		15% Formal Sector		100% Formal Sector		15% Formal Sector	
Parameters								
Monthly Salary	minimum wage	average	minimum wage	average	minimum wage	average	minimum wage	average
Amount (USD)	83	111	83	111	83	111	83	111
Discount rate (%)	5	5	5	5	5	5	5	5
Start paying taxes after age	18	18	18	18	18	18	18	18
Life expectancy	60	60	60	60	60	60	60	60
Income tax rate (%)	10	10	10	10	10	10	10	10
Costs								
Intervention costs	9500	9500	9500	9500	9500	9500	9500	9500
Teachers salaries	N	N	N	N	N	N	N	N
Tuition fees expenses for parents	N	N	N	N	N	N	N	N
Benefits								
Collected between years	2027-2069				2027-2069			
Tax receipts	68361	91423	10254	13713	187994	251414	28199	37712
Net Benefit (Benefits-Costs)	58861	81923	754	4213	178494	241914	18699	28212
Breakeven year	3	2	40	23	1	1	7	5
Internal Rate of Return (%)	620	862	8	44	1878	2546	196	297

Notes: All monetary values are in 2023 U.S. dollars. The table reports an accounting benchmark for lifetime tax-revenue externalities generated by higher lower-secondary completion. The calculations are scaled to 1,000 treated students, so program costs equal $9,500 (=1,000×$9.50). Panel A uses the full-sample grade-completion effect (4 percentage points); Panel B uses the effect for academically marginal students (12 percentage points). Tax revenues are computed from age 18 through age 60 and discounted at 5%. Monthly earnings are set either to the statutory minimum wage or to average earnings for lower-secondary graduates in the EHCVM survey; amounts are expressed in 2023 USD. To reflect limited income-tax collection in a high-informality setting, the table reports two tax-capture scenarios: (i) 100% of earnings generate income-tax revenue (upper bound), and (ii) 15% of earnings generate income-tax revenue (intermediate benchmark). The income tax rate is set to 10% and can be interpreted as an effective average tax take; all tax-receipt figures scale linearly in this parameter. The breakeven year is the first year (after labor-market entry) in which the present discounted value of tax receipts exceeds program costs. The internal rate of return (IRR) is the discount rate that equates the PDV of tax receipts and costs.

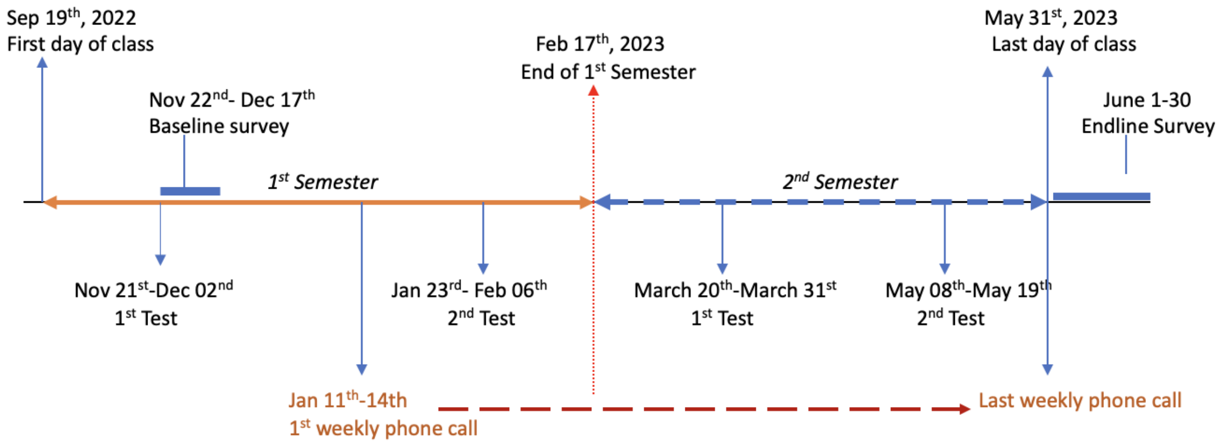
9 Figures

Figure I: Sampling



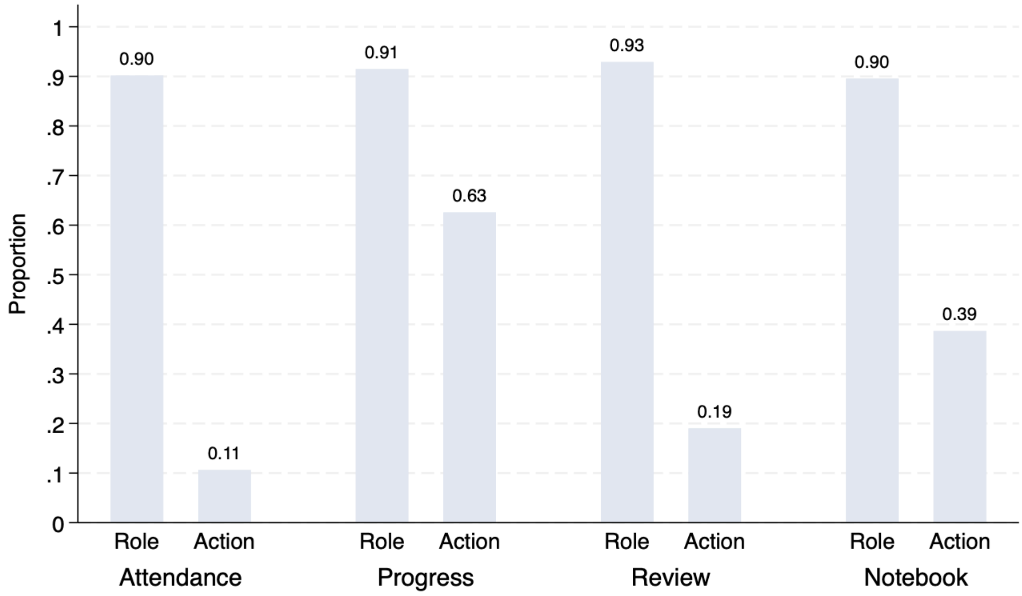
Notes: This figure depicts the sampling procedure in the intervention schools. In each of the 20 schools, there are, on average, four 8^{th} grade classes. In each class, I create literacy strata. Within each stratum, I randomly assign half of the students into treatment and the other half into control.
Alt text: Flow diagram showing schools, classes, literacy strata, and random assignment of students into treatment and control groups.

Figure II: Timeline of Activities



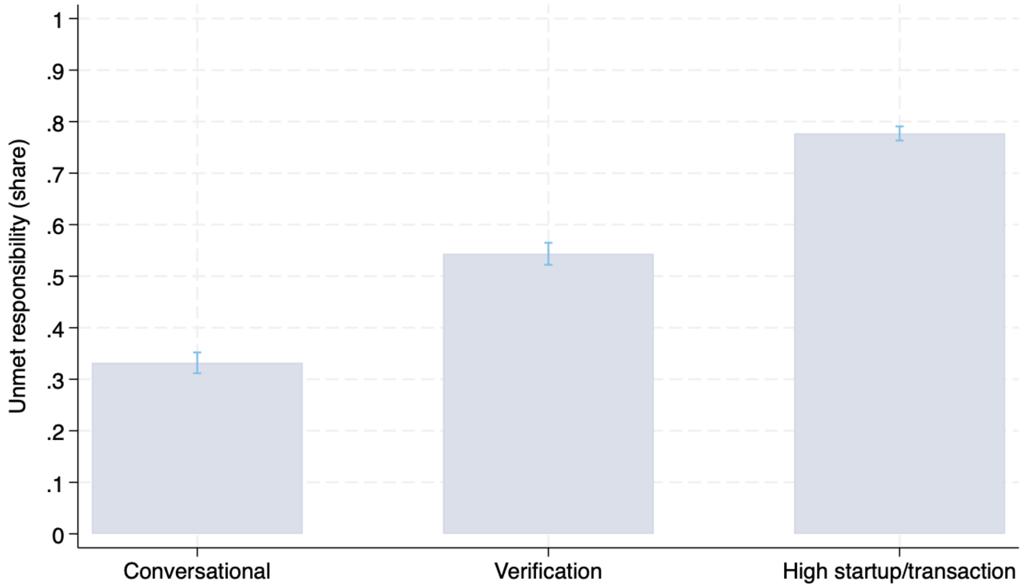
Notes: This figure shows the timeline of activities. At each test date, students take a test in all six subjects. Phone calls started on January 11^{th} and ended on May 20^{th}. Parents received on average 10 phone calls.
Alt text: Horizontal timeline marking test dates during the school year and the interval in which phone calls were delivered to parents.

Figure III: Role Endorsement vs Implementation of Parental Managerial Practices



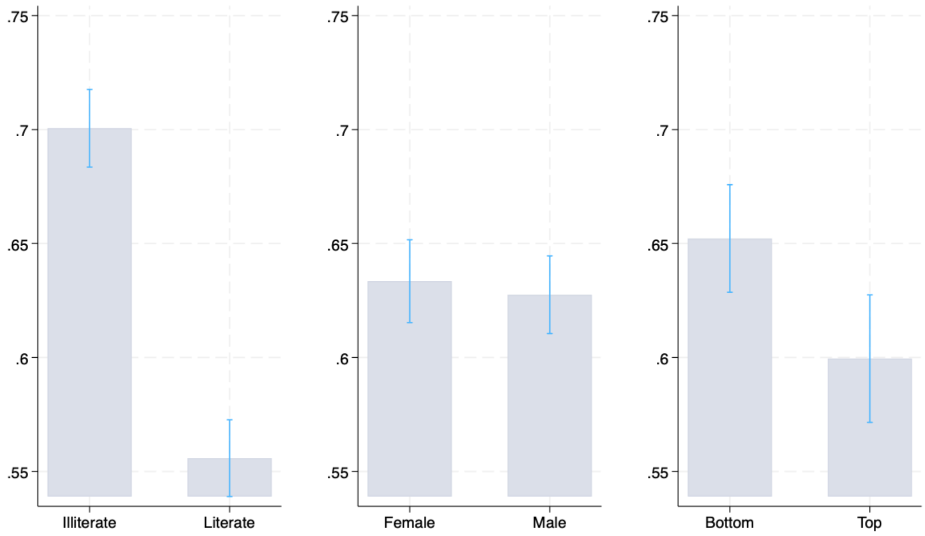
Notes: This figure compares baseline *role endorsement* (whether parents report that the domain is their responsibility) with *implementation* (whether the household reports taking a corresponding action). The four domains are attendance, progress monitoring, lesson review, and notebook checks. Role endorsement is near universal (0.90–0.93), while implementation varies substantially and is lowest for actions with higher transaction or start-up costs, such as school visits and structured study schedules.
Alt text: Four grouped bars show high values for responsibility across all domains and lower values for reported actions, with the largest gaps for school visits and structured study schedules.

Figure IV: Unmet-Responsibility Wedge by Implementation-Cost Category



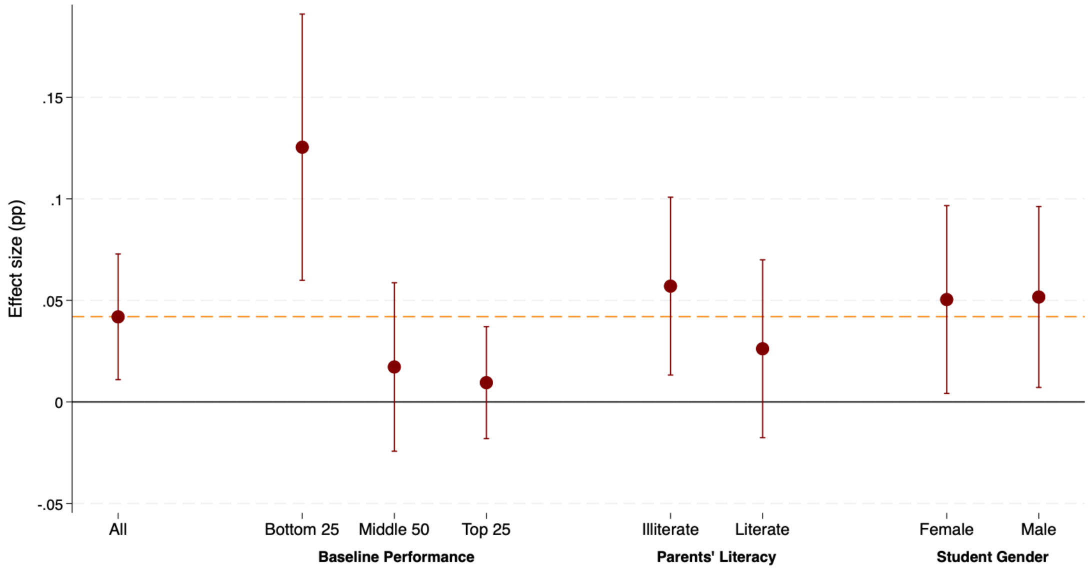
Notes: The figure reports baseline means (with 95% confidence intervals) of the unmet-responsibility wedge, aggregated into three implementation-cost categories. The wedge is constructed from the four baseline role–action pairs and equals one in a domain when parental responsibility is endorsed but the corresponding action is not reported, $g_{id} = \mathbf{1}\{r_{id} = 1,\ p_{id} = 0\}$. Category means average g_{id} across the relevant canonical domains.
Alt text: Three point estimates with confidence intervals comparing average unmet-responsibility gaps across low-, medium-, and high-cost implementation categories.

Figure V: Unmet-Responsibility Wedge by Subgroup



Notes: This figure reports the average unmet-responsibility wedge in parental managerial practices across subgroups defined by parental literacy, the gender of the focal student, and baseline student performance (bottom vs. top quartile). The wedge is constructed at the household–domain level as $g_{id} = \mathbf{1}\{r_{id} = 1,\ p_{id} = 0\}$, where r_{id} indicates that the respondent assigns *primary responsibility* to parents in domain d and p_{id} indicates that the household reports taking a concrete action in that domain. Thus $g_{id} = 1$ when parents are stated to be responsible but the corresponding action is not reported, and $g_{id} = 0$ otherwise (including when both are one or both are zero). The figure plots $G_i^{\text{unmet}} = \frac{1}{|D|}\sum_{d\in D} g_{id}$, the share of domains (attendance, academic progress, lesson review, and keeping notebooks up to date) in which responsibility is endorsed but not implemented.
Alt text: Point estimates with confidence intervals comparing average unmet-responsibility gaps across literacy, gender, and baseline performance subgroups.

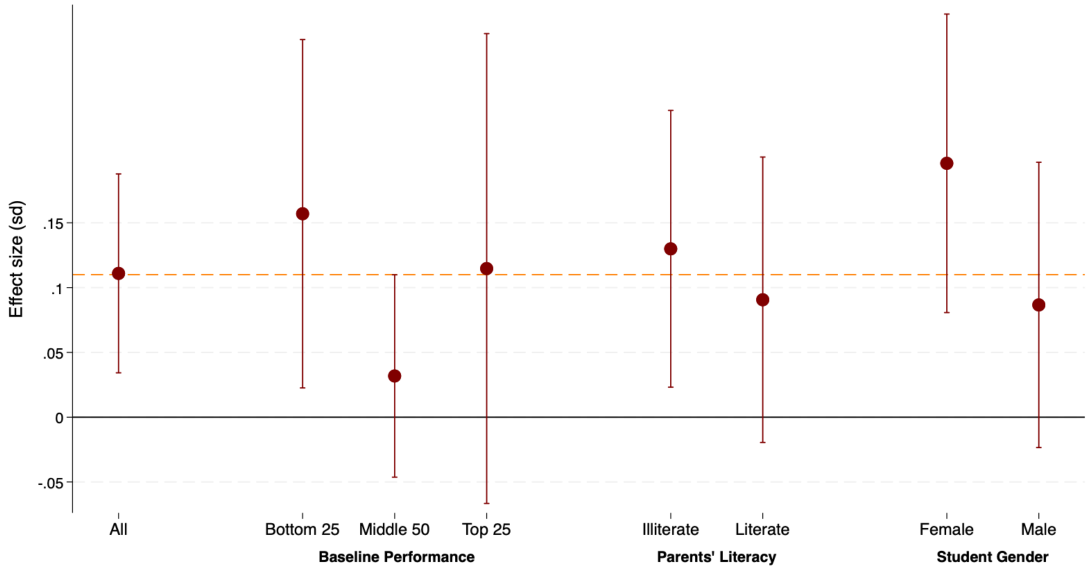
Figure VI: Effect on Grade Completion



Notes: This graph displays point estimates and 90% confidence intervals of the effect of the intervention on grade completion (passing from grade 8 to grade 9). A grade is completed if the year-end GPA is at least 10 out of 20. I measure the effect on the full sample and then investigate heterogeneity across baseline performance, parents' literacy, and student gender.

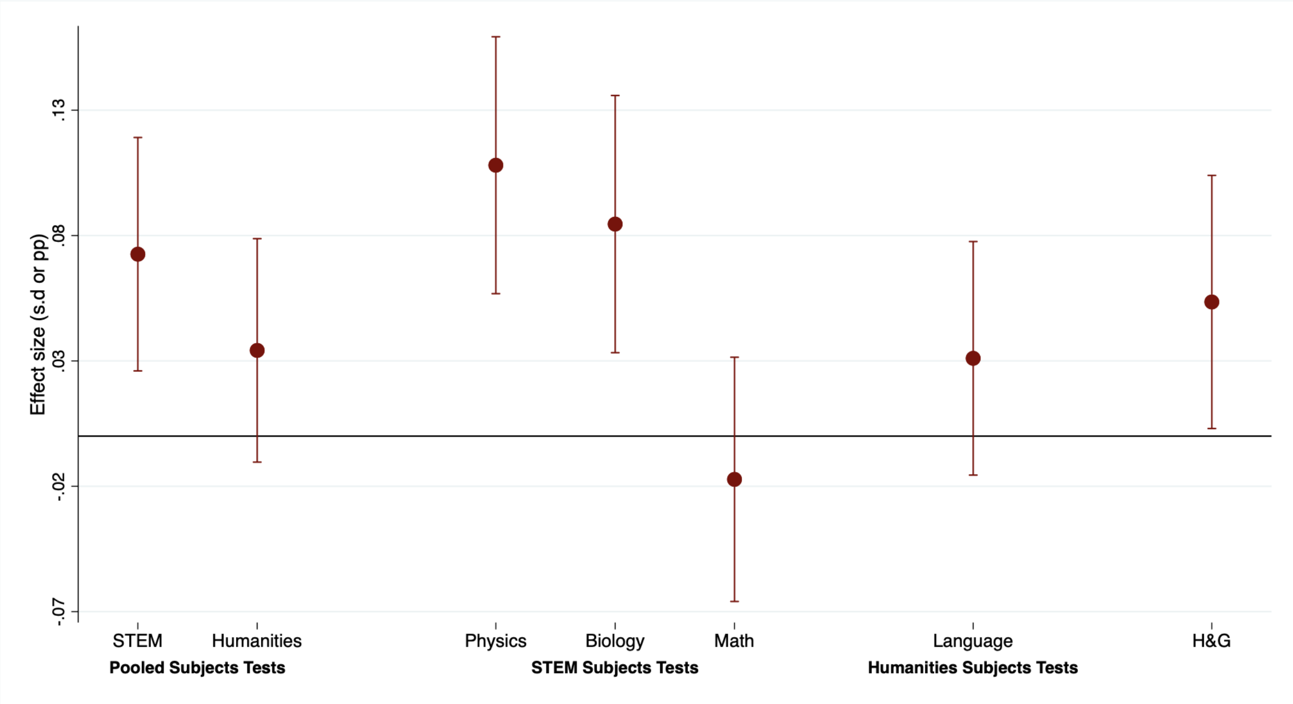
Alt text: Point estimates with confidence intervals comparing intervention effects on grade completion across the full sample and several subgroups.

Figure VII: Effect on year-end GPA



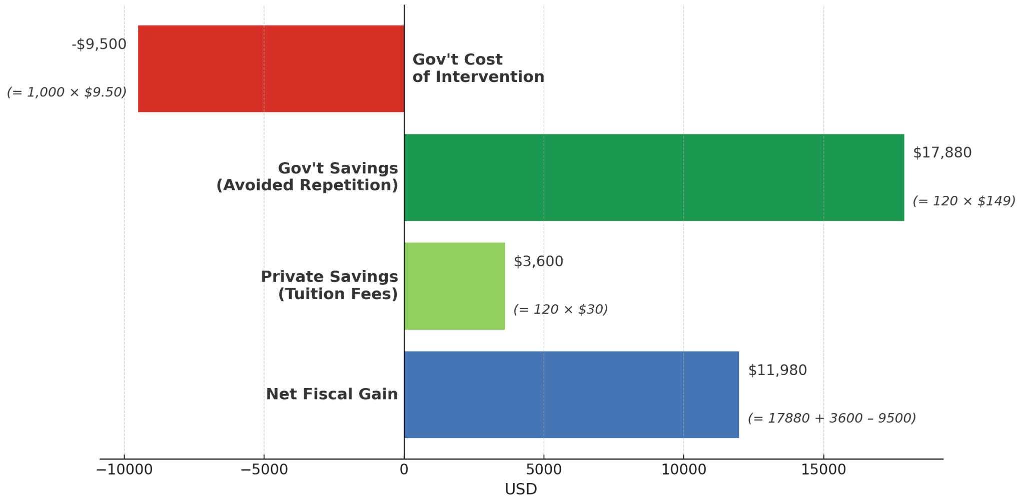
Notes: This graph displays point estimates and 90% confidence intervals of the effect of the intervention on students' standardized year-end GPA. I measure the effect on the full sample and then investigate heterogeneity across baseline performance, parents' literacy, and student gender.
Alt text: Point estimates with confidence intervals comparing intervention effects on standardized GPA across the full sample and several subgroups.

Figure VIII: Effect on Student Test Scores



Notes: This graph displays point estimates and 90% confidence intervals of the effect of the intervention on students' standardized test scores. H&G refers to History and Geography. Physics, Biology, and Math are grouped into STEM. Language and History-Geography are grouped into Humanities. All test scores are standardized. Standard errors are clustered at the individual level.
Alt text: Point estimates with confidence intervals comparing intervention effects on standardized test scores across subjects and subject groups.

Figure IX: One-Year Cost-Benefit Breakdown per 1,000 at-risk Students



Notes: This figure presents the cost of delivering the intervention to 1,000 at-risk students and compares it to the private and public savings from avoided grade repetition. To be conservative, private savings are restricted to tuition fees only, excluding additional expenses such as books, stipends, and foregone earnings from delayed entry into the labor market.
Alt text: Bar chart comparing program cost with private and public savings from avoided grade repetition.

A Appendix Tables

Table A.1: Household Characteristics

Variables	Mean	Standard Deviation
Household head is father	0.73	0.44
Household size	6.02	2.24
Mother years of schooling	2.29	3.57
Father years of schooling	4.9	4.75
Survey respondent is literate	0.49	0.5
Number of household members with some secondary education	1.56	1.36
Local language is spoken at home	0.98	0.13
Household has electricity(including solar)	0.76	0.43
Household has a mobile phone	0.97	0.16
Household has a television or radio	0.75	0.43
Household has a study table	0.55	0.5
Number of students	2094	

Notes: Data source is the baseline survey. I measured literacy by administering a literacy test to respondents during the baseline survey. The test was similar to that of DHS.

Table A.2: Effect on the Unmet Action Gap by Baseline Quartile (Endline)

	(1)	(2)	(3)	(4)
	Conversational (Percentage points)	Verificational (Percentage points)	High Cost (Percentage points)	Index (Percentage points)
Panel A: Bottom Quartile				
Treatment	-0.11***	-0.11**	-0.05	-0.09***
	(0.042)	(0.051)	(0.033)	(0.028)
Control Mean	0.23	0.51	0.62	0.50
Observations	497	496	497	497
Panel B: Middle 50				
Treatment	-0.10***	-0.12***	-0.07***	-0.09***
	(0.022)	(0.030)	(0.019)	(0.015)
Control Mean	0.21	0.49	0.57	0.46
Observations	1,031	1,029	1,031	1,031
Panel C: Top Quartile				
Treatment	-0.06	-0.07	-0.03	-0.05**
	(0.040)	(0.053)	(0.037)	(0.025)
Control Mean	0.15	0.45	0.52	0.41
Observations	393	393	393	393
p-value (Bottom = Middle)	0.65	0.87	0.70	0.99
p-value (Bottom = Top)	0.22	0.25	0.33	0.06

Notes: This table reports the effect of the intervention on the unmet gap between parents' stated responsibilities and implemented actions across baseline quartiles of the outcome distribution. Columns correspond to categories of parental actions: conversational, verificational, high-cost actions, and an overall index which is the average across all categories. Panels report treatment effects separately for the bottom quartile, the middle 50 percent, and the top quartile of the baseline distribution. The bottom rows report p-values for equality of treatment effects across quartiles. Standard errors are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.3: Effect on the Unmet Action Gap by Parental Literacy (Endline)

	(1)	(2)	(3)	(4)
	Conversational (Percentage points)	Verificational (Percentage points)	High Cost (Percentage points)	Index (Percentage points)
Panel A: Illiterate Parents				
Treatment	-0.09***	-0.09***	-0.08***	-0.09***
	(0.025)	(0.029)	(0.019)	(0.015)
Control Mean	0.25	0.57	0.60	0.51
Observations	990	990	990	990
Panel B: Literate Parents				
Treatment	-0.10***	-0.13***	-0.03	-0.07***
	(0.019)	(0.029)	(0.020)	(0.015)
Control Mean	0.15	0.40	0.54	0.41
Observations	931	928	931	931
p-value	0.80	0.26	0.09	0.55

Notes: This table reports the effect of the intervention on the unmet gap between parents' stated responsibilities and implemented actions, separately for illiterate and literate parents. Conversational, verificational, and high-cost categories correspond to the classification of parental actions defined in the main analysis. The index averages the four canonical actions. Standard errors are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.4: Effect on the Unmet Action Gap by Student Gender (Endline)

	(1)	(2)	(3)	(4)
	Conversational (Percentage points)	Verificational (Percentage points)	High Cost (Percentage points)	Index (Percentage points)
Panel A: Female Students				
Treatment	-0.10***	-0.11***	-0.07***	-0.09***
	(0.024)	(0.033)	(0.021)	(0.016)
Control Mean	0.20	0.49	0.58	0.47
Observations	906	905	906	906
Panel B: Male Students				
Treatment	-0.09***	-0.09***	-0.04**	-0.07***
	(0.024)	(0.030)	(0.020)	(0.015)
Control Mean	0.20	0.48	0.56	0.45
Observations	1,015	1,013	1,015	1,015
p-value	0.92	0.69	0.31	0.45

Notes: This table reports the effect of the intervention on the unmet gap between parents' stated responsibilities and implemented actions, separately by student gender. Conversational, verificational, and high-cost categories correspond to the classification of parental actions defined in the main analysis. Column (4) reports the average across the four canonical actions. The P-value row should report the p-value for the equality of treatment effects between female and male students. Standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Effect on Parental Managerial Actions: Low-/Medium-Cost Practices (Endline)

	(1)	(2)	(3)	(4)	(5)	(6)
	Talked to child (Percentage points)	Asked test scores (Percentage points)	Asked test dates (Percentage points)	Supervised homework (Percentage points)	Checked during lesson review (Percentage points)	Checked notebook (Percentage points)
Treatment	0.06***	0.11***	0.17***	0.11***	0.09***	0.12***
	(0.012)	(0.017)	(0.019)	(0.020)	(0.019)	(0.022)
Control mean	0.88	0.77	0.63	0.60	0.70	0.46
Observations	1,921	1,921	1,921	1,921	1,921	1,921

Notes: Each column reports the treatment effect on the outcome listed in the column header. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Effect on Parental Managerial Actions: High Startup Practices (Endline)

	(1)	(2)	(3)	(4)
	Has study schedule (Percentage points)	Visited school (Percentage points)	Talked to teacher (Percentage points)	Checked attendance at school (Percentage points)
Treatment	0.06***	0.08***	0.05***	0.06***
	(0.020)	(0.022)	(0.018)	(0.017)
Control mean	0.56	0.43	0.20	0.18
Observations	1,921	1,921	1,921	1,921

Notes: Each column reports the treatment effect on the outcome listed in the column header. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Effect on Parental Managerial Actions: Low/Medium Cost Practices from Students Survey at Endline

	(1)	(2)	(3)	(4)	(5)
	Talked to Child (Percentage points)	Supervised Homework (Percentage points)	Asked Test Scores (Percentage points)	Asked Test Dates (Percentage points)	Checked on the child during Lesson' Review (Percentage points)
Panel A: Extensive Margin					
Treatment	0.11***	0.13***	0.10***	0.18***	0.05***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Control Mean	0.85	0.59	0.83	0.73	0.91
Observations	1902	1902	1902	1902	1902
Panel B: Intensive Margin					
Treatment	3.75***	2.35***	3.38***	3.90***	2.26***
	(0.32)	(0.30)	(0.30)	(0.29)	(0.30)
Control Mean	10.33	5.36	6.77	5.27	13.56
Observations	1902	1902	1902	1902	1902

Notes: This table reports the effect of the intervention on low- and medium-cost parental managerial practices using information from the student survey. Panel A reports extensive margin effects (whether the action occurred at least once), while Panel B reports intensive margin effects (the number of times the action occurred). Standard errors are in parentheses and are clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.8: Effect on Parental Managerial Actions by Student Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Talked to Child (Percentage points)	Asked Test Scores (Percentage points)	Asked Test Dates (Percentage points)	Supervised Homework (Percentage points)	Checked on the Child During Lesson Review (Percentage points)	Visited School (Percentage points)	Checked Attendance (Percentage points)
Panel A: Female Students							
Treatment	0.008	-0.026	0.041	0.023	0.021	0.068*	0.004
	(0.023)	(0.027)	(0.031)	(0.032)	(0.023)	(0.035)	(0.034)
Control Mean	0.89	0.86	0.74	0.61	0.84	0.58	0.40
Observations	834	834	834	834	834	834	834
Panel B: Male Students							
Treatment	0.023	0.009	0.029	0.007	0.032	0.050	0.014
	(0.021)	(0.024)	(0.028)	(0.030)	(0.022)	(0.033)	(0.031)
Control Mean	0.87	0.81	0.73	0.58	0.82	0.57	0.35
Observations	949	949	949	949	949	948	948
P-value	0.779	0.541	0.535	0.625	0.851	0.678	0.842

Notes: This table reports the effect of the intervention on parental managerial actions separately by student gender. Columns correspond to specific parental actions measured in the follow-up survey. The P-value row reports the p-value for the equality of treatment effects between female and male students. Standard errors are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A.9: Spillovers Analysis using Grade Completion as Outcome

	(1)	(2)	(3)
	Treated students vs. students in non-intervention schools (Percentage points)	Control students in treated schools vs. students in non-intervention schools (Percentage points)	All students in treated schools vs. students in non-intervention schools (Percentage points)
Spillover	0.06***	0.02	0.04***
	(0.02)	(0.02)	(0.12)
Non-Intervention Schools Mean	0.66	0.66	0.66
Observations	5,136	5,146	6,179

Notes: This table estimates spillover effects by comparing grade completion outcomes in the 20 intervention schools to those in 20 randomly selected non-intervention schools. Column (1) compares treated students in intervention schools to students in non-intervention schools. Column (2) compares control students in intervention schools to students in non-intervention schools, identifying spillover effects on untreated students. Column (3) compares all students in intervention schools (treated and control) to students in non-intervention schools, capturing the overall school-level effect. Standard errors are in parentheses and clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A.10: Effect of the Intervention on Subject-Specific Test Scores

	(1)	(2)	(3)	(4)	(5)
	Physics (Standard deviations)	Biology (Standard deviations)	Math (Standard deviations)	Language (Standard deviations)	H&G (Standard deviations)
Panel A: Pooled					
Treatment	0.11***	0.09***	-0.02	0.03	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Control Mean	0.00	0.00	-0.00	0.00	0.00
Observations	6047	6107	6144	6177	6090
Panel B: 2nd Semester					
Treatment	0.10***	0.09***	-0.04	0.03	0.08**
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
Control Mean	0.00	0.00	-0.00	0.00	0.00
Observations	4011	4040	4073	4092	4040
Panel C: 1st Semester					
Treatment	0.12***	0.07	0.02	0.03	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Control Mean	0.00	0.01	0.00	0.01	0.01
Observations	2036	2067	2071	2085	2050

Notes: This table reports the effect of the intervention on standardized test scores by subject. H&G refers to History and Geography. Panel A pools test scores from both semesters to increase statistical power. Panel B restricts the analysis to second-semester exams (two tests per subject), and Panel C to first-semester exams (one test per subject). All regressions control for baseline performance in the corresponding subject. A double post-LASSO procedure selects control variables from those listed in Table I. Standard errors are in parentheses and clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A.11: Effect on Students Performance: Heterogeneity by Parents' Literacy

	(1)	(2)	(3)	(4)
	Grade Completion (Percentage points)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Illiterate Parents				
Treatment	0.06**	0.13**	0.07*	0.01
	(0.03)	(0.07)	(0.04)	(0.04)
Control Mean	0.66	-0.10	-0.09	-0.08
Observations	1047	1047	3133	3143
Panel B: Literate Parents				
Treatment	0.02	0.10	0.08*	0.06
	(0.03)	(0.07)	(0.04)	(0.04)
Control Mean	0.70	0.10	0.10	0.09
Observations	1009	1009	3041	3037
p-value	0.348	0.552	0.875	0.387

Notes: This table presents heterogeneity in treatment effects by parental literacy. Panel A shows estimates for students with illiterate parents, and Panel B for those with literate parents. Column (1) reports effects on grade completion, defined as a binary variable equal to 1 if the student passed to grade 9 (GPA of at least 10/20), and 0 otherwise. Column (2) reports effects on year-end GPA. Columns (3) and (4) show average standardized test scores in STEM and Humanities, respectively. STEM includes Physics, Biology, and Math. Humanities includes Language and History-Geography. Each regression (except Column 1) controls for the baseline value of the outcome and uses a double post-LASSO procedure to select covariates from Table I. All test scores are standardized. Standard errors are in parentheses and clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: Effect on Students Performance: Heterogeneity by Students' Gender

	(1)	(2)	(3)	(4)
	Grade Completion (Percentage points)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Female Students				
Treatment	0.05*	0.20***	0.07*	0.08**
	(0.03)	(0.07)	(0.04)	(0.04)
Control Mean	0.69	0.01	-0.04	0.05
Observations	969	969	2899	2897
Panel B: Male Students				
Treatment	0.05*	0.09	0.093**	0.02
	(0.03)	(0.07)	(0.04)	(0.04)
Control Mean	0.67	-0.01	0.04	-0.04
Observations	1087	1087	3275	3283
p-value	0.877	0.461	0.516	0.429

Notes: This table presents heterogeneity in treatment effects by student gender. Panel A shows estimates for female students, and Panel B for males. Column (1) reports effects on grade completion, defined as a binary variable equal to 1 if the student passed to grade 9 (GPA of at least 10/20), and 0 otherwise. Column (2) reports effects on year-end GPA. Columns (3) and (4) show average standardized test scores in STEM and Humanities, respectively. STEM includes Physics, Biology, and Math. Humanities includes Language and History-Geography. Each regression (except Column 1) controls for the baseline value of the outcome and uses a double post-LASSO procedure to select covariates from Table I. All test scores are standardized. Standard errors are in parentheses and clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.13: Effect on Students Performance: Heterogeneity by Parent Literacy and Student Gender

	(1)	(2)	(3)	(4)
	Grade Completion (Percentage points)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Illiterate Parents, Female Students				
Treatment	0.10**	0.24**	0.04	0.07
	(0.04)	(0.10)	(0.06)	(0.05)
Control Mean	0.67	-0.10	-0.10	-0.05
Observations	482	482	1439	1441
Panel B: Literate Parents, Female Students				
Treatment	0.01	0.14	0.09	0.10*
	(0.04)	(0.10)	(0.06)	(0.06)
Control Mean	0.72	0.12	0.02	0.14
Observations	487	487	1460	1456
p-value	0.127	0.399	0.66	0.787
Panel C: Illiterate Parents, Male Students				
Treatment	0.05	0.09	0.11**	-0.01
	(0.04)	(0.09)	(0.05)	(0.05)
Control Mean	0.65	-0.09	-0.08	-0.11
Observations	565	565	1694	1701
Panel D: Literate Parents, Male Students				
Treatment	0.05	0.07	0.07	0.05
	(0.04)	(0.10)	(0.06)	(0.05)
Control Mean	0.69	0.09	0.18	0.05
Observations	521	521	1578	1578
p-value	0.953	0.873	0.615	0.414

Notes: This table presents heterogeneity in treatment effects by student gender. Panel A shows estimates for female students whose parents are illiterate, Panel B for female students whose parents are literate, Panel C for male students whose parents are illiterate, and Panel D for male students whose parents are literate. Column (1) reports effects on grade completion, defined as a binary variable equal to 1 if the student passed to grade 9 (GPA of at least 10/20), and 0 otherwise. Column (2) reports effects on year-end GPA. Columns (3) and (4) show average standardized test scores in STEM and Humanities, respectively. STEM includes Physics, Biology, and Math. Humanities includes Language and History-Geography. Each regression (except Column 1) controls for the baseline value of the outcome and uses a double post-LASSO procedure to select covariates from Table I. All test scores are standardized. Standard errors are in parentheses and clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.14: Effect on Students Performance: Heterogeneity by the Person receiving the Calls (Mothers vs. Fathers)

	(1)	(2)	(3)	(4)
	Grade Completion (Percentage points)	GPA (Standard deviations)	STEM (Standard deviations)	Humanities (Standard deviations)
Panel A: Mother Receives Calls				
Treatment	0.09**	0.36***	0.11*	0.08
	(0.04)	(0.10)	(0.06)	(0.06)
Control Mean	0.66	-0.05	-0.03	-0.04
Observations	615	615	1857	1860
Panel B: Father Receives Calls				
Treatment	0.03	0.05	0.08**	-0.02
	(0.02)	(0.06)	(0.03)	(0.03)
Control Mean	0.69	0.02	0.02	0.02
Observations	1441	1441	4317	4320
p-value	0.491	0.041	0.731	0.077

Notes: Panel A shows results when mothers receive the calls, and Panel B when fathers receive the calls. Column (1) reports effects on grade completion, defined as a binary variable equal to 1 if the student passed to grade 9 (GPA of at least 10/20), and 0 otherwise. Column (2) reports effects on year-end GPA. Columns (3) and (4) show average standardized test scores in STEM and Humanities, respectively. STEM includes Physics, Biology, and Math. Humanities includes Language and History-Geography. Each regression (except Column 1) controls for the baseline value of the outcome and uses a double post-LASSO procedure to select covariates from Table I. All test scores are standardized. Standard errors are in parentheses and clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.15: Absenteeism by Parental Literacy and by Student Gender

	(1)	(2)	(3)	(4)
	Illiterate Parents (Percentage points)	Literate Parents (Percentage points)	Female Students (Percentage points)	Male Students (Percentage points)
Treatment	-0.061**	-0.072***	-0.070**	-0.069***
	(0.027)	(0.028)	(0.03)	(0.027)
Control Mean	0.61	0.49	0.52	0.67
Observations	1,061	1,024	975	1,110

Notes: This table reports treatment effects on student absenteeism, defined as a binary variable equal to 1 if the student was absent at least once. Columns (1) and (2) disaggregate effects by parental literacy; Columns (3) and (4) by student gender. All regressions control for baseline absenteeism and apply a double post-LASSO procedure to select covariates from the full list in Table I. Standard errors are in parentheses and clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.16: Heterogeneous Effects on Grade Completion and GPA by Parental Years of Schooling

	(1)	(2)
	Grade Completion (Percentage points)	GPA (Standard deviations)
Treatment	0.06**	0.13*
	(0.029)	(0.07)
Treatment*Years of Schooling	-0.002	-0.004
	(0.045)	(0.12)
Years of Schooling	0.008*	0.37***
	(0.004)	(0.01)
Observations	2076	2056

Notes: This table reports heterogeneous treatment effects by parental years of schooling. Column (1) uses a binary outcome indicating whether a student completed grade 8 (i.e., promoted to grade 9, which requires a GPA of at least 10 out of 20). Column (2) reports effects on the year-end GPA. All regressions include covariate selection via double post-LASSO from the full set of variables listed in Table I. Standard errors are clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.17: Parents' Financial Involvement

	(1)	(2)	(3)
	Total Tuition Fees Paid (CFA francs)	Tuition Fees Fully Paid (Percentage points)	Date of Tuition Fees Settlement (Days since Jan 1, 1960)
Treatment	-147.31	-0.021	-0.528
	(174.34)	(0.019)	(3.64)
Control Mean	14590	0.69	23017
Observations	1,769	1,695	1,056

Notes: This table presents the estimated effects of the intervention on parents' financial involvement in education, using tuition fee payments as proxies. Column (1) reports the total amount paid toward tuition fees. Column (2) shows results for a binary variable equal to 1 if the student's tuition fees were fully paid by the end of the school year. Column (3) reports the date on which full payment was completed, conditional on full payment being made. The date is coded as the number of days since January 1, 1960; thus, a negative coefficient indicates that treated households settled tuition fees earlier. All regressions use a double post-LASSO procedure to select covariates from the full set of baseline variables listed in Table I. Standard errors are in parentheses and clustered at the individual level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.18: Lee Bounds for Grade Completion and GPA

	Grade Completion			GPA		
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower	Higher	Main Estimate	Lower	Higher	Main Estimate
Panel A: Pooled						
Treatment	0.038**	0.043**	0.039**	0.08*	0.13***	0.11**
	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
Panel B: 2nd Semester						
Treatment	-	-	-	0.09**	0.14***	0.12***
	-	-	-	(0.05)	(0.05)	(0.05)

Notes: Columns 1, 2, 4, and 5 report the Lee Bounds (lower and upper bounds) from regression of grade completion and year-end GPA on treatment. To generate the lower (upper) bounds, I compare the proportion in treatment and control groups who respond to the survey and trim the excess respondents with the highest(lowest) values in the group with more respondents. Standards errors are clustered at the individual levels and are in parentheses. Significance: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.19: Lee Bounds for STEM and Humanities

	STEM			Humanities		
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower	Higher	Main Estimate	Lower	Higher	Main Estimate
Panel A: Pooled						
Treatment	0.06**	0.08***	0.07**	0.02	0.05*	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Panel B: 2nd Semester						
Treatment	0.05	0.07**	0.06**	0.03	0.06**	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Panel C: 1st Semester						
Treatment	0.09**	0.09***	0.09**	-0.01	0.01	0.00
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)

Notes: Columns 1, 2, 4, and 5 report the Lee Bounds (lower and upper bounds) from regression of STEM and Humanities test scores on treatment. To generate the lower (upper) bounds, I compare the proportion in treatment and control groups who respond to the survey and trim the excess respondents with the highest(lowest) values in the group with more respondents. Standards errors are clustered at the individual levels and are in parentheses. Significance: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A.20: Lee Bounds for Household Chores and Parents Information About their Children's Performance

	Student doing Less Chores			Parental Error about Student Performance			Parent Is Correct about Student's Passing/Failing		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Lower	Higher	Main Estimate	Lower	Higher	Main Estimate	Lower	Higher	Main Estimate
Treatment	0.10***	0.13***	0.11***	-0.13*	-0.08**	-0.09**	0.05***	0.07***	0.06***
	(0.02)	(0.02)	(0.02)	(0.08)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)

Notes: Columns 1, 2, 4, and 5 report the Lee Bounds (lower and upper bounds) from regression of reduced chores and parental information about their children's performance on treatment. To generate the lower (upper) bounds, I compare the proportion in treatment and control groups who respond to the survey and trim the excess respondents with the highest(lowest) values in the group with more respondents. Standards errors are clustered at the individual levels and are in parentheses. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.21: Students Endline Performance across Different Samples

	(1)	(2)	(3)
	Treat & Control	Treat & 45% sample	Control & 45% sample
Coefficient	0.04**	0.04**	-0.002
	(0.02)	(0.02)	(0.02)
Comparison group mean	0.68	0.68	0.68
Observations	2 056	2 563	2 574

Notes: This table examines the comparability between the main study sample and the 45% of 8th-grade parents who are not part of our study sample. The outcome is grade completion, defined as transitioning from grade 8 to 9 (i.e., earning a year-end GPA of at least 10/20). Column 1 reproduces the main treatment effect estimate from Table V. Column 2 compares students in the treatment group to those in the 45% non-invited group. Column 3 compares control students to the 45% group. Similar analyses using other outcomes such as year-end GPA yield consistent results. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.22: Randomization Inference for Table V

	(1)	(2)	(3)	(4)
	Grade Completion	GPA	STEM	Humanities
Panel A: Pooled				
Treatment	0.04**	0.11**	0.07***	0.03
	[0.033]	[0.016]	[0.006]	[0.24]
Control Mean	0.68	0.00	0.00	0.00
Observations	2056	2056	6174	6180
Panel B: 2nd Semester				
Treatment	-	0.12**	0.06**	0.05
	-	[0.011]	[0.049]	[0.11]
Control Mean	-	0.00	0.00	0.00
Observations	-	2056	4091	4094
Panel C: 1st Semester				
Treatment	-	0.07	0.09**	0.00
	-	[0.11]	[0.008]	[0.993]
Control Mean	-	0.01	0.01	0.01
Observations	-	2094	2083	2086

Notes: This repeats Table V results and provides randomization inference p-values in brackets. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$ based on the randomization inference p-values.

Table A.23: Effect of the Intervention on Student Performance using Restricted Sample

	(1)	(2)	(3)	(4)
	Grade Completion	GPA	STEM	Humanities
Panel A: Pooled				
Treatment	0.04**	0.11**	0.08***	0.02
	(0.02)	(0.05)	(0.03)	(0.03)
Control Mean	0.7	0.02	0.01	0.02
Observations	1 879	1 879	5 637	5 637
Panel B: 2nd Semester				
Treatment	-	0.12***	0.07**	0.02
	-	(0.05)	(0.03)	(0.03)
Control Mean	-	0.02	0.01	0.03
Observations	-	1 879	3 758	3 758
Panel C: 1st Semester				
Treatment	-	0.07	0.10***	0.01
	-	(0.05)	(0.04)	(0.04)
Control Mean	-	0.04	0.02	0.02
Observations	-	1 879	1 879	1 879

Notes: This table replicates Table V but restricts the sample to only students with all their test scores available for all the tests and every subject. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.24: Table V using ANCOVA

	(1)	(2)	(3)	(4)
	Grade Completion	GPA	STEM	Humanities
Panel A: Pooled				
Treatment	0.04**	0.09*	0.06**	0.02
	(0.02)	(0.05)	(0.03)	(0.03)
Control Mean	0.68	0.00	0.01	0.00
Observations	2056	2056	6174	6180
Panel B: 2nd Semester				
Treatment	x	0.10**	0.05*	0.04
	x	(0.05)	(0.03)	(0.03)
Control Mean	x	0.00	0.00	0.00
Observations	x	2056	4091	4094
Panel C: 1st Semester				
Treatment	x	0.05	0.09**	-0.01
	x	(0.05)	(0.04)	(0.03)
Control Mean	x	0.01	0.01	0.01
Observations	x	2094	2083	2086

Notes: The table shows the results in Table V using an ANCOVA specification. Standards errors are clustered at the individual levels and are in parentheses. Significance: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

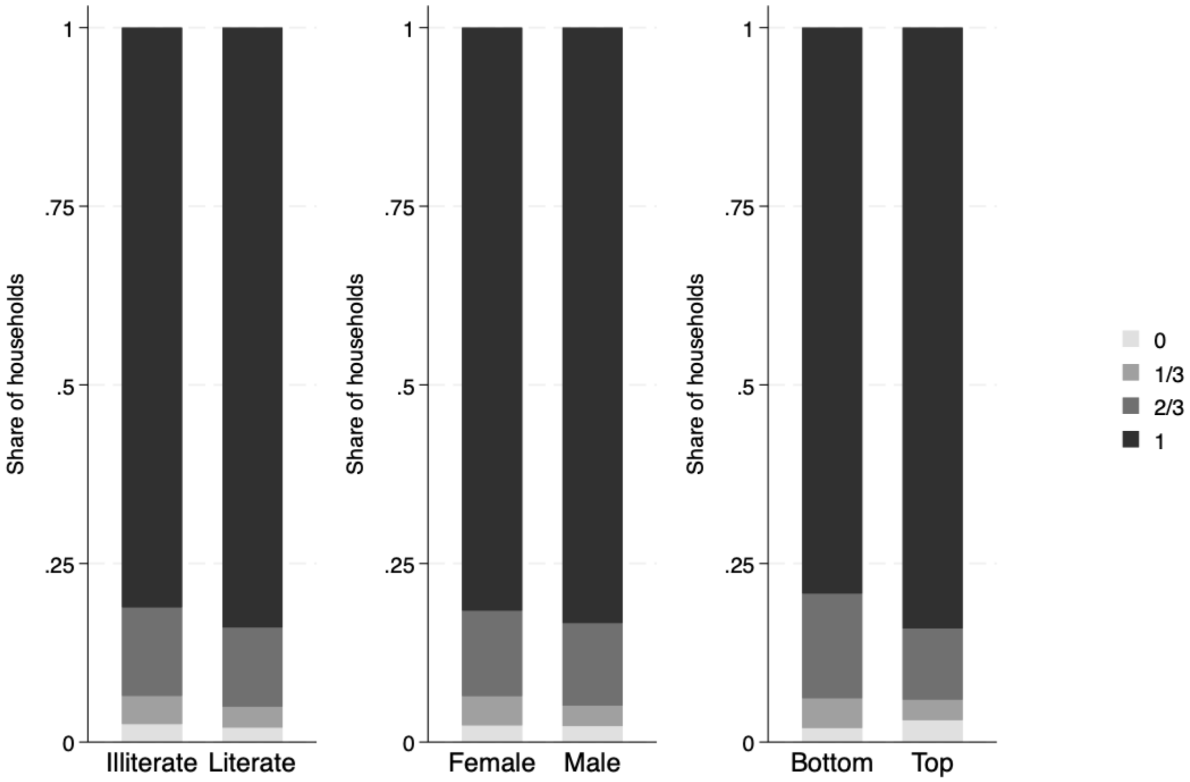
Table A.25: Cost Benefit Results adjusting for Inflation

Authors	**Interventions**	**Cost Benefits (SD per $100)**
Yedomiffi(2026)	Parents' Managerial Role	1.57
Duflo, Dupas, and Kremer 2011	Streaming, Kenya	16.44
Glewwe, Ilias, and Kremer 2010	Teachers Incentive, Kenya	2.97
Kremer, Miguel, and Thornton 2009	Scholarship Kenya	1.38
Duflo, Dupas, and Kremer 2011	ETP and streaming, Kenya	0.93
Baird, McIntosh, and Ozler 2010	CCT Malawi	0.65
Baird, McIntosh, and Ozler 2010	UCT Malawi	0.06

Notes: The table reports an illustrative cost-effectiveness benchmark: achievement gains per $100 spent, expressed in standard deviation units and scaled to 2011 USD. Program costs for this study are converted from 2023 USD to 2011 USD using the same CPI-based inflation adjustment used in Bhula et al. (2020). I accessed the report by Bhula, Mahoney, and Murphy 2020. Because outcomes and contexts differ across studies, these comparisons are descriptive rather than a definitive ranking. The primary policy-relevant benefits in this paper operate through progression outcomes (grade completion and dropout), which are not fully captured by SD-per-dollar calculations.

B Appendix Figures

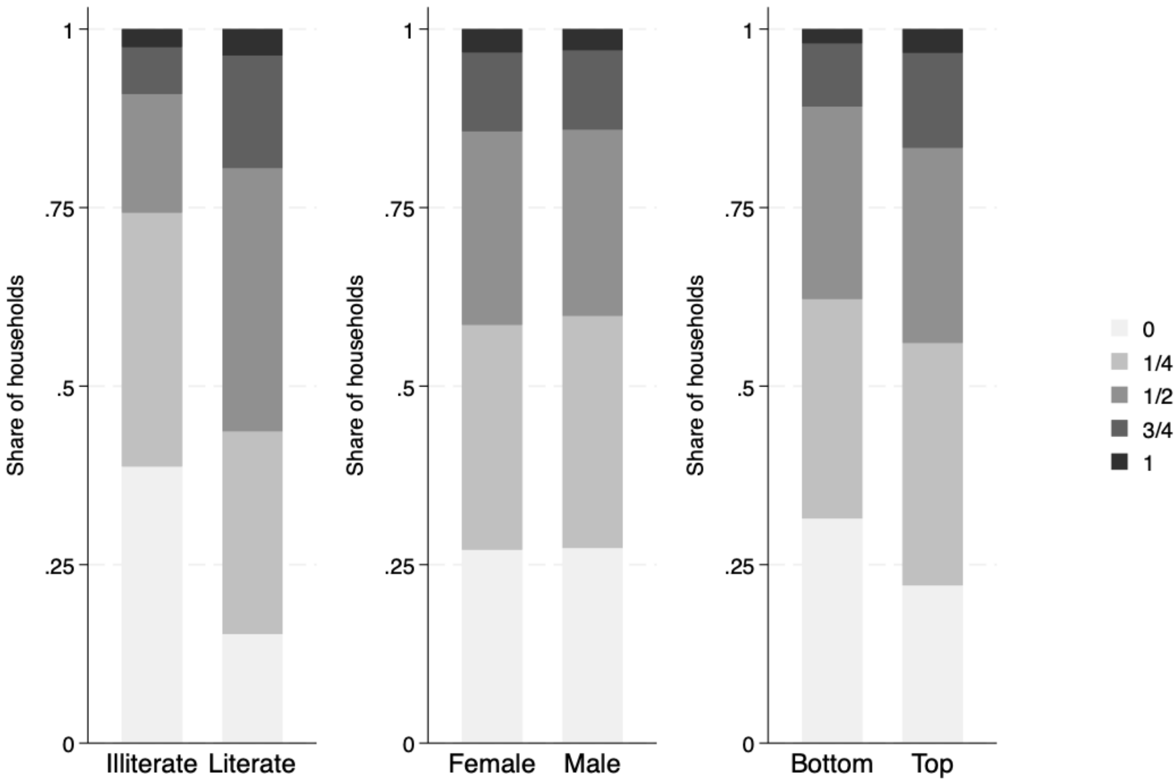
Figure A.1: Distribution of Endorsed Parental Roles at Baseline by Subgroup



Notes: This figure shows the distribution of the number of parental roles households report endorsing at baseline. The four canonical roles include parents being responsible for asking test scores, ensuring child review lessons, ensuring notebook is up to date, and ensuring child attend school. The bars show the share of households in each subgroup reporting that number of roles.

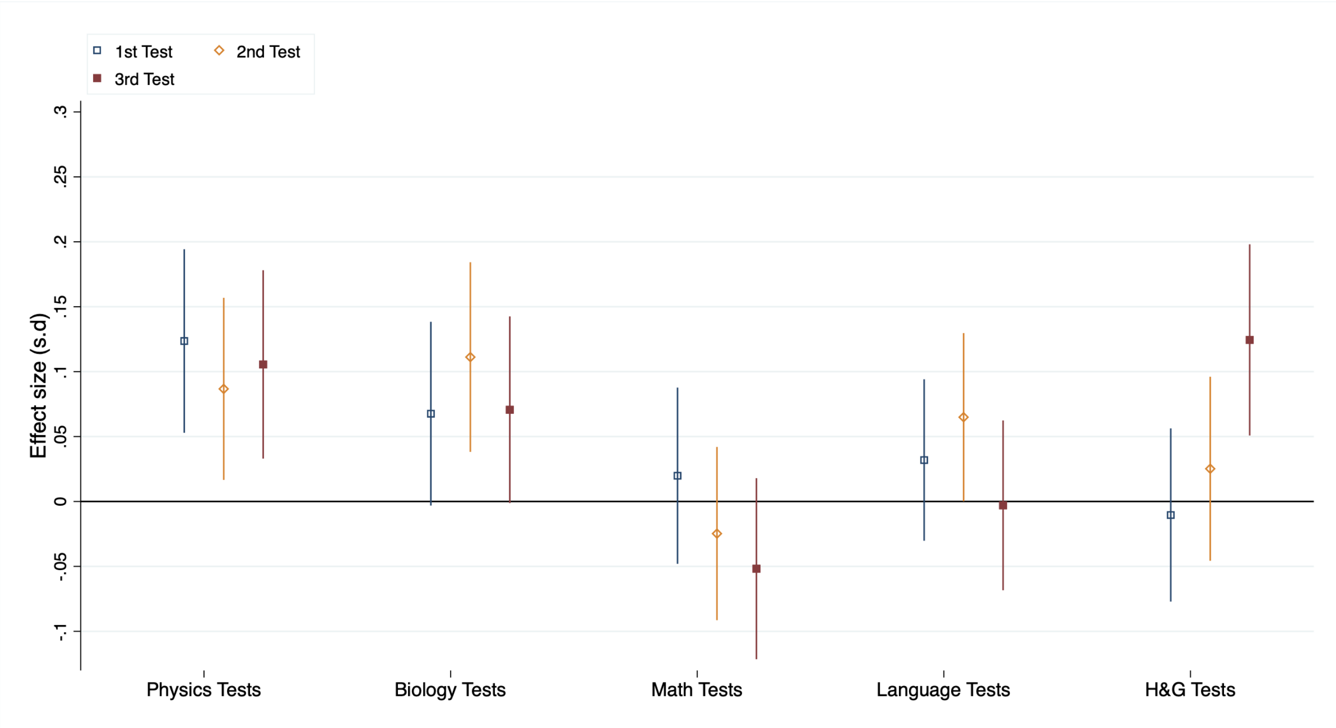
Alt text: Bar chart showing the distribution of the number of parental roles endorsed by households at baseline across subgroups defined by parental literacy, student gender, and baseline performance.

Figure A.2: Distribution of Implemented Parental Actions at Baseline by Subgroup



Notes: This figure shows the distribution of the number of parental roles for which households report taking concrete actions at baseline. The four canonical roles correspond to the same categories measured at baseline. The bars show the share of households in each subgroup.
Alt text: Bar chart showing the distribution of the number of parental actions implemented by households across subgroups defined by parental literacy, student gender, and baseline performance.

Figure A.3: Effect of the Intervention on Each of the Three Tests



Notes: This figure displays the estimated treatment effects of the intervention on subject-specific test scores across three examinations, along with 90% confidence intervals. The first test corresponds to the second exam of the first semester and occurred approximately ten days after the intervention began. The second and third tests were administered during the second semester. Subject-specific scores are grouped as follows: History and Geography (H&G); STEM (average of Math, Physics, and Biology); and Humanities (average of Language and H&G). Each test score is standardized. Each regression includes all three tests per subject to increase statistical power and controls for baseline performance in the corresponding subject. Additionally, regressions employ a double post-LASSO procedure to select control variables from those listed in Table I. Standard errors are clustered at the individual level.
Alt text: Point estimates with confidence intervals showing treatment effects on standardized test scores across three examinations for STEM, Humanities, and History–Geography subjects.