How Much Do I Matter?
Teacher Self-Beliefs, Effort, and Education Production*

Jalnidh Kaur†

November 29, 2024

Abstract

Teacher effort is critical for student learning. In many developing countries, however, teachers often perceive only a weak mapping between their effort and what students learn. I conduct an experimental evaluation of a psychosocial intervention in India that targets teachers' beliefs about *perceived control* – beliefs about one's ability to influence outcomes. The curriculum trains participants to build skills for navigating complex situations, avoiding any references to teaching or pedagogy. I study the extent to which this intervention affects teachers' beliefs about education production, their effort in class, and their students' academic performance. I devise a novel experimental task to elicit beliefs through revealed preference, about the relationship between their teaching effort and the performance of students in their classroom. I find that the intervention induced a 15% increase in teachers' beliefs about their ability to increase learning. Treated teachers exert greater effort at the intensive margin, scoring 0.13 SD higher on an index of classroom effort. They also spend more time grading student work and provide more detailed feedback to students. Finally, I find that the intervention raised student learning by 0.09 SD in classrooms taught by teachers in the treatment group. The gains are driven by students who are taught by teachers with low levels of perceived control beliefs at baseline. These findings suggest that teachers' beliefs about perceived control can serve as a powerful lever for changing teaching practice and raising learning levels in developing countries. *JEL Codes:* D91, I21, I24, J24

*The author is grateful to Alex Eble, Cristian Pop-Eleches, Supreet Kaur, Peter Bergman, and Mark Dean for invaluable support, mentorship and guidance. I thank Andreas de Barros, Sandra Black, Christina Brown, Karthik Muralidharan, Sabareesh Ramachandran, Juan Saavedra, Harshil Sahai, Judith Scott-Clayton, Abhijeet Singh, Lena Song, and Ying Xu for their feedback and suggestions across multiple stages of this project. I am grateful to Kalgidhar Trust for their support to run the study across Akal Academy schools. I thank WorldBeing, especially Kate Sachs Leventhal, for sharing the curriculum and providing helpful inputs regarding intervention design. Sukhmani Kaur, Jappanjot Singh and Sehajbir Singh provided excellent research assistance and field support. I acknowledge generous funding from the Weiss Fund for Research in Development Economics, the Agency Fund, and the Teachers College Economics of Education Program Research Grant. The experiment was approved by the Teachers College Institutional Review Board and was registered with a pre-analysis plan at AEA RCT registry (RCT ID: AEARCTR-0009803). All errors are my own.

†University of Glasgow, Adam Smith Business School. E-mail: jalnidh.kaur@glasgow.ac.uk

1 Introduction

Teacher effort is a key determinant of student learning. Systematic reviews of evidence show that the most effective interventions that influence student learning are driven by teachers (Béteille and Evans, 2021; Snilstveit et al., 2015).[1] Experimental studies find that the majority of gains in student learning due to teacher performance-pay interventions arise from changes in teacher effort in the classroom, beyond changes at the selection or attendance margins, establishing a causal link between teacher effort and student learning (Leaver et al., 2021, Muralidharan and Sundararaman, 2011).

However, in contrast, many teachers in low- and middle-income countries believe that their effort has only a limited impact on student achievement (Sabarwal and Abu-Jawdeh, 2018). For example, on an average, more than 40% of teachers across ten developing countries self-report that there is little they can do to help students learn if students come from disadvantaged backgrounds with financial constraints, low levels of parental education, or have prior academic deficits (Sabarwal et al., 2022).

Beliefs about the production function – specifically about one's role and influence over production – have substantial implications for effort decisions (Adams-Prassl et al., 2023; Attanasio et al., 2022; Ersoy, 2023; List et al., 2021). These beliefs have been extensively studied in psychology under the broad construct of *perceived control* – beliefs about one's ability to influence outcomes – and have been shown to be fundamental drivers of motivation and behavior (Bandura, 1977; Rotter, 1966; Skinner, 1996; Weiner, 1985).[2] In other words, if one believes that one has little control over outcomes, it is rational not to try. The fact that teachers in low-resource contexts perceive low levels of control in influencing learning generates worrying implications for teacher effort and its effects on student learning.

It is not unreasonable why teachers might develop such beliefs. In low-resource settings, teachers often face challenging work environments with many factors that are outside their control including high shares of first-generation learners, wide heterogeneity in learning levels, and limited institutional support (World Bank, 2018). These environments likely shape their perceptions of "how much they matter" for producing learning.

In this paper, I study the role of beliefs about perceived control in influencing teachers' effort, by directly targeting these beliefs using a psychosocial intervention. In particular, I investigate two research questions. First, to what extent are teachers' beliefs about perceived control over learning malleable? Second, what is the causal impact of a psychosocial intervention targeting teachers' control beliefs on teacher effort and student learning? To answer these questions, I conduct a randomized controlled trial across 83 schools, 292 teachers, and 7,570 students in a rural private school chain in India.

[1] Beteille and Evans (2021) summarize evidence showing that interventions driven by teachers had the highest gains for learning as opposed to those driven by the community, schools, or technology.

[2] Perceived control is an umbrella term capturing a variety of control-related beliefs such as locus of control, self-efficacy, and agency (Reich and Infurna, 2017). As per the agent-means-ends conceptualization by Skinner (1996), self-efficacy beliefs connect agent to means in the first part of the sequence, reflecting beliefs about competence; locus of control beliefs connect means to end in the second part of the sequence capturing beliefs about the causal relationship between one's actions and outcomes; while perceived control is an interaction between the two, capturing self-beliefs about one's ability to influence outcomes.

I randomize teachers to receive a psychosocial intervention targeting beliefs about perceived control in everyday situations. The curriculum is focused on general skill development with no references to teaching or pedagogy. The content uses skill-building and control-enhancing approaches that train participants to recognize differences between controllable and uncontrollable events, and to navigate different aspects of challenging situations. This is done through building awareness about internal strengths and capabilities, and imparting skills such as goal-setting and problem-solving to deal with stressful situations that may seem beyond one's control. In a setting where teachers perceive the effort-outcome relationship to be flat, the intervention, thus, targets the slope of this relationship by empowering teachers with skills that build their confidence to perceive a positive relationship between their effort and outcomes. The curriculum is designed with guidance from *WorldBeing*, an organization that creates evidence-based programs in psychology for adults and youth in developing countries. I partnered with *Sukrit* (a local NGO) to adapt the curriculum to the context and deliver it to teachers.

The intervention was delivered over ten sessions across five weeks. All training sessions were conducted after school hours to avoid crowding out of teacher time on other tasks. The curriculum intentionally avoids any references to classroom practices or pedagogy. This feature minimizes the scope for experimental demand effects. Further, to isolate the effect of the psychosocial component from any other effects of training attendance unrelated to perceived control beliefs, I used an active control (placebo) group. Teachers assigned to the placebo group also received training of a similar structure, duration, and format, but with content unrelated to personal development. The presence of a placebo group allows me to isolate the effect of the training content from effects due to interaction with a group or an external speaker during training sessions.

I devise a novel method to measure teachers' beliefs about their influence over student learning. Rather than relying on traditional survey questions, I develop a revealed preference measure of teachers' confidence about their ability to increase student learning through an incentivized real-stakes experimental task. This measure uses a price list approach and presents teachers with a sequence of choices between an unconditional flat pay or a performance-linked pay that conditions payment on test-score improvements of low-performing students.[3] The choices in each round keep the flat-pay option fixed but incrementally increase the stakes of the performance-pay contract with a higher minimum threshold for test-score improvement. The sequence of choices is designed to identify the value that makes teachers indifferent between unconditional pay and performance-linked pay. The switching point elicits teachers' degree of confidence in their abilities to exercise control over student learning. I document that this revealed preference measure is positively correlated with psychological measures of self-efficacy and locus of control.

I find that exposure to psychosocial intervention positively impacts teachers' beliefs. Compared to the control group, the treatment teachers exhibit a 15% increase (p-value = 0.024) in confidence in their abilities to raise student learning as elicited through the revealed preference measure. Teachers in the treatment group are more likely to forego a

[3] I define low-performing students as those in the bottom half of the baseline distribution. I specifically link incentives to test-score improvements of low-performers — who are more likely to lack parental and household inputs — to explicitly tease out teachers' beliefs about the role of *their* effort in influencing student learning.

monetary bonus received with certainty in exchange for a higher-paying monetary bonus that is conditional on the test-score improvement of low-performing students in their classroom—revealing a higher confidence in their ability to raise student scores by their own effort. Of note, this is a setting where all exams are externally graded, so teachers are unable to manipulate test scores in their favor. This shift in beliefs persists six months after the intervention and is not driven by changes in risk attitudes. I also elicit beliefs directly using standard psychological scales and find that the intervention raised teacher locus of control by 0.16 SD (p-value = 0.029), and generalized self-efficacy by 0.11 SD (p-value = 0.16).

I examine impacts across multiple dimensions of teacher effort including attendance, classroom effort, grading effort, and time use. I find no effect on the extensive margin of attendance which is already high in the setting. On the intensive margin, I find that teachers in the treatment group exert more effort in the classroom and perform 0.13 SD (p-value = 0.021) higher on a pre-registered index of measures capturing the quality of teaching gauged through classroom observations by independent observers. Direct physical observation of teacher activity may lead teachers to showcase their best practices in the presence of an observer. To tackle this concern, I also examine a measure of *past cumulative effort* by examining the homework notebooks of students and scoring these for the nature of feedback provided by teachers prior to the class visit. I find that teachers in the treatment group score 0.11 SD (p-value = 0.012) higher on this index of grading effort, and are more likely to provide detailed feedback on student work. Additionally, self-reported time use also shows consistent patterns. Teachers in the treatment group spend an additional 8 minutes (p-value = 0.002) on grading notebooks compared to control teachers.

These findings are interesting, especially given the fact that the training did not outline any expectations from teachers to exert more effort, nor did it provide any skills related to teaching. This was not an intervention about *what to teach* or *how to teach*. Yet, we see these behavioral effects on teacher effort suggesting that teachers had the innate content and pedagogical knowledge to begin with, but were not putting it into use in the status quo.

These increases in effort are meaningful if they translate into students' learning. I find that the psychosocial intervention was indeed effective in improving student learning. At the end of the academic year, students taught by treatment teachers performed significantly better than those taught by control teachers by 0.09 SD (p-value = 0.036), as reflected in math scores. Surprisingly, the gains are not driven by students at a particular end of the test-score distribution but are spread across baseline achievement levels. Thus, even though teacher incentives were designed to reward gains for the bottom half of the student distribution, students across the entire support of the distribution gained, suggesting the potential for efficiency gains from such interventions. Further, I find that treatment effects are higher for students taught by teachers with low levels of perceived control beliefs at baseline.

I adopted a range of strategies in the experimental design to address concerns about experimenter demand effects. First, the intervention content was purged of any references to classroom practices.[4] Second, teachers were not aware that their effort would be measured after the intervention. An independent field team conducted classroom observations, and this team was different from the team that conducted the intervention. Third, to deal with

[4]This was packaged as a general training for teachers covering content corresponding to general life situations. Any effects on teaching beliefs reflect "far transfer", that is, improvements in a different domain.

concerns regarding Hawthorne effects, I used multiple measures to capture dimensions of effort that were hard for teachers to manipulate on the spot. These included measures of past effort such as the nature of feedback provided in graded work.

Further, I show that it is unlikely that the intervention worked through related psychological mechanisms that I measure, such as growth mindset, risk preferences, and mental health. Instead, self-beliefs are a plausible mechanism explaining the effects on perceived control. Treatment teachers score 0.23 SD (p-value = 0.004) higher on an aggregate index of self-beliefs composed of self-efficacy, locus of control, and beliefs about the perceived importance of teacher inputs for education production.

My study contributes to three main bodies of literature. First, it contributes to a growing body of work across economics and psychology studying the role of beliefs in human capital investment decisions. In particular, my work relates to specific types of beliefs studied in the economics literature — production function beliefs about the effectiveness of inputs for producing human capital, and beliefs about perceived returns to effort. These have been studied primarily for parents (Attanasio et al., 2019; Bhalotra et al., 2020; Carneiro et al., 2019; Dizon-Ross, 2019) and students (Ersoy, 2023; Rury and Carrell, 2023), but not for teachers, who have a formative influence on human capital formation of children during school years. In contrast, a rich literature in psychology has examined the importance of these beliefs for teachers, however, existing work remains descriptive (Rose and Medway, 1981; Gibson and Dembo, 1984, Tschannen-Moran et al., 1998).[5] I provide the first evidence that teachers' beliefs about perceived control are malleable and can be influenced through targeted interventions – advancing work across economics and psychology.[6]

Second, I contribute to the education literature on how to improve teacher performance in developing countries, by advancing our understanding of interventions that raise teacher effort. Earlier works have used monetary incentives and monitoring to incentivize teacher effort (Muralidharan and Sundararaman, 2011; Glewwe et al., 2010; Duflo et al., 2012). Recent works have used alternative pedagogical approaches to improve student learning (Gray-Lobe et al., 2022; Nourani et al., 2023). These approaches have largely focused on targeting extrinsic motivation and external constraints, such as, the pace of the curriculum, heterogeneity in learning levels, and lack of accountability and standardization, which have been recognized as key structural constraints limiting the effectiveness of teacher effort (Muralidharan et al., 2019; Pritchett, 2013). My results provide one of the first pieces of evidence showing that alleviating internal, psychological constraints can raise teacher effort and student learning in the developing world. In that light, my work advances the literature by adopting a behavioral perspective on the teacher effort decision.

Third, my work relates to a growing literature in behavioral economics on the importance of psychological constructs for decision-making. Prior work has established the centrality of related self-beliefs, including, locus of control, self-efficacy, and beliefs about agency for

[5]This line of work has documented a robust correlation of teachers' beliefs about *self-efficacy* and *locus of control* with instructional behavior and student achievement across settings.

[6]Recent work in psychology has shown that experimental interventions can be used to enhance perceptions of perceived control for college students and middle-aged adults in the context of health and aging (Frazier et al., 2015; Hintz et al., 2015; Zautra et al., 2012; Lachman et al., 2011). However, there is no evidence for malleability of these beliefs for teachers in education settings. See Reich and Infurna (2017) for an excellent review of psychology literature on the concept of perceived control.

motivating human behavior, however, with limited evidence on malleability (Caliendo et al., 2023; Cobb-Clark, 2015; McKenzie et al., 2021).[7] My study provides one of the first pieces of causal evidence that self-beliefs about perceived control can be shaped, with substantive implications for field outcomes and behaviors. A notable exception is McKelway (2021), who also finds evidence for the malleability, of self-efficacy beliefs in particular, in a setting with adult women with effects on labor market outcomes. A distinctive feature of my study is the finding that self-beliefs can be shaped even among highly experienced economic agents: teachers who have been in their jobs for many years. The study also adds to a burgeoning literature that uses soft-touch psychological interventions to influence economic outcomes by highlighting the role of another important psychological construct for behavioral change.[8]

While situated in the context of schools and teachers, this paper also contributes to the broader literature on the impact of soft skills on worker productivity in firms (Adhvaryu et al., 2023; Campos et al., 2017; Groh et al., 2012, McKelway, 2021). My results are relevant for answering broader questions about the role of soft skills in influencing effort and productivity in organizations.

More generally, the findings of the study have important policy implications, especially in the context of teacher professional development programs across developing countries. While countries engage in massive spending to invest in teacher capacity building through in-service training, a wide range of evidence suggests that these programs fail to produce systematic improvements in instructional practice and student achievement, due to deficiencies in the content and delivery of these programs (Loyalka et al., 2019; Popova et al., 2018).[9] Existing curricula focuses on building content and pedagogical knowledge with limited emphasis on targeting teacher beliefs and motivation which may be essential for ensuring that teachers make use of the skills and apply these in their teaching. My results suggest that using targeted psychological content that enables teachers to perceive themselves as active agents who are able to influence learning may provide a promising option to influence teacher productivity at scale by being incorporated into traditional professional development programs.[10]

The rest of this paper is organized as follows: Section 2 motivates the empirical relevance of the research questions. Section 3 describes the experimental design, including the data and outcomes, and Section 4 discusses the empirical strategy. Sections 5 and 6 present results and discussion. Section 7 presents conclusions.

---

[7] In its extreme, loss of agency and control have been studied in works targeting depression among specific populations, and psychotherapy interventions have been found to be effective (Baranov et al., 2020; Bhat et al., 2023). There is limited work exploring the importance of control-related beliefs in less extreme settings.

[8] These works have used psychological interventions that target aspirations (Orkin et al., 2023; Riley, 2024), growth mindset (Ganimian, 2020; Yeager et al., 2022), grit (Alan et al., 2019), build visualization, and planning skills (John and Orkin, 2022; Ashraf et al., 2022) through short-term training.

[9] Loyalka et al. (2019) show that the ineffectiveness of a national PD program in China stemmed from the content being overly theoretical with limited relevance, emphasis on rote learning, and passive delivery.

[10] The intervention also provides a more cost-effective option for raising teacher productivity compared to approaches using incentives and monitoring, which are costly and hard to enforce due to a lack of trust in the fairness of evaluation systems.

2 Motivation

2.1 Teachers' Beliefs Across Low-Income Contexts

In this section, I establish empirical support for the hypothesis that teachers perceive a weak mapping between their ability, effort, and student learning in low-resource contexts by compiling descriptive evidence from teacher surveys across multiple contexts.

First, Sabarwal et al. (2022) document teachers' beliefs using survey data from 20,000 teachers at the primary and lower-secondary level across nine developing countries including Afghanistan, Nepal, Pakistan, Senegal, Nigeria, Tanzania, Argentina, Indonesia, Myanmar and Tajikistan. They find that averaging across countries, around 40% of teachers report that there is not much they can do to help students learn if students come from disadvantaged backgrounds with financial constraints, low levels of parental education, or have prior academic deficits. They find striking and consistent patterns across lower-income countries – for example, the share of teachers in Nigeria, Pakistan and Zanzibar who believe that their influence is limited for first-generation learners are 52, 53, and 50 percent respectively.

Second, I compile data from the Young Lives (YL) school surveys from 2016-17 across India and Ethiopia. The surveys covered 281 teachers in India spread across 205 schools and 257 teachers across 63 schools in Ethiopia.[11] More than 60% of teachers in both countries agreed that a student's home environment directly limited the influence of their teaching. Around 50% of teachers in India, and more than 80% of teachers in Ethiopia agreed that a student's capacity to learn was limited by family background (Table 1). The broad consensus on these statements reflects the wide acceptability of these views among teachers. Given that self-reports can be subject to social conformability bias, these figures likely represent lower bounds on teachers' beliefs about their perceived role in education production.

Further, anecdotal evidence from Pritchett (2013) also reflects these patterns of beliefs. In a public meeting between a school principal and villagers in Uttar Pradesh, India, the Principal responds to complaints from dissatisfied low-income parents stating,

> "*It is not our fault. We do what we can with your children. You are [offensive term]. The children of [offensive term] are also [offensive term]. We cannot be expected to teach your children.*"

Taken together, these patterns show that teachers routinely undermine the importance of their effort, and perceive the effort-learning relationship to be flat, especially in light of disadvantages at the student and household levels.

2.2 Psychological Underpinnings

Literature in psychology has studied an array of constructs related to control – including self-efficacy, locus of control, agency, and learned helplessness. Perceived control is

[11] Young Lives is a study of child poverty in India, Peru, Ethiopia and Vietnam. The study follows two cohorts of children in these countries through household surveys. YL conducted school surveys in 2010-11 and 2016-17 where schools were chosen to be representative of sites used for household surveys.

an umbrella term encompassing two key constructs: self-efficacy and locus of control. Self-efficacy is defined as the belief in one's ability to engage in specific behavior or execute actions (Bandura, 1977), whereas, locus of control centers around causal beliefs about action-outcome contingencies (Rotter, 1966). According to Skinner's (1996) agent-means-end sequence, self-efficacy links agents to means, whereas locus of control connects means to ends. Perceived control is conceptualized as the interaction between the two, linking agents to ends (Reich and Infurna, 2017).

Beliefs about perceived control have been shown to be strong determinants of motivation and behavior. When people perceive situations as controllable, they exhibit high levels of interest, exert effort, and persist in the face of failures. In contrast, when people perceive low levels of control, they withdraw, retreat, or give up. Further, beliefs about perceived control are self-reinforcing. Those with low levels of perceived control tend to avoid difficult tasks and forego opportunities, eventually reinforcing their beliefs about low perceived control.

Where do these beliefs come from? Prior work has documented that aversive environments with low levels of control can condition individuals to develop low perceived control with behavioral effects (Sherrod et al., 1977; Hiroto and Seligman, 1975).[12] In developing country contexts, teachers often face challenging work environments involving high shares of first-generation learners, wide heterogeneity in student learning within classrooms, and resource constraints – situations that are largely outside teachers' control. Repeated exposure to these low-control environments might be a potential pathway for teachers to develop low perceived control in the first place.

To what extent is perceived control malleable? Recent work in psychology has shown that perceived control can be manipulated using targeted interventions that train participants on recognizing differences between controllable and uncontrollable events, and focus on developing coping skills as pathways out of difficult situations. These interventions have been shown to enhance perceptions of control, mainly in the context of health and aging (Frazier et al., 2015; Hintz et al., 2015; Zautra et al., 2012; Lachman et al., 2011). McKelway (2021) implemented a psychosocial curriculum among adult women in India to target labor supply, and showed that it increased self-efficacy. While these works relate to settings outside of education, they provide us some grounding to think of perceived control as potentially malleable in the classroom setting.

My intervention builds on these works and uses a curriculum that is focused on building perceived control by targeting beliefs about competence (self-efficacy) as well as beliefs about effort-outcome contingencies (locus of control), using tools from positive psychology. Notably, the curriculum targets beliefs about perceived control in everyday situations, rather than beliefs about teaching per se. The intervention design intentionally excluded any references to teaching skills or classroom practices. Consequently, any effects on teachers' effort in the classroom are unlikely to be driven by the intervention having encouraged teachers to increase effort.

[12]The extreme case is that of learned helplessness where a person perceives a complete lack of control, leading to withdrawal. The person does not even try even when opportunities are available.

3 Experimental Design

3.1 Setting

I conduct the study in partnership with Akal Academies – a large chain of rural, affordable private schools in north India.[13] With an annual tuition of $300 a year and with subsidies and scholarships for lower-income students, the schools primarily cater to middle- and lower-income households in rural areas. Most students in the setting come from agricultural backgrounds with limited parental support and involvement in learning. Close to a third of teachers report that a majority of students in their classroom are first-generation learners. As is typical in many developing countries, teachers teach classrooms with wide variation in achievement levels.

The study sample consists of 292 teachers and their students in grades 2, 4, 6, and 8 across 83 schools. Table A1 shows descriptive characteristics of schools and teachers in the study sample. Schools have an average of 500 students, a student-teacher ratio of 18 students per teacher, and are led by an experienced school principal with more than ten years of teaching experience. Teachers are mostly female (82%), young (33 years on average), with a median experience of 6 years. Nearly all teachers have a BA, around 70% hold a Masters degree or higher, and around 58% have a permanent (as opposed to contractual) status.

3.2 Intervention

I design an intervention that targets teachers' perceptions of control in everyday situations, in other words, beliefs about the slope of the perceived effort-outcome relationship. All teachers in the study were invited to attend a five-week training consisting of 10 sessions, delivered after school hours over Zoom. Teachers were randomized to a treatment or a placebo arm. Both arms received training of the same duration, structure, and format but with different content.

Treatment Arm. Teachers in the treatment arm were exposed to a psychosocial curriculum that uses skill-building and control-enhancing approaches to target perceptions of control. The curriculum was designed with guidance from *WorldBeing* - an organization that develops evidence-based content from positive psychology for disadvantaged communities. I collaborated with *Sukrit*, a local NGO that develops motivational programs for youth and adults in the region, to adapt the content and deliver to teachers. The adaptation process involved the addition of modules emphasizing strategies for secondary control, incorporating culturally relevant examples, along with translation to the local language.[14]

[13] Akal Academies are a chain of 129 rural schools spread across the states of Punjab, Haryana, Uttar Pradesh, with headquarters in Himachal Pradesh.

[14] Psychology literature distinguishes between strategies to exercise control. Primary control (attempt to change the situation) is different from secondary control (relinquishing control to a powerful other). The latter is especially relevant to Eastern cultures with less individualistic orientations (Schulz and Heckhausen, 1999).

To build people's perceptions of control, the curriculum works through three main components: (a) building awareness of personal strengths, (b) training people to differentiate between elements of situations that are inside vs. outside their control, and (c) teaching appropriate strategies to deal with each type of situation.

The first part is focused on building awareness of personal strengths. The session introduces different types of character strengths (e.g. curiosity, creativity, forgiveness), and asks teachers to practice identifying their own strengths and those of others around them. In the second part, covered across two modules, participants learn to distinguish between situations that are inside their control and those that are not. These modules use examples of challenges in everyday life (e.g. financial loss, health issues, balancing personal and professional commitments), and then use guided worksheets that ask participants to reflect on these situations and identify elements that were inside/outside their control. The core message is that, while problems may seem insurmountable at first, individuals can learn to focus on aspects that they can influence. The third part then delves into appropriate strategies to deal with situations that are inside vs. outside one's control. For situations that are *inside* one's control, teachers learn to practice skills like goal-setting, planning, and problem-solving. For situations that are *outside* their control, teachers learn practices such as emotion management and benefit-finding, that can help them manage their response to challenges. The overarching theme that runs through the sessions is that there are *different pathways* out of challenging situations, and individuals can *actively choose* how to emerge out of a challenge – one may choose defeat, go into a survival mode, bounce back to the original state, or at best, emerge stronger and thrive. See Figure 2 for an example of a content worksheet. Appendix C provides additional details on the intervention.

These concepts were illustrated through interactive lectures led by a facilitator, using contextualized examples and narratives, guided reflection activities, and group discussions in each session. In addition, motivational visuals and affirmations were sent over WhatsApp groups to reinforce content each week. Sessions were conducted by the NGO staff. I hypothesize that better psychological skills for dealing with complex situations may affect teacher behavior through a change in self-beliefs about perceived control.

Placebo arm. Teachers in the placebo arm also gathered in their respective Zoom groups at the same time, and received content consisting of psychologically inactive topics not related to personal development. These sessions were led by experts at a local agricultural university and covered informational topics about the environment such as plant growth, soil erosion etc.[15] Appendix C provides additional details on the content of each session for both arms. The presence of an active control group allows me to avoid confounding effects of the treatment with effects of group attendance unrelated to the underlying psychological mechanism, but due to interaction with a group or an external speaker during training.

Additional considerations. To incentivize take-up and attendance in treatment and placebo arms, teachers were mandated by school principals to attend all sessions. Sessions were conducted virtually, and all participants were asked to keep their videos on during the entire

[15]The choice of these topics builds on placebo content used in prior literature (John & Orkin, 2022).

session. Weekly attendance, participation (verbal and chat), and assignment completion were monitored. Every week, top three participants were recognized in each group using a leaderboard format. At the end of the training, one best trainee was awarded in each group. These incentives for attendance were provided to both the treatment and placebo groups.[16]

For the purpose of the training, teachers were divided into 10 groups, with 29-30 teachers in each group. Five of these groups were treatment groups, while the others were placebo. The structure of the sessions was similar across all groups. Each week included a main session (1 hour) on Sunday and a debrief session (1 hour) on a weekday evening. The main session introduced a new concept and involved an instructor-driven lecture with reflection-based activities. To ensure reinforcement of content, a weekly assignment was provided after the main session, which was expected to be completed before the debrief session of that week. During the debrief session, the facilitator led teachers into a group discussion about the assignment response. To prevent the risk of crossovers, treatment and placebo sessions for teachers of a given school were scheduled at the same time, and attendance in each session was monitored by the research team.

To minimize the scope for demand effects, the curriculum in both groups had no discussion about teaching skills or classroom practices. Teachers were informed that this training is solely for their personal development. Neither teachers nor students were informed about classroom observations or the collection of data on school exams later in the school year. A separate research team conducted classroom observations at endline. Having an active control group also minimizes the risk of concerns due to experimenter demand effects and social desirability bias.

3.3 Implementation and Protocols

Math teachers in grades 2, 4, 6, and 8 across study schools were enrolled in the study. I randomized at the teacher level stratified by grade pairings within schools. Pairwise randomization led to increased statistical power within the existing design. First, schools were randomized to all possible combinations of non-adjacent grade pair combinations.[17] Then, grades within a grade-pair were randomized to treatment and control. This process ensured stratified random assignment at the school- and grade-pair-level. I work with non-adjacent grades to minimize the possibility of spillovers in this within-school randomization design.[18]

Given that randomization is at the school-grade (or teacher) level, at least two teachers were needed in every school. This was the school selection criteria used to select study schools out of the full sampling universe of 129 schools. Based on this criteria, the sample consisted of 83 schools out of which 54 schools had 4 teachers, 18 schools had 3 teachers,

[16]Scores were calculated based on attendance, assignment completion, chat, and verbal participation. Chat texts were analyzed using text analysis and teacher-level counts of participation were generated.

[17]There are 3 sets of grade pairings to which 4-teacher schools are randomized: (a) (2, 4) and (6, 8), (b) (2, 6) and (4, 8), and (c) (2,8) and (4, 6). 3-teacher schools are randomized to one of the following three pairings: (a) 2 and (4, 6); (b) 4 and (2, 6); (c) 6 and (2, 4).

[18]Using a network survey for pilot schools, I found that teacher interactions were higher for teachers of adjacent grades, as compared to those from non-adjacent grades. This guided the choice of grades.

and 11 schools had 2 teachers. In each of the study schools, one math teacher per grade was recruited for the study. If multiple teachers were teaching different classrooms of the same grade, one teacher (and classroom) was randomly chosen. In total, the study sample comprises 83 schools, 292 teachers, and 7,570 students who are taught by these teachers.

Spillovers can occur if treated teachers within a school talk to their control colleagues and share ideas from the intervention which impacts beliefs, or alternatively, if students from treated and control classrooms communicate. If the intervention has a positive effect, then any intent-to-treat (ITT) estimate of the treatment effect will be downward biased, in the presence of spillovers. In principle, given the intensive nature of the intervention, the possibility of spillovers through mere verbal interactions is low. Since teacher interactions are most likely to occur across adjacent grades, I attempt to minimize the possibility of spillovers by including only non-adjacent grades in the sample.

Data were collected by an independent field team through four school visits over the academic year – baseline and three rounds of endline through the academic year 2022-23. The full timeline is in Figure 3. Each school visit was conducted by an independent survey team (different from the team that monitored the intervention) and involved teacher surveys, classroom observations, and the collection of student administrative records. Members of the survey team were blind to teachers' treatment status.

3.4 Outcomes and Measures

I pre-registered three primary sets of outcomes: teachers' beliefs, teachers' effort, and student learning, apart from secondary outcomes to better understand mechanisms. The measurement of all outcomes is detailed below.

Teachers' Beliefs. To elicit teachers' beliefs about perceived control, I use (1) self-reported measures including psychological scales, vignettes, and a direct survey question, and (2) a revealed preference measure using an incentive-compatible elicitation mechanism.

(1) Self-Reported Measures. First, I use standard psychological scales for locus of control and self-efficacy. For locus of control, I use an adapted version of Rose and Medway (1981) teacher locus of control scale that captures teachers' perception of control within the classroom environment using 10 items. Each item presents a classroom scenario, and asks respondents to choose between alternative attributions for the outcomes in each scenario. Generalized self-efficacy is measured using the standard Schwarzer and Jerusalem (1995) scale. The scale consists of 10 items that ask respondents to rate their level of agreement with statements about their ability to handle challenges and adverse situations in life. Further details are in Appendix D.1.

Second, to directly elicit beliefs, I use a survey question that asks teachers to rank the relevance of various inputs for education production. This measure is adapted from the China Education Panel Survey (CEPS). Teachers are presented with a list of nine inputs and asked to rank these from most important to least important for student learning. The list includes four student-level inputs (student's talent, effort, family background, prior learning), two teacher-level inputs (teachers' teaching method, attention to students), and three school-level inputs (staff salary, school management, and school's teaching facilities). From these rankings, I construct an indicator for whether a teacher-level factor was ranked as the top most important input.

Third, I present vignettes to teachers that describe a student with a given score in mid-term exams. The question asked teachers to state their expectation of how much the given student's score will increase in response to an additional hour of effort. Variants of the question altered whether the student is high-performing, low-performing, or average.

(2) Revealed Preference Measure. Using self-reports may not always be credible as teachers might have incentives to misreport. To tackle this concern, I devise an incentive-compatible elicitation mechanism using Multiple Price List (MPL) procedure.[19]

I ask teachers to make a sequence of 10 binary choices across contract types to receive a bonus at the end of the year. Teachers choose between a flat bonus and a performance-linked bonus that links payoff to the test score improvement of a randomly chosen below-average student in their class. In each decision round, teachers choose between receiving Rs 1000 as flat bonus and Rs 2000 conditional on the test-score improvement of X points, where X = {0, 2, 4, 6, 8, 10, 12, 14, 16, 50}. Thus, the choices in each round keep the flat-pay option fixed but incrementally increase the stakes of the performance-pay contract with a higher minimum threshold required for the payment to be received. The full experimental task and script is in Appendix B.

I use the switching point as a revealed preference measure for teachers' beliefs about their perceived control over education production. Later switching points in the task indicate a higher level of confidence in teacher abilities to generate student learning. The administration of the MPL task involved two guided examples, an explanation of rules for compensation using a pre-recorded script, and comprehension questions to ensure that teachers understood the task. This was followed by the actual task. Teachers completed the MPL task at baseline, one month, and six months after the intervention.

Given that the real-stakes behavioral task used for belief elicitation presents choices between risky and risk-free options for bonus pay, teachers' decisions may be reflective of their risk preferences instead of self-beliefs about agency and influence over student learning. To explore this channel, I experimentally measure risk attitudes through a second hypothetical price list task. Participants choose between a fixed payment and a lottery that yields 100,000 rupees with probability 0.5 and zero otherwise. The lottery remains the same in all rows while the safe payment increases. Earlier switching points indicate a lower certainty equivalent than later switching points. Details on the task are in Appendix Table D.3. When analyzing treatment effects on teachers' beliefs, I control for baseline risk preferences. I also investigate effects on risk preferences as a secondary outcome.

Teacher Effort. I use a set of pre-registered measures to capture multiple dimensions of teachers' effort across the extensive and intensive margins, observed both inside and outside the classroom.

[19]Multiple price list was chosen over a Becker-DeGroot-Marschak approach (Becker et al., 1964) for relevance to the belief measure in the context and to facilitate ease of understanding.

First, I capture effort at the extensive margin through teacher attendance, data on which was accessed through school administrative records. Second, I capture the intensive margin of effort using (1) classroom observations, and (2) reviews of student homework. During classroom visits, trained observers who were blind to treatment status scored teachers on objective measures of effort. The observers used a standard observation tool which I curated using items from widely-used international tools such as World Bank TEACH, Service Delivery Indicators (SDI) and Stallings instruments, and adapted after extensive piloting. The items capture the quantity and quality of teacher effort in the classroom across multiple domains such as classroom climate (captured by the number of students who raised their hand to ask questions), effort in facilitating engagement (captured by the number of students who were called by name), pedagogical practices (captured by the nature of questions asked: stimulating, recall-related, application-based), among others.[20] The measures were carefully chosen so that they: (i) made it hard for teachers to manipulate on the spot, and (ii) reduced the amount of observer subjectivity in scoring teacher behavior; all measures required a binary or numeric response, and observers were trained to use the tool.

Given that direct classroom observation by surveyors may be subject to Hawthorne effects, I used reviews of homework notebooks as a measure of *past teacher effort*, on the lines of Muralidharan and Singh (2020). In each classroom, homework notebooks of two randomly selected students were reviewed and scored for (a) whether the notebook had been checked, (b) the nature of feedback provided (general/question-specific/none), and (c) whether any encouraging feedback had been provided. To avoid problems associated with multiple hypothesis testing, I generate a single summary index (one for classroom observations, and another for past effort) by calculating inverse-covariance-matrix-weighted averages (following Anderson, 2008). This measure was pre-registered in my pre-analysis plan. I also present decomposed results and report on domain-wise indices for classroom effort and for individual items for homework grading effort.

Additionally, I complement the above measures with teacher-reported measures of time-use that capture time spent on class preparation and grading notebooks apart from other activities. Teachers also reported on whether they engaged in after-school tutoring.

Student Learning. I measure student learning in mathematics after one year, using standardized test scores in mathematics at end-of-year exams. The school chain conducts centralized end-of-year assessments during the school year enabling comparability of student performance across schools. The assessments test for mathematical competence on curriculum taught in the academic year. These are externally graded by teachers from other academies. While the primary outcome is scores on end-of-year Math exams, I also collect data on previous year final Math scores, as well as assessments throughout the year including pre-midterm, midterm, and post-midterm scores. The timeline of these assessments is in Figure 3.

Additionally, I collected data on secondary outcomes including teachers' mental health, growth mindset, risk preferences, and perceptions about the importance of different inputs

[20] A total of 29 items were pre-grouped into six domains: materials and content, effort at facilitating engagement, effort at making the class accessible, pedagogical practices, teacher demeanor, and classroom climate. The full observation tool is in Appendix Table D.2.

for student learning. Details on corresponding measures are in Appendix Section D.3. I also accessed administrative data on school characteristics, teacher background, and student characteristics. To examine the severity of social conformability bias among self-reported measures, I administered the Marlowe-Crowne survey module to all teachers at baseline, which captured respondents' propensity to give socially desirable answers (Crowne and Marlow, 1960).

3.5 Experimental Integrity

Randomization successfully balanced teacher and student observables across treatment and control arms (Tables 5 and 3). Only one covariate, teacher education is statistically significant at 10% level, no more than what would be expected by chance. The null hypothesis of joint significance cannot be rejected (p-value of F-test = 0.165), indicating that there is no evidence to suggest that baseline variables are jointly related to the treatment status. I control for teacher education in all specifications.

Attrition was low and balanced across experimental arms (Table A2). The attrition rate among teachers was 10%, 11% and 13% across the first, second and third endline rounds, respectively. 96% of teachers were surveyed at baseline and at least one endline, while 82% were surveyed across all rounds. Attrition was mainly driven by contractual teachers who left school or those who went on an extended leave, and is typical of the turnover rate in this rural setting.[21]

Attendance tracking for the intervention shows compliance with the experimental assignment. Table A3 shows that 85.3% of teachers attended at least one session, and this rate is balanced by treatment status. The mean attendance rate was 52% indicating that on average, teachers attended five out of 10 training sessions. Figures 5 and 6 show the distributions of total and session-wise attendance, by treatment status. Both groups see a dip in attendance after the first session, however, note that attendance in later sessions is endogenous to the nature of the content. Importantly, attendance in the first session is balanced across both groups (Table A3).

4 Empirical Strategy

I estimate the effects of the intervention on outcomes using the specification below.[22]

$$y_{igjt} = \beta_0 + \beta_1 T_{gj} + \alpha_1 y_{igj0} + X_{gj}\Gamma + \nu_j + \epsilon_{igjt} \quad (1)$$

y_{igjt} is the outcome for student i in grade g and school-grade-pair j at endline t, T_{gj} is an indicator for whether the teacher corresponding to school-grade g was treated, y_{igj0}

[21]Table A4 shows the reasons for attrition. The most common reason was that the teacher left school, followed by long-term absence (on account of a medical leave). In the latter case, some teachers returned in a later endline and were tracked. Those on short-term leave were contacted over phone and administered surveys remotely. Appendix E presents further details regarding correlates of attrition in line with the pre-analysis plan.

[22]The pre-analysis plan outlined three specifications which sequentially added baseline value of outcome and controls selected using the LASSO procedure. In the main tables, I present results using the full specification only for brevity. Appendix tables show the robustness of the findings with alternative specifications.

is the baseline value of the outcome, ν_j is a matched-pair fixed effect, X_{gj} is a a vector of sociodemographic controls selected via double-selection LASSO to maximize precision following Belloni et al. (2014), and ϵ_{igjt} is an idiosyncratic error term.[23] Given multiple endline rounds throughout the year, I pool across endline rounds to present average impacts unless otherwise stated. I estimate an analogous specification for teacher-level outcomes.

The coefficient of interest is β_1 which estimates the intent-to-treat (ITT) effect of the intervention. All specifications with pooled data control for round fixed effects to account for any period-specific changes and time trends. Since treatment is assigned at the teacher (school-grade) level, I cluster standard errors at the teacher level.

To address multiple inference when analyzing effects on primary outcomes, I construct summary indices to reduce dimensionality. Consistent with my pre-analysis plan, I also report adjusted p-values to control for the false discovery rate, using two approaches: the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) and the more conservative Benjamini-Yekutieli step-up method (Benjamini and Yekutieli, 2001). These adjustments account for multiple hypotheses within each table.

5 Results

5.1 Teacher Beliefs

Figure 7 shows teachers' self-reported beliefs at baseline. Teachers display a low locus of control with a right-skewed distribution. 73% of teachers score less than 5 on the 10-point scale, suggesting a predominant tendency to attribute causality to external rather than internal factors. Less than a third of teachers perceive teachers to be the most important inputs for student learning. In contrast, self-efficacy is relatively high, with most teachers rating themselves very favorably on the scale, the average score is 3.5 on a 4-point scale [Table A1]. These patterns suggest that while teachers believe in their ability to exert effort, they perceive a weak link between effort and outcomes.

The multiple price list (MPL) task provides a revealed preference measure of teachers' beliefs about perceived control in influencing student outcomes. Figure 8 shows the distribution of switching points in the MPL task. Consistent with low perceived control, around a quarter of teachers opted for the flat-pay option in the first row. Most teachers exhibited well-behaved preferences, characterized by a unique switching point from performance-linked to flat pay. However, as is typical in experimental tasks, a fifth of teachers made inconsistent choices at baseline — either switching multiple times or unique switches in the reverse direction. This proportion dropped to 8% in subsequent rounds, likely due to improved comprehension in the repeated administration of the task. Table A6 shows the full distribution of responses across rounds. I find that the grade taught by the teacher is correlated with consistency – teachers who teach primary grades (2 and 4) made on average 28% inconsistent choices, whereas teachers who teach secondary grades (6 and 8) made only 14% inconsistent choices at baseline (difference significant at 1% level).

[23]Appendix Table A5 lists the full set of controls that the double LASSO procedure chose from. In line with the pre-analysis plan, baseline value of outcomes and strata fixed effects are not penalized in these regressions.

Other demographic and background variables, including education, experience, gender, permanent (vs. contractual) status, gender, financial situation, and workload, did not correlate with consistency.

I use the switching point as the pre-registered outcome measure to capture beliefs. I define the switching point as the first row in which a teacher chooses the safe option (flat bonus). Note that the construction of this measure takes into account all teachers (including those with multiple switching points). In practice, this is akin to a design where the survey stops after the first switch is made and has been widely used in the experimental literature (Andersen et al., 2006). I also check the robustness of results to using a conservative measure that drops teacher-round observations with inconsistent choices. Notably, the switching point is positively correlated with measures of locus of control and general self-efficacy elicited using standard psychometric scales, suggesting that the choice on the multiple price list captures meaningful variation in the psychological constructs.

Table 4 shows the effect of the intervention on teachers' beliefs. The results are shown for the pooled sample using the main specification in equation (1), sequentially adding risk preferences. I find that the switching point for treated teachers is higher than that for control teachers by 0.45 (p-value = 0.012). Given that the switch points correspond to discrete test-score brackets that quantify the level of teachers' perceived control, I use a second specification that uses the midpoint of the bracket as the dependent variable.[24] Cols (3)-(4) show that the treatment effect in terms of teachers' predicted test-score increase is 1.17 points on the test, which is around 5.5% of a standard deviation for student test scores.[25] Compared to the control group, the treatment teachers exhibit a 15% increase in confidence over their abilities to raise student learning as elicited through their choices in the revealed preference task. Tables A7 and A8 show that these results are robust even without using LASSO controls, with estimates that are similar in magnitude and significance. Columns (4)–(9) present round-wise estimates, which, though constrained by limited sample sizes for each round, consistently show positive effects. Notably, these effects persist six months after the intervention.

In Table A9, I show robustness of main results to restricting the sample to teachers with well-behaved preferences. Given that around 21% teachers exhibited inconsistent preferences at baseline, controlling for baseline value of outcome leads to a drop in the sample moving from col (1) to (2). While the magnitude and significance of the treatment estimate drops in cols. (2) and (3), this is primarily driven by the change in sample size.[26] In cols. (4) and (5), I add flags for missing baseline values and show estimates for the same sample as in column (1). The effect size is 0.42 (p-value < 0.05), and the magnitude remains stable and consistent across specifications.

[24] The midpoints for the ten rows are: -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 50. For never-switchers who always chose flat bonus, the test-score equivalent may lie anywhere in the interval (-100, 0). For those who always chose the performance-linked bonus, it may lie anywhere between (50, 100). On the lower end, -1 would be the next midpoint if the MPL had included a choice between performance-pay conditional on score higher than -2 and Rs 1000. The regression results reported here use -1 as the lowest, and 50 as the highest test-score equivalent. For the penultimate row switchers, 17 would be the midpoint if the price list included a choice between 18 points versus Rs 1000. In line with related work (Dean & Sautmann, 2021), I show the most conservative estimates using the lowest values on the upper end, and also including negative values on the lower end.

[25] End-of-year math scores have a mean of 64 and a standard deviation of 21.

[26] The p-values for the treatment estimate in cols. (2) and (3) are 0.114 and 0.083 respectively.

Given the design of the price list procedure, later switching points by teachers in the multiple price list procedure are reflective of higher confidence in their abilities to influence scores of low-performing students. The above patterns suggest that the intervention led teachers to positively update their beliefs about perceived control in education production.

Additionally, I investigate the effects on self-reported beliefs. Treatment teachers are 6 percentage points more likely to rank teacher inputs as most important for student learning, relative to the control mean of 32%. Treatment teachers score 0.16 SD higher on locus of control (p-value = 0.029) and 0.1 SD (p-value = 0.16) higher on self-efficacy. The effect on self-efficacy is not statistically significant, which is unsurprising given the high baseline levels of self-efficacy among teachers. Overall, treatment teachers score 0.23 SD (p-value = 0.004) higher on an aggregate index that combines the three self-reported measures.

In line with the pre-analysis plan, I also attempted to elicit beliefs using a set of vignette questions that asked teachers to estimate scores for three types of students (low, medium, and high performers) following an additional hour of effort each week. I show in Table A10 that this measure likely suffers from mismeasurement. Specifically, when exploring correlations between belief measures and the first principal component extracted from all measures, the vignette measure exhibits a negative loading, unlike other measures which display positive loadings consistent with a shared latent construct. This suggests that the measure may fail to reliably capture teachers' beliefs about perceived control. Despite these concerns, I present results in Appendix Table A11.

5.2 Teacher Effort

Table 6 presents the results for both extensive and intensive margins of teacher effort. Col (1) shows effects on teacher attendance during the academic year.[27] I find no difference in teacher attendance across treatment and placebo groups. This is not surprising in light of the fact that average attendance rates are already high in this private school setting with mean and median of 0.84 and 0.9 respectively.

Cols (2)-(3) show results for the intensive margin of effort captured through classroom observations by treatment-blind observers for pooled data combined across all survey rounds. Treated teachers score 0.13 standard deviations higher on the summary index of in-class effort compared to control teachers (p = 0.021). The effects are high in the first month after the intervention and persist until 3 months later, with magnitudes ranging between 0.13-0.15 standard deviations, as seen in Table A14.

Table 7 further decomposes the effects on classroom effort by examining pre-defined sub-indices. The psychosocial treatment significantly increased teacher effort in three key areas: facilitating student engagement (0.19 SD, $p < 0.01$), adopting pedagogical best practices (0.15 SD, $p < 0.05$), and making a better use of materials and content (0.12 SD, $p < 0.05$). Effort in facilitating engagement was captured using three measures: the number of students called by name, the number of individual visits to students, and whether teachers actively

[27] Attendance data was accessed from school registers. While most schools maintained hard copy records for attendance, some schools transitioned to a biometrics system of attendance mid-way through the year. Due to technical glitches, biometric attendance data of 15 teachers could not be retrieved.

attempted to engage backbenchers. Pedagogical best practices were assessed through indicators related to types of questions asked in class (e.g., related to recall, application, or encouraging creativity) and curricular techniques, such as starting lessons with stimulating questions and concluding with summaries. Effective use of materials and content was evaluated using indicators for blackboard, textbook, and teaching aide utilization, as well as numeric measures like the number of in-class questions solved, practice questions assigned, and whether homework was given.

Table 6, column 3 shows results for teachers' grading effort as reflected in the quality of feedback provided in homework notebooks of students in the teachers' classrooms. Treatment teachers score 0.11 standard deviations higher on a summary index of grading effort compared to control teachers, and the effect is significant at 5% level of significance (p-value = 0.012). Further, Table 8 shows that the effects are driven by the provision of detailed question-level feedback by teachers in the treatment group. Treatment teachers are 6 percentage points more likely to provide detailed as opposed to generic feedback on student work. While the classroom observations may be subject to Hawthorne effects where teachers (across both treatment and control) are likely to portray their best behavior in the presence of an external observer, this alternative measure for effort at the intensive margin using grading practices offers a relatively clean and undistorted measure of past effort. One concern is that given that the measure used information from only two randomly sampled students per teacher, this could be a noisy. However, aggregating individual measures into an index reduces noise as random errors that are uncorrelated across indicators are likely to cancel each other.

In addition to the above measures collected through direct physical observation of teacher activity and records, I collect self-reports of teacher time-use through teacher surveys at all data collection rounds. Table 6, column (4) shows effects on self-reported minutes spent on teaching-related activities, on an average day. While there are no differences in average time spent on class preparation, treatment teachers spend an average of 8.4 minutes more on checking notebooks compared to control teachers, which translates to an increase of 10 percentage points (p-value < 0.01). This result is consistent with previous results on the provision of more detailed feedback by treatment teachers, as recorded by independent observers. I also find that treated teachers are 10 percentage points (p-value < 0.05) more likely to engage in after-school tutoring. While I do not have information on who attends these after-school sessions, it is not uncommon for teachers to also be tutors to their own students after school hours (Jayachandran, 2014), in which case, this might be a potential alternative channel through which teachers are exerting additional effort outside classroom.

Overall, these results show that the treatment impacts teacher behavior and raises effort conditional on attendance. Another salient takeaway from the above findings is that teachers are aware of different dimensions of effort and ways to target student learning. Since the intervention did not allude to classroom or pedagogical practices, the observed increase in effort is unlikely to stem from the acquisition of any new pedagogical skills.

5.3 Student Learning

The student dataset is comprised of 318 classrooms taught by 292 teachers in the study. 91% of teachers teach one classroom, while 9% of teachers teach two classrooms either in the same grade or another study grade. At the beginning of the academic year, I collected student information for 7,570 students who were taught by study teachers. At the end of the study, 7,425 students appeared for the end-of-year exam. 21 teachers left during the academic year, so their classrooms were assigned to an alternative teacher in the school.

To estimate the impact on student learning, I implement equation (1) by regressing end-of-year scores on treatment indicator, lagged test scores and strata fixed effects, apart from double-lasso selected controls. Given that the intervention occurred into the school year, both pre-midterm and previous year final scores are available as measures of lagged scores. The previous year scores are externally graded, however, these are available for 87% of the sample. In contrast, the pre-midterm scores are internally graded; these are available for 99% of students. I show results using pre-midterm scores since they are available for a larger sample. Figure 10 shows that the distributions are similar.

Table 9 shows the results for student learning. Students taught by treatment teachers score 0.09 SD (p-value = 0.042) higher than those taught by control teachers at the end-of-year Math exams. Accounting for teacher education which was imbalanced at baseline increases the magnitude of the estimate to 0.1 SD (p-value = 0.028). Col (3) shows that the estimate remains stable after accounting for double-lasso selected controls. The analysis is restricted to students whose teachers had not attrited before the last endline. For classrooms that experienced teacher attrition, an alternative teacher was assigned to teach the students.[28] In Table A15, I show results for the full sample of students including the classrooms whose teachers had attrited midway through the year. Estimates are slightly reduced in magnitude ranging between 0.06-0.08 SD (p-value = 0.068), and can be interpreted as a lower bound on the treatment effect.

An important concern is whether performance-linked bonuses induced perverse incentives for teachers to increase scores for their students. However, this concern is largely addressed by the nature of the examination system. All end-of-year exams are centrally administered with external proctoring and grading. Further, since incentives are offered to both groups, it is unlikely that one group would be systematically more likely to distort scores compared to the other.

Given the nature of teacher incentives offered to both groups as part of the belief-elicitation task, teachers stand to gain from performance improvement for students in the bottom-half of the distribution only, and not the top-half. One concern is that this may divert effort away from better-performing students. To test if this is the case, I examine heterogeneity in treatment effects by baseline ability. I find no evidence of differential treatment effects by student baseline performance (Table 10, column 3).

Figure 9 plots nonparametric treatment effects by percentile of baseline score. We see that while gains are larger for the lower tail, these are also present for the top tail, and the

[28] The alternative teacher could be a study teacher with the same or different treatment status, or a non-study teacher from the same school.

null of equality cannot be rejected. This indicates that the learning gains are broad-based. Even though teacher incentives are designed to reward gains for the bottom-half of the student distribution and the bottom-half is better off, the top-half is not worse-off. This indicates that the intervention led to a Pareto improvement compared to the status quo. Additionally, I do not find any heterogeneity in treatment effects by student characteristics including gender, socioeconomic status, and grade (Table 10).

Interestingly, I find that the effects on student learning are driven by students who are taught by teachers with low levels of baseline beliefs about perceived control (p-value for the interaction coefficient = 0.05, Table 11). While estimates on interaction term are not statistically significant for other teacher covariates including baseline experience, effort and value-added, these are negative suggesting that gains are driven by students of teachers with lower levels of effort, ability, and experience to begin with.

6 Mechanisms and Discussion

In this section, I provide evidence that beliefs about influence over student learning are the most plausible mechanism for the observed effects of the intervention on teacher effort and student learning. Given the content of the intervention, it is possible that the psychosocial training impacted effort and learning through other channels. I examine the effects on pre-registered secondary outcomes, including teachers' growth mindset, mental health, and risk preferences, and rule out that the effects operated through these channels. In addition, I also rule out some alternative non-psychological mechanisms.

Growth Mindset. Growth mindset is a related construct that captures beliefs about the malleability of intelligence. Teachers were administered Implicit Theories of Intelligence scale (Blackwell et al., 2007) after the intervention, and six months later in the last endline. Col. 1 of Table 12 shows results for pooled data combining both endline rounds. The treatment had no discernible impact on teachers' growth mindset.

Mental Health. To explore the possibility that the intervention affected teachers' mental health, I administered the standardized Center for Epidemiological Studies - Depression (CES-D) Scale (Radloff, 1977) used in the literature.[29] I reverse code the scale so that higher scores better mental health; in this case, the absence of depression symptoms. The intervention had a positive but not statistically significant effect of 0.08 standard deviations on symptoms of depression (Table 12, col. 2). This is not surprising as the workshop – while facilitating debrief sessions and forums for group discussion – did not specifically target depression or had any key elements of a group therapy session.

[29] The scale consists of 20 items that ask respondents to indicate the frequency with which they have experienced symptoms related to stress, burnout, and depression, including restless sleep, poor appetite, feeling lonely, etc.

Risk Preferences. I test whether the intervention influences teachers' risk preferences by administering a multiple price list with ten hypothetical choices between a safe payment and a lottery. I calculate the certainty equivalent using the switching point in the price list, and construct an indicator for risk aversion that takes the value one if the certainty equivalent is less than the expected value of the lottery. I find that the intervention had no discernible impact on teachers' risk aversion (Table 12, col. 3).

In contrast, results from multiple measures of beliefs (across a range of meausures including both revealed preference and self-reported responses on pscyhometric scales) provide suggestive evidence that the intervention impacted teacher effort and student learning by influencing teachers' perceptions about relative roles of inputs in education production and spurring them into action.

6.1 Other concerns

Experimenter Demand. It is possible that teachers reported answers that they believed the surveyors wanted to hear, leading to a concern about inducing experimenter demand effects. However, this is unlikely in my setting due to three reasons. First, all primary outcomes had objective measures as opposed to self-reports – beliefs measure is an incentivized real-stakes task, effort is captured using classroom observations by treatment-blind observers, and student learning is captured using externally graded exam scores. Second, the intervention did not contain any references to classroom-related practices or pedagogy and was imparted as a broad-based training for personal growth of teachers in all areas of life. More specifically, teachers were not aware that their effort would be measured after the intervention and linked to the training. An independent field team conducted classroom observations, different from the team that conducted the intervention.

Group Effects. The existence of a placebo group implies that interaction with a group or a speaker should not drive the effects of the treatment. The structure of the training (including format, duration, timings, and group size) was the same for both groups. The treatment effect is unlikely to be driven by anything else apart from the training content.

Spillovers. While the randomization design took account of the possibility of spillovers by including only non-adjacent grades in the study, it is possible that the study teachers across treatment status within a school communicated about the intervention. At the outset, given the medium-touch nature of the intervention with debrief sessions, and activities to reinforce content, the motivational effect is hard to diffuse across teachers through communication and interaction. While it is hard to test and rule out the possibility completely, the expected direction of spillovers is positive, due to which any ITT estimate of the treatment effect is likely to be downward biased. In this case, the observed effect can be interpreted as a lower bound for the true effect.

7 Conclusion

This paper presents the first evidence on estimating the impact of removing psychosocial barriers for teachers in developing countries on the quality of teacher effort and student learning. I document empirical support for the idea that teachers in resource-constrained environments perceive a weak mapping between their effort and student learning. I present evidence that these beliefs about teachers' perceived control over education production are malleable and can be improved through training. Further, I show that an intervention that shifts these beliefs also leads to marked improvements in teacher effort and productivity. Teachers learn how to navigate complex learning environments and direct their efforts on inputs that translate into student learning. These results show that alleviating psychosocial constraints for teachers improves student learning.

The magnitude of the effects on student learning is substantial, especially given the scale and low-cost nature of the intervention: learning gains of 0.09 standard deviations at a cost of $8 per teacher and $0.32 per student over a period of six months. Given the online mode of instruction, the costs primarily cover instructor time and the cost of Zoom licenses for delivery of online sessions. Other teacher-level or pedagogical interventions achieve gains of this size either over a longer duration or by incurring a substantially higher cost per pupil. For example, additional school inputs (cash grant and contract teacher) in India raised student learning by 0.08 SD over two years at a cost of $1.56 per pupil, while introducing teacher performance-pay raised student learning by 0.27 SD over two years at a cost of $1.25 per pupil (Muralidharan and Sundararaman, 2011). Mbiti et al. (2019) show that a combination of school inputs and teacher incentives raised student learning by 0.23 SD over two years at a cost of $11.56 per student. In contrast, the gains in this study were achieved over a duration of six months, with a total of ten hours of teacher training. The intervention demonstrates a particularly low-cost way of inducing higher effort into teaching and learning gains in the natural environment of the classroom.

Additionally, while alternative teacher-focused interventions using performance pay and incentives have been shown to be effective, these interventions are not always economically and politically feasible to scale. Given the low costs of training and the potential for the curriculum to be embedded in regular teacher professional development, these results are promising from a policy perspective, especially given the fact that developing countries engage in massive yearly spending to invest in the capacity-building of teachers through in-service training.

An important limitation of the study is that this was done in a private as opposed to a public school setting. While the socioeconomic demographics of students and teachers in this rural, private setting are comparable with those of public schools, an important difference is the pay for teachers. Public school teachers earn higher salaries, which impacts the selection margin by affecting the composition of teachers. It is possible that teachers across public and private settings differ in terms of their baseline motivation. It remains to be seen if these findings also replicate in government-run schools, though the low levels of baseline effort suggest higher scope for the impact of such an intervention.

References

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2023). Perceived returns to job search. *Labour Economics*, 80, 102307.

Adhvaryu, A., Kala, N., & Nyshadham, A. (2023). Returns to on-the-job soft skills training. *Journal of Political Economy*, 131(8), 000-000.

Alan, S., Boneva, T., Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3), 1121-1162.

Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9, 383-405.

Anderson, M.L., (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103, 1481–1495.

Ashraf, N., Bryan, G., Delfino, A., Holmes, E., Iacovone, L., & Pople, A. (2022). Learning to see the world's opportunities: The impact of imagery on entrepreneurial success. London School of Economics & Political Science, Working Paper.

Attanasio, O., F. Cunha, and P. Jervis (2019). Subjective parental beliefs. their measurement and role. Technical report, National Bureau of Economic Research.

Attanasio, O., Boneva, T., & Rauh, C. (2022). Parental beliefs about returns to different types of investments in school children. *Journal of Human Resources*, 57(6), 1789-1825.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.

Baranov, V., Bhalotra, S., Biroli, P., Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review*, 110(3), 824-859.

Becker, G.M., DeGroot, M.H., Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9 (3), 226–232.

Belloni, A., Chernozhukov, V., Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29-50.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1), 289-300.

Benjamini, Y., & D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29: 1165–1188.

Béteille, Tara; Evans, David K. 2021. Successful Teachers, Successful Students: Recruiting and Supporting Society's Most Crucial Profession. World Bank, Washington, DC.

Bhalotra, S. R., A. Delavande, P. Gilabert, and J. Maselko (2020). Maternal investments in children: The role of expected effort and returns. IZA Discussion Paper.

Bhat, B., De Quidt, J., Haushofer, J., Patel, V. H., Rao, G., Schilbach, F., & Vautrey, P. L. P. (2022). The long-run effects of psychotherapy on depression, beliefs, and economic outcomes (No. w30011). National Bureau of Economic Research.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246-263.

Caliendo, M., Kritikos, A. S., Rodriguez, D., & Stier, C. (2023). Self-efficacy and entrepreneurial performance of start-ups.*Small Business Economics*, 61(3), 1027-1051.

Campos, F., M. Frese, M. Goldstein, et al. 2017. "Teaching Personal Initiative Beats Traditional Training in Boosting Small Business in West Africa." *Science* 357 (6357): 1287–90.

Carneiro, P. M., E. Galasso, I. X. Lopez Garcia, P. Bedregal, and M. Cordero (2019). Parental beliefs, investments, and child development: Evidence from a large-scale experiment. World Bank Policy Research Working Paper (8743).

Cobb-Clark, D. A. (2015). Locus of control and the labor market. *IZA Journal of Labor Economics*, 4(1), 1-19.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349.

Dean, M., & Sautmann, A. (2021). Credit constraints and the measurement of time preferences. *Review of Economics and Statistics*, 103(1), 119-135.

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

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4), 1241-78.

Ersoy, F. (2023). Effects of perceived productivity on study effort: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, 207, 376-391.

Frazier, P., Meredith, L., Greer, C., Paulsen, J. A., Howard, K., Dietz, L. R., & Qin, K. (2015). Randomized controlled trial evaluating the effectiveness of a web-based (p.197) stress management program among community college students. *Anxiety, Stress, & Coping*, 28, 576–586.

Ganimian, A. J. (2020). Growth-mindset interventions at scale: Experimental evidence from Argentina. *Educational Evaluation and Policy Analysis*, 42(3), 417-438.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569-582.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3), 205-227.

Gray-Lobe, G., Keats, A., Kremer, M., Mbiti, I., & Ozier, O. W. (2022). Can education be standardized? Evidence from Kenya. Evidence from Kenya (September 16, 2022). University of Chicago, *Becker Friedman Institute for Economics Working Paper*, (2022-68).

Groh, M., N. Krishnan, D. J. McKenzie, and T. Vishwanath. (2012). "Soft Skills or Hard Cash? The Impact of Training and Wage Subsidy Programs on Female Youth Employment in Jordan." Policy Res. Working Paper no. 6141, World Bank, Washington, DC.

Hintz, S., Frazier, P. A., & Meredith, L. (2015). Evaluating an online stress management intervention for college students. *Journal of Counseling Psychology*, 62(2), 137.

Hiroto, D. S., & Seligman, M. E. P. (1975). Generality of learned helplessness in man. *Journal of Personality and Social Psychology*, 31(2), 311-327.

Jayachandran, S. (2014). Incentives to teach badly: After-school tutoring in developing countries. *Journal of Development Economic*s, 108, 190-205.

John, A., & Orkin, K. (2022). Can simple psychological interventions increase preventive health investment?. *Journal of the European Economic Association*, 20(3), 1001-1047.

Lachman, M., Neupert, S. D., & Agrigoroaei, S. (2011). The relevance of control beliefs for health and aging. In K. W. Schaie S. L. Willis (Eds.), Handbook of the psychology of aging (7th ed., pp. 175–190). San Diego, CA: Academic Press

Langer, E. J., & Rodin, J. (1976). The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology*, 34(2), 191–198.

Leaver, C., Ozier, O., Serneels, P., & Zeitlin, A. (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. *American Economic Review*, 111(7), 2213-2246.

List, J. A., Pernaudet, J., & Suskind, D. L. (2021). Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nature Communications*, 12(1), 5765.

Loyalka, P., Popova, A., Li, G., & Shi, Z. (2019). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics*, 11(3), 128-54.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania. *The Quarterly Journal of Economics*, 134(3), 1627-1673.

McKelway, M. (2021) Women's Employment in India: Intra-Household and Intra-Personal Constraints. [Job Market Paper]. Available at: `https://drive.google.com/file/d/1Fi8Y8A0o4j-NwqOmROowAV0kLcuCC1Ai/view`

McKenzie, D., Mohpal, A., & Yang, D. (2021). Aspirations and Financial Decisions: Experimental Evidence from the Philippines (No. w28607). National Bureau of Economic Research.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39-77.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4), 1426-60.

Nourani, V., Ashraf, N., & Banerjee, A. (2023). Learning to teach by learning to learn. University of Chicago and Makerere University.

Popova A, Evans D, Breeding Mary E, Arancibia V. (2018) Teacher Professional Development around the world: The gap between evidence and practice. World Bank Policy Research Working Paper 8572

Orkin, K., Garlick, R., Mahmud, M., Sedlmayr, R., Haushofer, J., & Dercon, S. (2023). Aspiring to a better future: can a simple psychological intervention reduce poverty? (No. w31735). National Bureau of Economic Research.

Pritchett, L. (2013). The rebirth of education: Schooling ain't learning. CGD Books.

Radloff, L. S. (1977). The CES-D scale: A self report depression scale for research in the general population. *Applied Psychological Measurements*, 1, 385-401.

Reich, J. W., & Infurna, F. J. (Eds.). (2017). Perceived control: Theory, research, and practice in the first 50 years. Oxford University Press.

Riley, E. (2024). Role models in movies: The impact of Queen of Katwe on students' educational attainment. Review of Economics and Statistics, 106(2), 334-351.

Rose, J. S., & Medway, F. J. (1981). Teacher locus of control, teacher behaviour, and student behaviour as determinants of student achievement. *Journal of Educational Research*, 6, 375-381.

Rury, D., & Carrell, S. E. (2023). Knowing what it takes: The effect of information about returns to studying on study effort and achievement. *Economics of Education Review*, 94, 102400.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80, 1–28.

Sabarwal, S., & Abu-Jawdeh, M. (2018). What teachers believe: mental models about accountability, absenteeism, and student learning. World Bank Policy Research Working Paper, (8454).

Sabarwal, S., Abu-Jawdeh, M., & Kapoor, R. (2022). Teacher beliefs: Why they matter and what they are. *The World Bank Research Observer*, 37(1), 73-106.

Schulz, R., & Heckhausen, J. (1999). Aging, culture, and control: Setting a new research agenda. Journal of Gerontology: Psychological Sciences, 54(3), 139-145.

Schwarzer, R, and Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs. , ed. John Weinman, Stephen C. Wright and Marie Johnston, 357. Windsor, England: NFER-NELSON.

Sherrod, D. R., Hage, J. N., Halpern, P. L., & Moore, B. S. (1977). Effects of personal causation and perceived control on responses to an aversive environment: The more control, the better. *Journal of Experimental Social Psychology*, 13(1), 14-27.

Skinner, E.A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology* 71: 549–570.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., & Eyers, J. (2015). Interventions for improving learning outcomes and access to education in low-and middle-income countries: a systematic review. 3ie Systematic Review, 24.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202-248.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573.

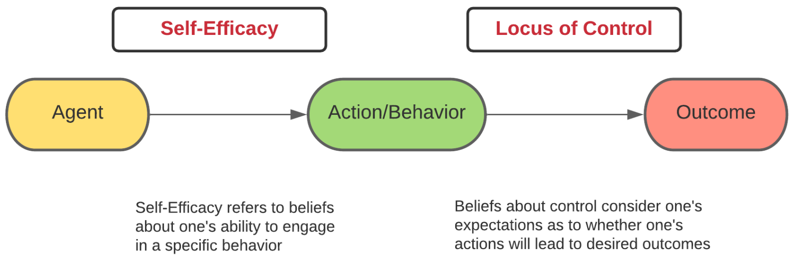
World Bank (2018). World Development Report 2018 : Learning to Realize Education's Promise. Washington, DC: *World Bank*

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Woody, S., Crosnoe, R., ... & Dweck, C. S. (2022). Teacher mindsets help explain where a growth-mindset intervention does and doesn't work. Psychological Science, 33(1), 18-32.

Zautra, A. J., Davis, M. C., Reich, J. W., Sturgeon, J. A., Arewasikporn, A., & Tennen, H. (2012). Phone-based interventions with automated mindfulness and mastery messages improve the daily functioning for depressed middle-aged community residents. *Journal of Psychotherapy Integration*, 22(3), 206.

Figures

Figure 1: *Self-Efficacy vs. Locus of Control*

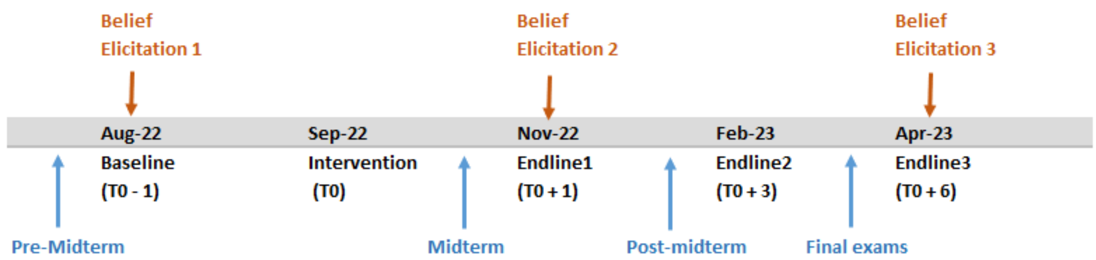


Notes: This figure illustrates the conceptual relationship between different constructs of control as highlighted by Skinner (1996). Self-efficacy links agents to means in the first part of the sequence, whereas locus of control concerns beliefs connecting means to ends. Perceived control is an interaction between self-efficacy and locus of control beliefs, reflecting beliefs about people's ability to influence outcomes.

Figure 2: *Example of Intervention Activity*

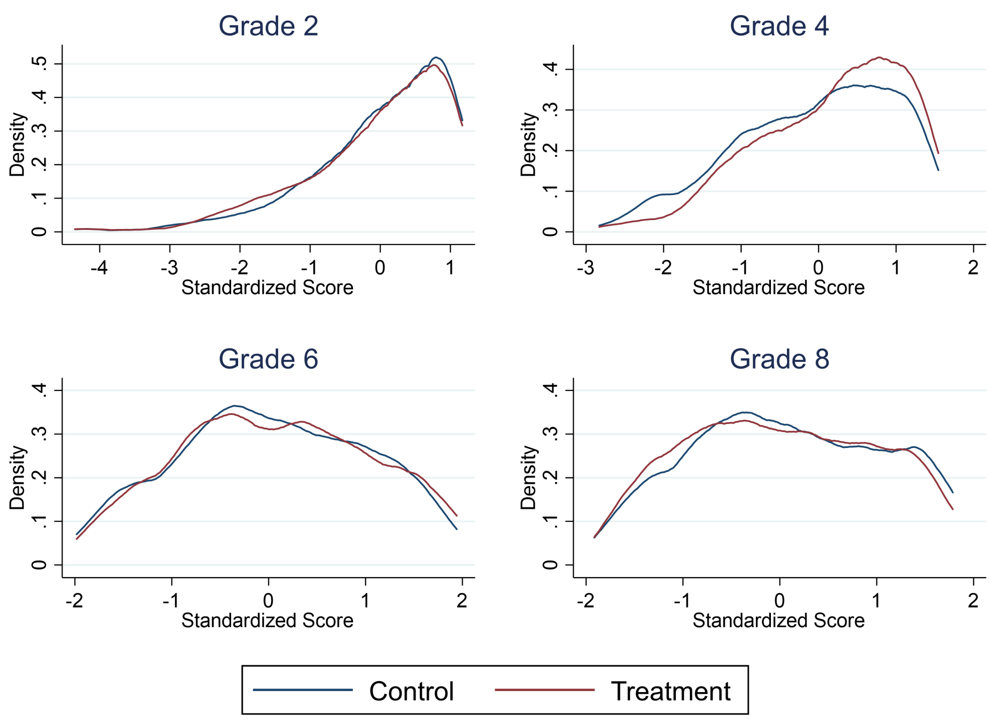
Notes: This figure shows an example of an activity for teachers presented as a guided worksheet to teach them to identify parts of problems that were inside vs. outside their control.

Figure 3: *Timeline of Data Collection*



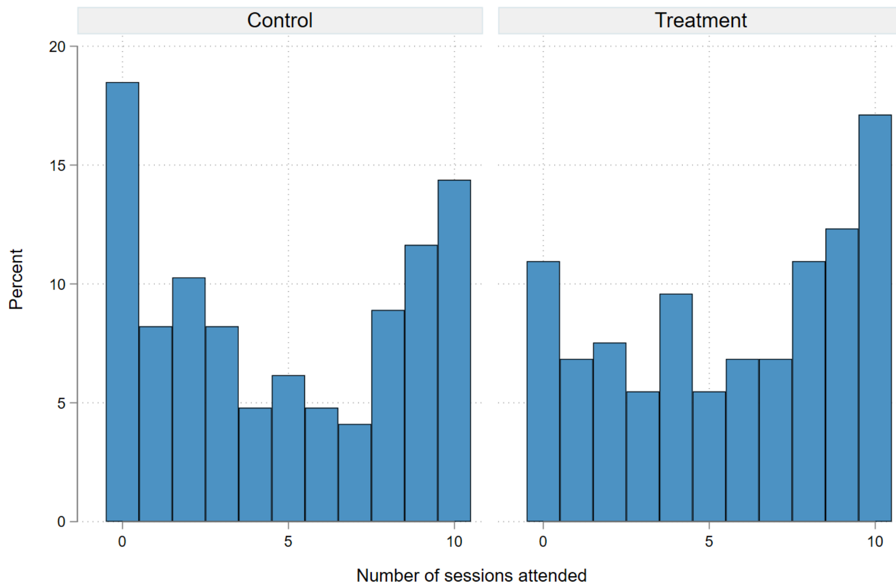
Notes: This figure details the timeline for the baseline data collection, the intervention, the endline data collection rounds, as well as the timeline for school-level examinations for the academic year 2022-23.

Figure 4: *Distribution of student test-scores at baseline (pre-midterm examinations)*



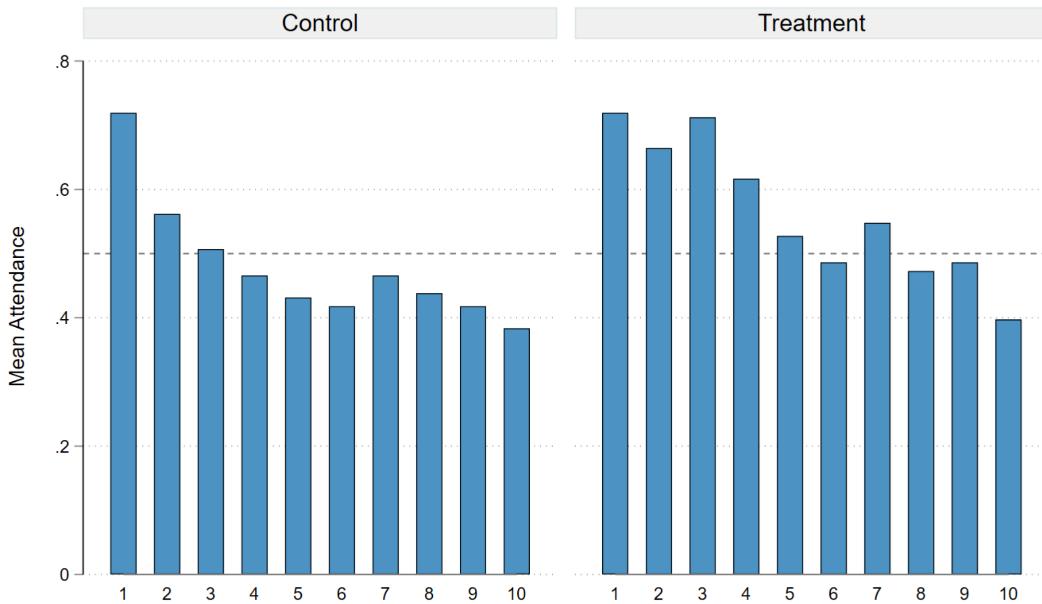
Notes: This figure shows the distribution of student test-scores for grades 2, 4, 6, and 8 in mathematics from administrative data on pre-midterm examinations across schools. These are centralized examinations across all academies. The examinations are externally graded by teachers from academies in the same cluster. Each cluster consists of 4-6 academies. The distributions are skewed for lower grades but roughly normal for higher grades.

Figure 5: *Distribution of Total Attendance by Treatment Status*



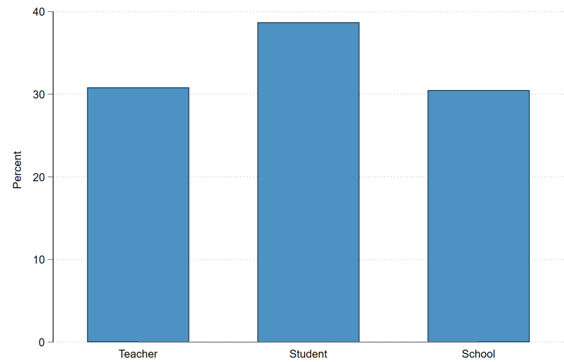
Notes: This figure presents the distribution of the number of sessions attended by teachers by treatment status. A total of 10 training sessions were conducted. Each bar in the figure represents the percentage of teachers of a given treatment status who attended a given number of total meetings. The p-value for Kolmogorov-Smirnov test for equality of distribution functions is 0.098.

Figure 6: *Session-wise Attendance by Treatment Status*

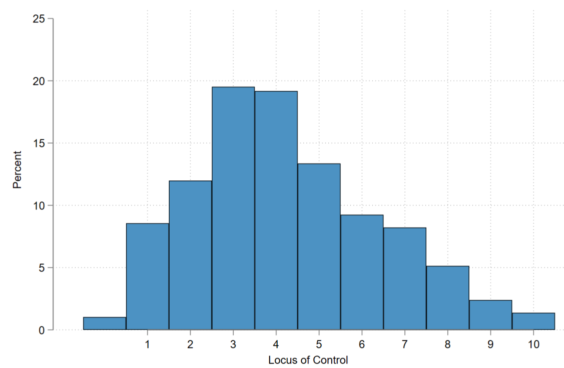


Notes: This figure shows the average attendance at each training session by treatment status. Each bar in the figure represents the percentage of teachers of a given treatment status who were present for a given session.

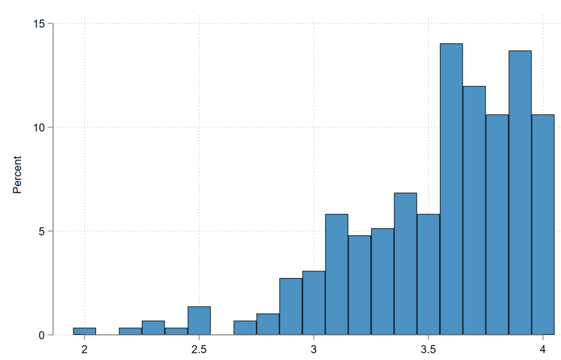
Figure 7: *Teachers' Beliefs at Baseline*



(a) *Top ranked input for student learning*



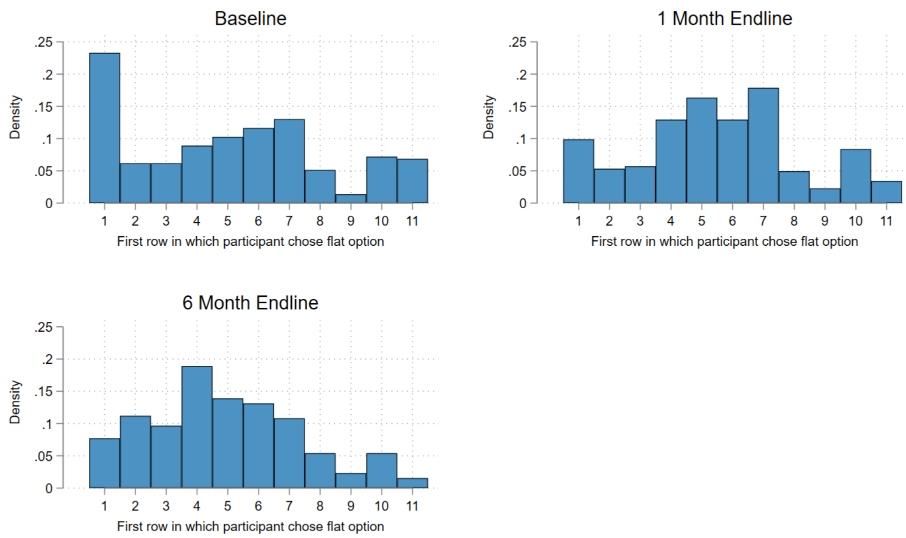
(b) *Locus of Control*



(c) *Generalized Self-Efficacy*

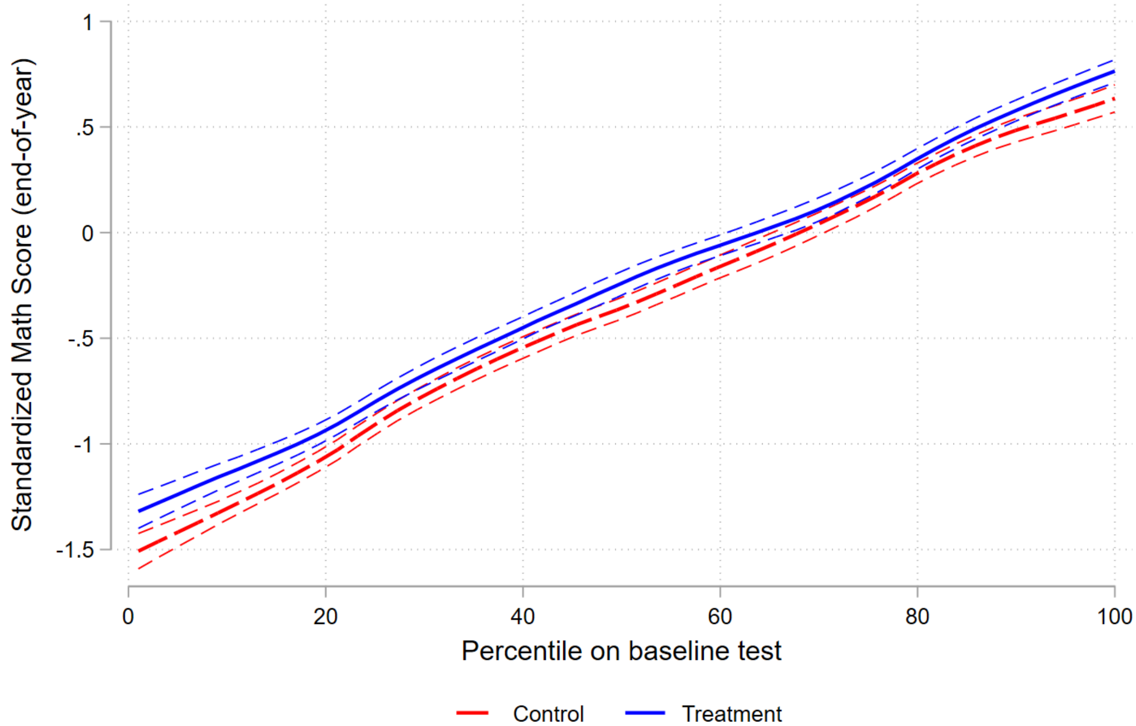
Notes: The figures show the distribution of teachers' self-reported beliefs at baseline. The top panel shows the distribution of teachers' rop-ranked input for learning out of a list of nine inputs, presenting in a ranking task. The bottom two panels show distributions of locus of control and self-efficacy beliefs captured by standard psychometric scales, that is, Rose and Medway (1981) and Schwarzer and Jerusalem (1991) respectively.

Figure 8: *Distribution of Teachers' Beliefs elicited through MPL (Switching Point)*



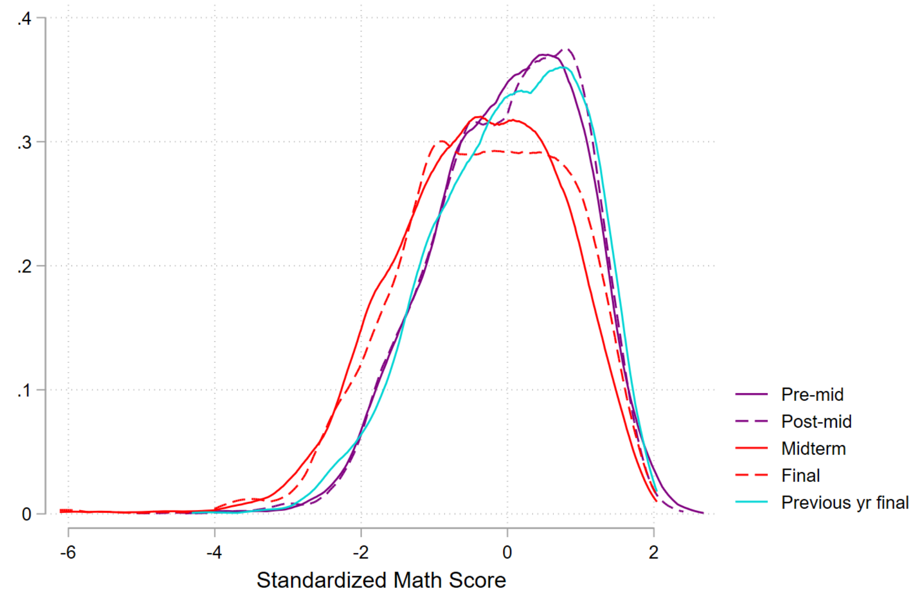
Notes: The figures show the distribution of switching points in the multiple price list belief elicitation task which has 10 rows with binary choices between a flat and a performance-linked bonus. Switching point is defined as the first row in which a teacher chooses the safe option (flat bonus). The horizontal axis shows the first row in which the participant chose the flat pay option. If the flat pay option was never chosen, the switch point is coded as 11. The number of observations at baseline, first and last endline are 292, 263, and 259 respectively.

Figure 9: *Non-Parametric Treatment Effects By Baseline Percentiles*



Notes: The figure shows kernel-weighted local mean smoothed plots relating endline test scores to percentiles in baseline achievement, alongside 95 percent confidence intervals, for treatment and control groups.

Figure 10: *Density of Student Test-Scores Across Assessments*



Notes: The figure shows kernel density plots of student test scores across school-administered assessments. Mid-term, final, and previous Year scores are standardized within grade to have a mean of zero and standard deviation of one for the control group at baseline. Pre-mid term and post-mid term scores are standardized within a school cluster x grade.

Tables

Table 1: *Distribution of teachers' responses to belief statements*

	Strongly Disagree	Disagree	Agree	Strongly Agree
The amount a student can learn is primarily related to family background.				
India	4.46	43.87	49.44	2.23
Ethiopia	4.30	14.45	59.38	21.88
I am very limited in what I can achieve because a student's home environment is a large influence on his/her achievement				
India	2.59	34.81	60.01	2.59
Ethiopia	3.91	32.81	45.70	17.58
Even a teacher with good teaching abilities may not make a difference for many students.				
India	14.10	46.56	35.08	3.93
Ethiopia	10.08	34.11	39.53	16.28

Notes: Data from Young Lives school survey (India and Ethiopia, 2016-17). The surveys covered 281 teachers across 205 schools in India, and covered 271 teachers across 63 schools in Ethiopia. The schools were chosen to be representative of Young Lives sites in each country. The teacher questionnaire included a belief module that asked teachers to indicate their level of agreement with various statements on a four-point likert scale. Young Lives is a study of child poverty in India, Peru, Ethiopia and Vietnam.

Table 2: *Baseline Balance on Teacher Observables*

	Mean/SE (Placebo)	Mean/SE (Treat)	Diff in means (Placebo - Treat)	N (Placebo)	N (Treat)
Age	32.897	33.486	-0.589	146	146
	[0.559]	[0.586]			
Years of Experience	7.178	7.740	-0.562	146	146
	[0.401]	[0.475]			
Female	0.788	0.849	-0.061	146	146
	[0.034]	[0.030]			
Education Level: Bachelors	0.233	0.370	-0.137**	146	146
	[0.035]	[0.040]			
Education Level: Masters and above	0.753	0.623	0.13**	146	146
	[0.036]	[0.040]			
Teacher training degree/diploma	0.863	0.904	-0.041	146	146
	[0.029]	[0.024]			
Permanent	0.548	0.610	-0.062	146	146
	[0.041]	[0.041]			
Number of periods taught per day	6.760	6.815	-0.055	146	146
	[0.119]	[0.125]			
Number of periods taught per week	40.836	39.432	1.404	146	146
	[0.800]	[0.915]			
Teaches subjects other than Math	0.699	0.651	0.048	146	146
	[0.038]	[0.040]			
Class size	25.890	25.651	0.239	146	146
	[0.641]	[0.639]			
Grade 2	0.240	0.281	-0.041	146	146
	[0.035]	[0.037]			
Grade 4	0.260	0.240	0.02	146	146
	[0.036]	[0.035]			
Grade 6	0.199	0.260	-0.061	146	146
	[0.033]	[0.036]			
Grade 8	0.301	0.219	0.082	146	146
	[0.038]	[0.034]			
F-test of joint significance (*p*-value)					0.165
F-test, number of observations					292

Notes: This table presents summary statistics for baseline covariates at the teacher-level by assignment and tests whether there is any statistically significant difference between experimental arms at baseline. Standard deviations are in brackets. Data was collected during the teacher survey at baseline. The *p*-value for the F-test of joint orthogonality is shown at the bottom of the table. * p<0.10, ** p<0.05, *** p<0.01..

Table 3: *Baseline Balance on Student Observables*

	Mean/SE (Placebo)	Mean/SE (Treat)	Diff in means (Placebo - Treat)	N (Placebo)	N (Treat)
Female	0.463	0.468	-0.006	3641	3697
	(0.008)	(0.008)			
Grade	5.082	5.000	0.082	3748	3822
	(0.037)	(0.036)			
Receive financial assistance	0.110	0.100	0.010	3748	3822
	(0.005)	(0.005)			
Baseline Math Scores					
Pooled previous year score	0.002	0.032	-0.030	3323	3295
	(0.017)	(0.018)			
Grade 2	0.000	-0.032	0.032	721	812
	(0.037)	(0.035)			
Grade 4	-0.000	0.080	-0.080	862	771
	(0.034)	(0.038)			
Grade 6	-0.000	0.017	-0.017	773	845
	(0.036)	(0.036)			
Grade 8	0.007	0.065	-0.057	967	867
	(0.032)	(0.034)			
Pooled pre-mid score	0.001	0.001	-0.001	3702	3771
	(0.016)	(0.016)			
Grade 2	0.017	-0.015	0.033	858	953
	(0.035)	(0.031)			
Grade 4	-0.011	0.012	-0.023	965	913
	(0.032)	(0.033)			
Grade 6	-0.009	0.008	-0.017	868	960
	(0.034)	(0.032)			
Grade 8	0.007	0.001	0.006	1011	945
	(0.031)	(0.032)			
F-test of joint significance (*p*-value)					0.21
F-test, number of observations					7,246

Notes: This table presents summary statistics for baseline covariates at the student-level by assignment and tests whether there is any statistically significant difference between experimental arms at baseline. Student baseline ability is proxied by average math scores in final exams of previous grade (AY 2021-22), which are available for 87% of the sample. These are centralized examinations across all academies. Scores are standardized at the grade level, with a mean of zero and standard deviation of one for the control group at baseline. Student gender was determined based on their names. For around 3% of the sample, the student name was gender-neutral, so the variable takes a missing value for these students. Administrative records indicate whether a student receives need-based financial assistance from these school. The bottom of the table reports the *p*-value from F-test regressing indicator for assignment on gender, grade, indicator for financial assistance and standardized pre-midterm math scores, accounting for strata fixed effects. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: *Treatment Effect on Teachers' Beliefs (Revealed Preference Measure)*

	Switching Point		Test-score equivalent	
	(1)	(2)	(3)	(4)
Treat	0.600***	0.447**	1.538***	1.167**
	(0.175)	(0.177)	(0.521)	(0.517)
p-value	0.001	0.012	0.003	0.024
Benjamini Hochberg q-value	0.002	0.023	0.010	0.024
Yekutieli p-value	0.005	0.032	0.013	0.050
Risk Preferences	No	Yes	No	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes
Baseline Value of Outcome	Yes	Yes	Yes	Yes
Control Mean	5.10	5.10	7.81	7.81
Observations	522	522	522	522

Notes: The dependent variable in columns (1) and (2) is a revealed preference measure of teachers' perceived control captured by the row in which the teacher first switched from the performance-linked option to the flat pay option in the multiple price list task and ranges from 1 to 11. The dependent variable in columns (3) and (4) is the same measure in units of predicted test-score increase over a year, and corresponds to the mid-point of test-score bracket mapped to each row in the multiple price list. Treat is an indicator of whether the teacher was assigned to the treatment group. All specifications include controls selected using post-double selection LASSO procedure following Belloni et al. (2014). The full set of controls is in Appendix Table A5. Risk preferences capture risk attitudes captured at baseline using a multiple price list. Baseline value of outcome, strata fixed effects, and risk preferences are not penalized across regressions. Data is pooled across endline rounds, and is at the teacher x endline level. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. I report adjusted p-values using Benjamini-Hochberg and Benjamini-Yekutieli step-up methods. Asterisks on coefficients correspond to unadjusted p-values. * p<0.1, ** p<0.05, *** p<0.01.

Table 5: *Treatment Effect on Teachers' Beliefs (Self-Reported)*

	(1) Aggregate Index	(2) Beliefs about teaching inputs	(3) Locus of Control	(4) Self-Efficacy
Treat	0.232***	0.064**	0.156**	0.105
	(0.080)	(0.032)	(0.072)	(0.075)
p-value	0.004	0.043	0.029	0.161
Benjamini Hochberg q-value	0.016	0.087	0.087	0.161
Yekutieli p-value	0.033	0.121	0.121	0.336
Control Mean	-0.12	0.32	0.00	0.00
Observations	782	782	782	782

Notes: This table presents the treatment effect on self-reported beliefs of teachers elicited using surveys. The dependent variable in col (2) is an indicator for whether the teacher ranked a teacher-level input as the most important input for learning, from a list of nine inputs including student, school, and teacher-level inputs. Locus of control is captured using an adapted version of the Rose and Medway (1981) scale. Self-efficacy is measured using the Schwarzer and Jerusalem (1995) scale. Responses on both indices are aggregated into a weighted index, with weights constructed using the inverse of variance-covariance matrix on the lines of Anderson (2009). The dependent variable in cols (3) and (4) is standardized with respect to the control group to have a mean of zero and a standard deviation of one. All specifications include controls selected using post-double selection LASSO procedure following Belloni et al. (2014). The full set of controls is in Appendix Table A5. Baseline value of outcome, and strata fixed effects are not penalized across regressions. Data is pooled across endline rounds, and is at the teacher x endline level. Robust standard errors in parenthesis. I report adjusted p-values using Benjamini-Hochberg and Benjamini-Yekutieli step-up methods accounting for false discovery rate. Asterisks on coefficients correspond to unadjusted p-values. * p<0.1, ** p<0.05, *** p<0.01.

Table 6: *Treatment Effect on Teacher Effort*

	Admin data	Observations		Self-Reports	
	(1)	(2)	(3)	(4)	(5)
	Attendance	Classroom Effort	Grading Effort	Timeuse: Checking HW	After-school tutoring
Treat	-0.012	0.134**	0.111**	8.449***	0.097***
	(0.013)	(0.058)	(0.044)	(2.322)	(0.030)
p-value	0.376	0.021	0.012	0.000	0.001
BH q-value	0.376	0.042	0.035	0.001	0.005
Yekutieli p-value	0.858	0.060	0.044	0.003	0.007
Control Mean	0.87	0.00	0.00	76.90	0.18
Observations	277	752	752	782	782

Notes: The dependent variable in column (1) is teacher attendance (percentage of days present) which was collected using school administrative records and a centralized biometrics database. Some schools transitioned from a paper system of attendance collection to a biometric system during the middle of the academic year 2022-23, data on 15 teachers were lost due to technical glitches. As a result, the number of observations (277) is lower than the full teacher sample of 292. The dependent variable in columns (2) and (3) are pre-registered indices of classroom and grading effort constructed using GLS weighted procedure (Anderson, 2008) for measures captured during classroom observations by trained surveyors at four points in the academic year. The dependent variable in col (4) is self-reported minutes spent on checking homework notebooks, on an average day. The dependent variable in col (5) is an indicator for whether the teacher conducted after-school tutoring. Data on self-reports comes from four surveys conducted with all study teachers throughout the academic year 2022-23 - at baseline, and at three endline rounds. Data in cols (2) and (3) come from classroom observations done at four points during the year - at baseline, and three endline rounds. Data in col (1) is at the teacher level, and in cols (2) - (5) is at the teacher x endline level. All specifications include controls selected using post-double selection LASSO procedure following Belloni et al. (2014). The full set of controls is in Appendix Table A5. Baseline value of outcome, and strata fixed effects are not penalized across regressions. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. I report adjusted p-values using Benjamini-Hochberg (BH) and Benjamini-Yekutieli step-up methods accounting for false discovery rate. Asterisks on coefficients correspond to unadjusted p-values. * p<0.1, ** p<0.05,

Table 7: *Treatment Effect on Disaggregated Measures of Teachers' In-Class Effort*

	(1) Pooled Index	(2) Materials and Content	(3) Classroom Climate	(4) Engagement	(5) Accessibility	(6) Demeanor	(7) Pedagogical Practices
Treat	0.129**	0.120**	-0.013	0.189***	0.091	0.057	0.148**
	(0.065)	(0.058)	(0.068)	(0.055)	(0.069)	(0.069)	(0.061)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	752	752	752	752	752	752	752
R-squared	0.24	0.24	0.22	0.37	0.24	0.25	0.29

Notes: The dependent variables are standardized weighted indices, generated using GLS weighted procedure for measures captured during classroom observations by trained surveyors at four points in the academic year. The pooled index is a weighted index of classroom effort constructed using Anderson (2008) using all measures in the teacher observation tool, shown in Table D.2. Treat is an indicator for whether the teacher was assigned to the treatment group. Results are pooled OLS for data combined across all baseline and endline surveys. All columns control for strata fixed effects, round fixed effects, controls for teacher education, and baseline value of the outcome. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table 8: *Treatment Effect on Disaggregated Measures of Grading Effort*

	(1) Pooled Index	(2) Checked	(3) Detailed Feedback	(4) Encouraging Feedback
Treat	0.099*	0.001	0.056*	0.033
	(0.051)	(0.010)	(0.029)	(0.033)
Controls	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes
R-squared	0.53	0.89	0.27	0.32
Observations	752	752	752	752

Notes: Pooled Index is a standardized weighted index of past grading effort, generated using GLS procedure for three measures captured using surveys of homework notebooks of two random students in teachers' classroom. The measures captured whether the notebook was checked, had detailed feedback, and had any encouraging feedback. Control group means for checked, detailed feedback and encouraging feedback are 0.73, 0.44, and 0.35, respectively. All specifications show results for pooled OLS on data combining all survey rounds. All columns control for strata fixed effects, round fixed effects, teacher education and baseline value of the outcome. Standard errors are clustered at the teacher level. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 9: *Treatment Effect on Student Learning*

	Standardized Math Scores		
	(1)	(2)	(3)
Treat	0.091^{**}	0.101^{**}	0.096^{**}
	(0.045)	(0.046)	(0.046)
Pre-mid score	0.768^{***}	0.769^{***}	0.761^{***}
	(0.017)	(0.017)	(0.017)
p-value for Treat = 0	0.042	0.028	0.036
Controls	No	Yes	PDS Lasso
Observations	6941	6941	6941

Notes: This table reports treatment effects on student learning outcomes. Student learning is measured using test-scores on end-of-year Math exam. All exams were centrally designed and externally graded across academies. The dependent variable is the student's math score on end-of-year exams, standardized within grade to have a mean of zero and SD of one for control group at baseline. Pre-mid term scores are standardized within cluster x grade, not with reference to control group. This is because not all cluster x grades have a control, and some clusters have only one school. Missing values for past scores are imputed with the mean value, and regressions include a flag for missing values. Column (2) includes control for teacher education which was imbalanced at baseline. Column (3) includes controls selected using post-double selection LASSO procedure following Belloni et al. (2014). The full set of controls is in Appendix Table A5. Lagged test score, and strata fixed effects are not penalized across regressions. All regressions include strata (school-grade-pair) fixed effects. Standard errors are clustered at the teacher level (unit of randomization). *p<0.1, ** p<0.05, *** p<0.01.

Table 10: *Heterogeneity in Treatment Effects by gender, SES, grade-level, and baseline ability*

	(1) Female	(2) Low-SES	(3) High BL score	(4) Primary grades
Treat	0.068	0.098**	0.113**	0.044
	(0.050)	(0.046)	(0.054)	(0.078)
Covariate	0.127***	-0.010	0.050	-0.311***
	(0.030)	(0.043)	(0.054)	(0.084)
Interaction	0.050	0.018	-0.027	0.082
	(0.041)	(0.075)	(0.059)	(0.133)
R-squared	0.51	0.50	0.50	0.51
Observations	6687	6895	6895	6895

Notes: This table reports heterogeneity in treatment effects on student learning by student covariates. The dependent variable is the standardized z-score on end-of-year Math scores. Student gender was determined based on their names. For 3% of the sample, the student's name was gender-neutral, I leave these missing. Low-SES is an indicator that takes the value 1 if the student received need-based financial assistance, based on school records. High baseline score is an indicator for an above-median pre-midterm score. Primary grade is an indicator that takes the value 1 for grades 2 and 4. All regressions include strata (school-grade-pair) fixed effects. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis * p<0.1, ** p<0.05, *** p<0.01

Table 11: *Heterogeneity in Effects on Student Learning by Baseline Teacher Characteristics*

	(1) Experience	(2) Baseline beliefs	(3) Baseline effort	(4) Value-added
Treat	0.126*	0.217***	0.156**	0.167**
	(0.069)	(0.069)	(0.073)	(0.070)
Covariate	-0.159*	0.062	0.037	-0.150*
	(0.085)	(0.086)	(0.087)	(0.087)
Interaction	-0.038	-0.213*	-0.123	-0.146
	(0.106)	(0.108)	(0.119)	(0.119)
R-squared	0.51	0.50	0.50	0.51
Observations	6895	6895	6895	6766

Notes: This table reports heterogeneity in treatment effects on student learning by teacher characteristics. The dependent variable is the standardized z-score on end-of-year Math scores. Column titles indicate the covariate used for the regression. High experience is an indicator for above-median years of experiences. Baseline beliefs is an indicator that takes the value 1 if the teacher has an above-median score on the belief index at baseline. Belief index is an aggregate index of locus of control, self-efficacy, and self-reported beliefs about the importance of teaching inputs at baseline. Baseline effort is an indicator that takes the value 1 if the teacher has an above-median score on the effort index at baseline. Effort index is an aggregate index of classroom effort constructed using a pre-registered measure. Teacher value-added is constructed using pre-intervention scores for students taught by teachers, constructed by estimating the teacher fixed effects on regression of pre-mid term scores on previous year scores, student SES, and gender. Value added is an indicator for above-median value-added. All regressions control for past score, teacher education, and strata (school-grade-pair) fixed effects. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis * p<0.1, ** p<0.05, *** p<0.01

Table 12: *Treatment Effect on Secondary Outcomes*

	(1) Growth Mindset	(2) Mental Health	(3) Risk Preferences
Treat	-0.059	0.085	-0.002
	(0.093)	(0.074)	(0.037)
Round FE	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes
R-squared	0.43	0.46	0.33
Observations	522	781	584

Notes: Growth Mindset is captured using Blackwell et al. Implicit Theories of Intelligence scale. The dependent variable is the z-score for growth mindset. Mental health is captured using CES-D Radloff (1977) scale. The dependent variable is the z-score constructed from the summed score. Risk preferences were elicited using a multiple price list with ten hypothetical choices between a safe payment and a lottery. The dependent variable is an indicator for risk aversion and takes the value 1 if the certainty equivalent elicited through the first switching point is less than the expected value of the lottery in the multiple price list task. All specifications are for pooled data combining endline rounds. All regressions include round fixed effects, strata fixed effects, baseline controls (if available). Standard errors are clustered at the teacher level. Number of observations differ as some measures were captured at 2 or 3 follow-up rounds. Data is at the teacher x endline level. Robust standard errors in parenthesis * p<0.1, ** p<0.05, *** p<0.01

Appendix A: Supplementary Analyses

Table A1: *Sample Characteristics*

	Mean	SD	Min	Max	N
School Characteristics					
Total enrollment	502.76	341.06	130	1800	83
Total number of teachers	26.55	15.07	9	82	83
Teacher-student ratio	18.39	3.47	10.73	25.26	83
Proportion of girls	0.46	0.04	0.38	0.59	83
Years since establishment	13.11	5.84	4	29	83
Principal's years of teaching experience	13.34	6.93	1	32	83
Principal's years of leadership experience	7.46	4.83	0	25	83
Teacher Characteristics					
Age	33.19	6.92	21	64	292
Years of Experience	7.46	5.31	0	29	292
Female	0.82	0.39	0	1	292
Education Level: Masters and above	0.69	0.46	0	1	292
Teacher training degree/diploma	0.88	0.32	0	1	292
Permanent	0.58	0.49	0	1	292
Number of periods taught per day	6.79	1.47	1	9	292
Number of periods taught per week	40.13	10.39	5	72	292
Teaching subjects other than Math	0.67	0.47	0	1	292
Workload is too much	0.32	0.47	0	1	292
Teacher Beliefs					
At least half of my class					
- consists of first-generation learners	0.32	0.47	0	1	292
- comes from financially constrained backgrounds	0.12	0.33	0	1	292
- have parents who don't engage in child's schooling	0.46	0.5	0	1	292
Ranked teacher-level input as most important for student learning	0.31	0.46	0	1	292
Locus of Control	4.24	2.17	0	10	292
Generalized Self-Efficacy	3.54	0.38	2	4	292

Notes: This table presents summary statistics for schools and teachers in the study sample. School characteristics were provided by schools at the time of the baseline school visit. Teacher characteristics and beliefs are based on self-reports by teachers in the baseline survey. Teachers were presented with a list of nine inputs that are important for student learning (including student-level inputs such as student effort, ability, prior learning, family background; teacher-level inputs such as teachers' teaching method and attention to students, and school-level inputs such as school management, facilities, and staff salaries), and asked to rank the inputs from most important to least important for student learning. Locus of Control is measured using a 10-item adapted version of the Rose and Medway (1981) teacher locus of control scale. Higher values on the 10-point scale indicate a more internal locus of control. Generalized Self-Efficacy is measured using the 10-item Schwarzer & Jerusalem (1995) scale. Higher values on the scale indicate higher perceived self-efficacy.

Table A2: *Teacher Attrition*

	Missing in pooled data		Missing at any round		Missing at all rounds	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	-0.032	-0.031	-0.027	-0.013	-0.021	-0.023
	(0.030)	(0.021)	(0.046)	(0.046)	(0.024)	(0.024)
Constant	0.115***	0.053*	0.199***	0.092	0.055***	0.015
	(0.025)	(0.029)	(0.033)	(0.061)	(0.019)	(0.026)
Controls	No	Yes	No	Yes	No	Yes
R-squared	0.003	0.400	0.001	0.584	0.002	0.530
Observations	876	876	292	292	292	292

Notes: This table shows results from regressing indicator for attrition on treatment status. Cols (1) and (2) show data pooled across rounds, the dependent variable is an indicator for missing in surveys at the teacher x endline level. Cols (3) - (6) show data at the teacher level. The dependent variable in cols (3) and (4) is an indicator for missing at any endline round conditional on being present at baseline. The dependent variables in cols (5) and (6) is an indicator for missing at all endline rounds conditional upon being present at baseline. The specifications in cols (2), (4) and (6) control for strata fixed effects and teacher education which was imbalanced at baseline. Robust standard errors are in parentheses and clustered at the teacher level.

Table A3: *Intervention Compliance*

	(1) Attended first session	(2) Attended at least one session	(3) Percentage of sessions attended
Treat	0.000	0.066	0.067
	(0.057)	(0.044)	(0.044)
Control Mean	0.72	0.82	0.48
R-squared	0.525	0.531	0.564
Observations	292	292	292

Notes: This table presents compliance in terms of attendance at intervention sessions by treatment status. The dependent variable in col. (1) is an indicator for attendance at the first session. The dependent variable in col. (2) is an indicator for having attended at least one session. The dependent variable in col. (3) is the percentage of sessions attended (out of a total of 10). All columns control for strata fixed effects. Robust standard errors in parentheses clustered at the school and training group levels.

Table A4: *Reasons for Attrition*

Type	Baseline	Endline 1	Endline 2	Endline 3
Tracked and surveyed	292	263	260	259
In school	292	256	255	241
Over virtual meeting	0	7	5	18
Tracked but could not be surveyed	0	29	32	33
Teacher left	0	12	23	22
Extended leave (medical/maternity)	0	9	4	5
Teacher not in school	0	8	5	5
Teacher refused	0	0	0	1

Notes: Notes: The sample analyzed in this table are 292 teachers who were surveyed at baseline and tracked in each endline round. For all teachers who were not present in school on the day of the visit, a follow-up visit was attempted to conduct a classroom observation. For those on extended leave, a virtual meeting was conducted over Zoom to administer the teacher survey and conduct belief elicitation. The number of observations in the first row correspond to sample of teachers who responded in the surveys. The number of observations in the second row correspond to the N in all results on classroom effort measurements since these could not be conducted after school.

Table A5: *Full set of controls for Double-LASSO*

S. No.	Variables
1	School: Enrollment
2	School: Total number of teachers
3	School: Teacher-student ratio
4	School: Proportion of girls
5	School: Years since establishment
6	School: Principal's years of teaching experience
7	School: Principal's years of leadership experience
8	Teacher: Age
9	Teacher: Years of experience
10	Teacher: Female
11	Teacher: Education level (masters and above)
12	Teacher: Holds training degree/diploma
13	Teacher: Permanent
14	Teacher: Number of periods taught per day
15	Teacher: Number of periods taught per week
16	Teacher: Teaches subjects other than math
17	Teacher: Size of average class
18	Teacher: Teaches grade 2
19	Teacher: Teaches grade 4
20	Teacher: Teaches grade 6
21	Teacher: Teaches grade 8
22	Classroom: More than half are first-gen learners*
23	Classroom: More than half come from financially constrained backgrounds*
24	Classroom: More than half the parents don't engage in child's schooling*
25	Baseline: Social desirability score (Marlowe Crowne index)
26	Baseline: Locus of Control (Anderson index)
27	Baseline: Generalized Self-Efficacy (Anderson index)
28	Baseline: Risk Aversion (from multiple price list)
29	Baseline: Teacher-level input ranked as most important*
30	Baseline: Timeuse – time spent on travel
31	Baseline: Timeuse – time spent on admin work
32	Baseline: Timeuse – time spent on class preparation
33	Baseline: Timeuse – time spent on checking notebooks
34	Baseline: Teacher expectation for a weak student
35	Baseline: Teacher expectation for a below-average student
36	Baseline: Teacher expectation for an above-average student
37	Baseline: Teacher expectation for an intelligent student
38	Student: Female*
39	Student: Received financial assistance
40 - 44	Flags for missing (x 5)

Notes: The table shows the list of 45 potential control variables from which the double LASSO procedure of Belloni, Chernozhukov, and Hansen (2014) chose the final set of controls. Five variables (marked with *) had missing values. For these, flags for missing values are also included. All regressions control for the baseline value of outcome and strata fixed effects, which are not penalized in line with the pre-analysis plan.

Table A6: *Distribution of Response Patterns on Multiple Price List for Belief Elicitation*

	Round			
Type	Baseline	Endline 1	Endline 3	Total
Well-behaved preferences				
Unique switch	189	223	232	644
Performance-linked option always	20	9	4	33
Flat option always	21	10	5	36
Preference-reversal				
Reverse switch	5	6	4	15
Multiple switches	57	15	14	86
Total	292	263	259	814

Notes: This table shows the distribution of different types of responses on the belief elicitation task at each round of data collection.

Table A7: *Round-Wise Effects for Teachers' Beliefs (Revealed Preference Measure)*

	Dependent Variable: Switching Point								
	Pooled			1-Month Endline			6-Month Endline		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat	0.466**	0.510**	0.430**	0.491	0.511	0.379	0.459	0.528	0.489
	(0.213)	(0.212)	(0.207)	(0.375)	(0.380)	(0.378)	(0.334)	(0.323)	(0.319)
Baseline controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Risk Preferences	No	No	Yes	No	No	Yes	No	No	Yes
Control Mean	5.10	5.10	5.10	5.42	5.42	5.42	4.79	4.79	4.79
Observations	522	522	522	263	263	263	259	259	259
R-squared	0.43	0.44	0.45	0.60	0.61	0.63	0.65	0.67	0.67

Notes: The dependent variable is a revealed preference measure of teachers' perceived control captured by the row in which the teacher first switched from the performance-linked option to the flat pay option in the multiple price list task and ranges from 1 to 11. Treat is an indicator of whether the teacher was assigned to the treatment group. Baseline controls include an indicator for teacher education (masters or above) that was imbalanced at baseline, and the baseline analog of the outcome. Risk preferences capture risk attitudes captured at baseline using a multiple price list. Cols (1)-(3) show effects for pooled data from both endlines. All regressions include strata fixed effects. Pooled specifications also include round fixed effects. Standard errors are clustered at the teacher level. Cols (4)-(6) show results at the first endline, one month after the intervention. Cols (7)-(9) show results at the last endline, six months after the intervention. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A8: *Round-Wise Effects for Teachers' Beliefs (Revealed Preference)*

	Dependent Variable: Test-Score Value Mapped to Switch Point								
	Pooled			1 Month Endline			6 Month Endline		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat	1.509**	1.474**	1.307**	1.259	1.115	0.729	1.771*	1.812*	1.837*
	(0.633)	(0.640)	(0.644)	(1.274)	(1.298)	(1.294)	(0.968)	(0.976)	(1.049)
Baseline controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Risk Preferences	No	No	Yes	No	No	Yes	No	No	Yes
Control Mean	7.62	7.62	7.62	8.43	8.43	8.43	6.81	6.81	6.81
R-squared	0.37	0.38	0.38	0.61	0.61	0.63	0.65	0.65	0.66
Observations	522	522	522	263	263	263	259	259	259

Notes: The dependent variable is the revealed preference measure of teachers' perceived control in units of predicted test-score increase over a year, and corresponds to the mid-point of test-score bracket mapped to each row in the multiple price list. Treat is an indicator for whether the teacher was assigned to the treatment group. Baseline controls include an indicator for teacher education (masters or above) that was imbalanced at baseline, and the baseline analog of the outcome. Risk preferences capture risk attitudes captured at baseline using a multiple price list. Cols (1)-(3) show effects for pooled data from both endlines. All regressions include strata fixed effects. Pooled specifications also include round fixed effects. Standard errors are clustered at the teacher level. Cols (4)-(6) show results at first endline, one month after the intervention. Cols (7)-(9) show results at last endline, six months after the intervention. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A9: *Robustness of Treatment Effects: Teacher Beliefs*

	(1)	(2)	(3)	(4)	(5)
Treat	0.445**	0.398	0.428*	0.459**	0.417**
	(0.221)	(0.251)	(0.245)	(0.205)	(0.201)
Baseline controls	No	Yes	Yes	Yes	Yes
Risk preferences	No	No	Yes	No	Yes
Missing value adjustment	No	No	No	Yes	Yes
Observations	483	384	384	483	483
R-squared	0.47	0.59	0.59	0.51	0.52

Notes: This table shows the robustness of treatment effects restricting sample to teachers who exhibited well-behaved preferences. The dependent variable is the switching point in the multiple price list task. Baseline controls include an indicator for teacher education (masters or above) that was imbalanced at baseline, and the baseline analog of the outcome. Risk preferences capture risk attitudes captured at baseline using a multiple price list. Given that around 21% teachers exhibited inconsistent preferences at baseline, controlling for the baseline value of outcome leads to a drop in the number of observations to 384 in columns 2 and 3. The *p*-value of the coefficient estimate on the indicator for treatment in columns (2) and (3) is 0.11, and 0.13 respectively. Col.(4) and (5) impute the missing value for baseline switching point as zero while adding a missing value dummy for these observations. All specifications are for pooled data combining all rounds and control for both strata and round fixed effects. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A10: *Correlation between Belief Measures and First Principal Component*

Measure	Type	First Principal Component
Beliefs about effort	Vignette	-0.126
Locus of control	Psychometric scale	0.618
Generalized self-efficacy	Psychometric scale	0.404
Beliefs about teaching inputs	Ranking task (CEPS)	0.263
Switching point from Multiple Price List	Real-stakes task	0.608

Notes: The table shows loadings for the first principal component, representing the correlation between each belief measure and the first principal component from all measures, for the control group. The vignette measure is constructed using responses to vignettes presented to teachers about their estimates of how much the scores of their students (low-/medium-/high-types) will increase in response to an additional hour of effort each week. The measure is a summary index across the three questions using Anderson (2008). Locus of control and self-efficacy were measured using standard psychological scales: Rose and Medway (1981) for locus of control, and Schwarzer and Jerusalem (1995) for self-efficacy. Beliefs about teaching inputs corresponds to an indicator for whether a teacher ranked teacher-level inputs as the most important out of a list of nine inputs. The switching point is a revealed preference measure of teachers' perceived control captured using the multiple price list task detailed in Appendix B that asked teachers to make binary choices across flat and performance-linked contract types to receive a bonus at the end of the year. Data comes from teacher surveys.

Table A11: *Treatment Effect for Vignette Measure*

	(1) Average	(2) Low	(3) Medium	(4) High
Treat	-0.633	-0.618	-0.515	-1.001*
	(0.392)	(0.568)	(0.313)	(0.558)
Baseline controls	Yes	Yes	Yes	Yes
Strata fixed effects	Yes	Yes	Yes	Yes
Observations	782	782	782	782
R-squared	0.55	0.51	0.40	0.49

Notes: This table reports treatment effects on teacher beliefs elicited through a set of vignette questions. The vignette measure is constructed using responses to hypothetical scenarios presented to teachers about their estimates of how much the scores of their students (low-/medium-/high-types) will increase in response to an additional hour of effort each week. The dependent variable in col (1) is a simple average of variables in columns (2)-(4). The dependent variable in columns (2)-(4) is the difference between the teachers' estimate of the total score and the baseline score mentioned in the vignette for the respective student type. All specifications control for the baseline analog of outcome, teacher education, a z-score for social conformability at baseline, strata, and round fixed effects. Data is pooled across endline rounds and is at the teacher x round level. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A12: *Effect on Teachers' Risk Preferences*

	Dependent Variable: Indicator for Risk Aversion		
	Pooled	1-Month Endline	6-Month Endline
	(1)	(2)	(3)
Treat	-0.002	0.055	-0.059
	(0.037)	(0.061)	(0.059)
Strata FE	Yes	Yes	Yes
Round FE	Yes	No	No
Baseline value of outcome	Yes	Yes	Yes
Control Mean	0.39	0.54	0.62
Observations	584	292	292
R-squared	0.33	0.57	0.57

Notes: Risk preferences are elicited using a multiple price list with ten hypothetical choices between a safe payment and a lottery. The dependent variable is an indicator for risk aversion and takes the value 1 if the certainty equivalent elicited through the first switching point is less than the expected value of the lottery in the multiple price list task. Treat is an indicator for whether the teacher was assigned to the treatment group. All columns control for baseline risk preferences, indicator for teacher education, and strata fixed effects. Column 1 shows the effect for pooled data from both endlines; this specification also includes round fixed effects. Standard errors are clustered at the teacher level. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A13: *Treatment Effect on Teacher Attendance*

	Percentage of Days Present		Number of Days Present	
	(1)	(2)	(3)	(4)
Treat	-0.002	-0.005	-1.106	-1.704
	(0.021)	(0.019)	(5.603)	(5.108)
Controls	No	Yes	No	Yes
Control Mean	0.84	0.84	231.47	231.47
R-squared	0.57	0.58	0.67	0.67
Observations	277	277	277	277

Notes: This table reports treatment effects on teacher attendance, which was collected using school administrative records and a centralized biometrics database. Note that some schools transitioned from a paper system of attendance collection to a biometric system during the middle of the academic year 2022-23. While biometric attendance was independently retrieved at the end of the year, data on 15 teachers were lost due to technical glitches. As a result, the number of observations is lower than the full teacher sample.

Table A14: *Round-wise Treatment Effect on Teacher Effort in Classroom*

	Pooled		1 Month Endline		3 Month Endline		6 Month Endline	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat	0.133**	0.129**	0.121	0.130	0.167	0.147	0.035	0.040
	(0.066)	(0.065)	(0.141)	(0.145)	(0.162)	(0.161)	(0.137)	(0.137)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	752	752	256	256	255	255	241	241
R-squared	0.24	0.24	0.70	0.70	0.59	0.60	0.64	0.64

Notes: The dependent variable is a standardized weighted index of classroom effort, generated using GLS weighted procedure for measures captured during classroom observations. Identical weights are used across all endlines, to ensure comparability of the outcome variable. The index is normalized with reference to the control group for the relevant round, and has an effect size interpretation. Treat is an indicator for whether the teacher was assigned to the treatment group. Baseline effort is the classroom effort at baseline. Controls include dummies for bachelors and master degrees, teacher-level variables that were imbalanced at baseline. Cols (1)-(3) show effects using pooled OLS on data from three endlines. Standard errors are clustered at the teacher level. Cols (4)-(6) show results at first endline, one month after the intervention, cols (7)-(9) show results at second endline, three months after the intervention, cols (10)-(12) show results at third endline, six months after the intervention. Robust standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A15: *Treatment Effects on Student Learning: Robustness across Full Sample*

	Final Score		
	(1)	(2)	(3)
Treat	0.074*	0.067	0.082*
	(0.043)	(0.047)	(0.045)
Pre-mid term score	0.766***	0.465***	0.791***
	(0.017)	(0.024)	(0.017)
p-value for Treat = 0	0.089	0.154	0.068
Observations	7425	6526	6526
R-squared	0.50	0.61	0.53

Notes: This table presents effects on student learning for the full sample of students, including those whose teachers left in the middle of the study. These students were taught by another teacher in the school who may or may not have been a study teacher. Given the potential for contamination, these effects reflect a lower bound on treatment effect. All specifications control for teacher education. Col (3) runs the same specification as in col (1) restricted to subsample in col (2), that is, students with non-missing previous year scores. All specifications include strata fixed effects. Standard errors are clustered at the teacher level.

Appendix B: Experimental Task

B.1. Task Design

Teachers were presented with the below table with a set of choices to receive bonus pay for the current academic year. For each row, they were asked to choose the left or the right option. For the right-hand-side option, teachers were informed that a computer would randomly choose a below-average student in their class. The payment in the right-hand-side would be linked with the performance of this randomly chosen student. At the end of the school year, based on a lucky draw, one teacher per school would be awarded payment, based on a randomly selected row.

Table B.1. Belief Elicitation through Multiple Price List

Choice	**Option 1**	**Option 2**
Choice 1	Rs 1000 flat	Rs 2000 for maintaining at least the same test score across exams (either score remains same or increases, but does not decrease)
Choice 2	Rs 1000 flat	Rs 2000 for an increase in test score of at least 2 marks
Choice 3	Rs 1000 flat	Rs 2000 for an increase in test score of at least 4 marks
Choice 4	Rs 1000 flat	Rs 2000 for an increase in test score of at least 6 marks
Choice 5	Rs 1000 flat	Rs 2000 for an increase in test score of at least 8 marks
Choice 6	Rs 1000 flat	Rs 2000 for an increase in test score of at least 10 marks
Choice 7	Rs 1000 flat	Rs 2000 for an increase in test score of at least 12 marks
Choice 8	Rs 1000 flat	Rs 2000 for an increase in test score of at least 14 marks
Choice 9	Rs 1000 flat	Rs 2000 for an increase in test score of at least 16 marks
Choice 10	Rs 1000 flat	Rs 2000 for an increase in test score of at least 50 marks

Notes: To reduce noise due to different priors about student ability distribution, this choice is presented to teachers after they know their student allocation for 2022-23 school year. I specifically target lower-performing students because teachers are more likely to perceive external constraints (family background etc.) as binding for these students. In this setting, all exams across academies are designed by a centralized examinations cell, and teachers do not have control over the design and grading.

Interpretation of responses: Someone who chose Option 1 over Option 2 in Choice 2 believes that she is incapable of triggering even a 2 point increase in test scores for a random weak student. Someone who chose Option 2 in Choice 2 but switched to Option 1 in Choice 3 believes she can trigger increase in student test scores of between 2-4 points (excluding upper bound), but no more. Similarly, someone who chose the risky option in choices 1, 2, and 3 but switched to safe option in choice 4 perceives her control to be up to 4-6 percentage point increase in student test scores.

I use the switching point as my key outcome that elicits teachers' beliefs about their perceived control over education production for lower-performing students in their classroom. I expect to see a shift in the distribution of switching points for the treatment group relative to the control group at endline, relative to the baseline. Given that my intervention didn't change the costs of effort, I argue that the shift in switching points is driven by beliefs about returns to effort.

B.2. Task Administration Script

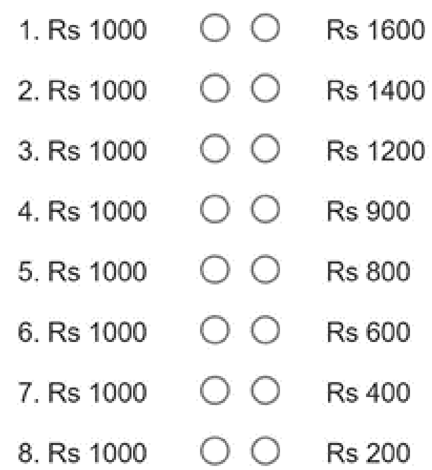
The task is presented to teachers at baseline where the choice is intended to be implemented at the end of the academic year. I present the same choices at the first endline where teachers are presented with an opportunity to revise their choices for receiving payment at the end of the year. They are then presented the same choices at the third endline, where choices are elicited for the next academic year. Below is the full text of the pre-recorded script that played at each school site. Teachers completed a qualtrics survey on their mobile devices. When teachers reached the last section, the script was played.

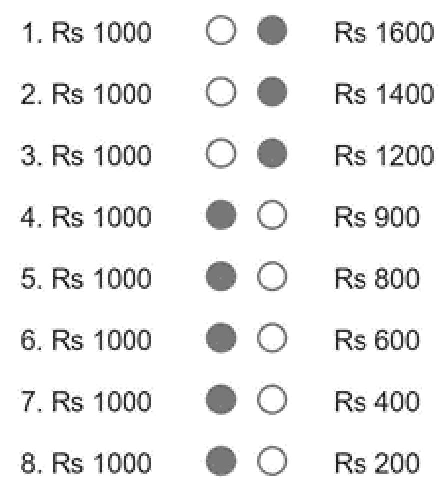
Good Morning Everyone! Today we will play a task that requires you to take certain choices which will decide your chances of winning a bonus payment. Your responses will be confidential and your school management will not be able to see them. You must give all answers without fear or hesitation, based on whatever feels right to you. Ok? Now we will give some instructions to play this task. Please listen to them carefully.

Baseline. *We have reached the last section in the teacher survey which is very interesting to play. Section 7 is a unique type of section that will involve making choices which could win you a bonus payment. You will be shown a few options, and asked to choose one of them. First of all, we will show some examples to familiarize you with the format, then we will move to the actual question. The passcode for this section is 789789, I repeat 789789.*

On your screen, you have a table with 8 rows with numbers in it. Look carefully. On the left is always a fixed number – 1000. On the right, the numbers are decreasing – 1600, 1400, 1200 and so on. These numbers are actually payments or cash prizes. Now, lets look at each row. If you are given an option to take a cash prize worth Rs 1000 cash or Rs 1600, what would you choose? It is obvious that you will choose the option in which the amount is more. E.g. 1600 is more than 1000, so it makes sense to choose 1600 – the right option.

The next table is a filled table. A choice has been made in each row: between 1600 and 1000, 1600 was chosen. Between 1000 and 1400, 1400 was chosen. Between 1000 and 1200, 1200 was chosen. Between 1000 and 900, 1000 was chosen. After a point you see, initially, right hand option was chosen and after a point left side option was chosen. Why did this switch happen? Because the choice becomes attractive on that side.





So, following this pattern of logic, we have to make our choices between left- and right-hand side options. While choosing the options, please choose as per your calculated profit or loss. Please don't hesitate to maximize your payoff in this game – there is no judgement about greed or any related intention. Your goal is to maximize your award. Is this clear? Any questions?

Now let's do a question, it is available on your screen. It has two choices. On the right side, there is a fixed choice: either you will get 1 lakh rupees, or you will not get anything, you have (50-50) chances of winning it. On the left-hand side, there is a fixed amount of Rs.5000. If you are given this choice, what will you choose?

Similarly, in the 2nd row, the choice is between a fixed amount of Rs 10,000 on the left versus a 50/50 chance of getting 1 lakh or nothing on the right side. Similarly, for each row, you have to choose between left and right options. Now you can fill the table.

Till now the tables were imaginary. Moving to the next question, where you have the opportunity to get an actual payment. There is going to be real money involved. Listen carefully.

1. Rs 5,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
2. Rs 10,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
3. Rs. 20,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
4. Rs 30,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
5. Rs 40,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
6. Rs 50,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
7. Rs 60,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
8. Rs 70,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
9. Rs 80,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0
10. Rs 90,000	○ ○	50/50 chance between Rs 1 lakh and Rs 0

There are two important rules of payment involved: 1. One row will be randomly selected for payment. You will get paid according to your choice in the row which has been selected. If you selected left, you will get the LHS payment and if right then you will get the RHS payment. The implication is that you must make a careful and informed choice in each row, because any row can be selected for actual payment. 2. Next, one teacher will get payment in every school. Given that you four/three are part of this study - one of you will be randomly chosen for payment. This random choice of a teacher and a row will be done by a computer.

Before the final question, lets do a small test to check your understanding of the payment rules.

Q1. If row no. 3 is selected, what will you be paid? (a) 1200 (b) 1000

Q2. If row no. 5 is selected, what will you be paid? (a) 800 (b) 1000

Q3. Given 4 teachers in a school, what is your chance of receiving a payment?
(a) 1 in 4 (b) 1 in 6

If anyone has questions, please ask. «Surveyor to answer all possible queries. »

The passcode for the final question is: 312456 Now we will give you the final question in which payment is involved. You can get a bonus other than your salary and there is a scheme for this.

There is a fixed amount on the left side (Rs 1000). There is a bonus on the right side, which is subject to a condition. A computer will randomly select a below-average student from your class and compare his marks in the final exam with the current scores (meaning mid-term scores). If the student's math score registers an increase of the given amount at the end of the year, then you will qualify to receive this bonus.

For example, if you see the first row, there is a a fixed amount on the left side and a deal on the right side which is related to the marks of the same below-average student. If the student's scores remain the same or increase by any amount, then you will get 2000 rupees. If the scores decrease,

1. Rs 1000	○ ○	**Rs 2000** for maintaining at least the same test score across exams (**either score remains same or increases, but does not decrease**)
2. Rs 1000	○ ○	**Rs 2000** for an increase in test score of a low-performing student of **at least 2 marks**
3. Rs 1000	○ ○	**Rs 2000** for increase of **at least 4 marks**
4. Rs 1000	○ ○	**Rs 2000** for increase of **at least 6 marks**
5. Rs 1000	○ ○	**Rs 2000** for increase of **at least 8 marks**
6. Rs 1000	○ ○	**Rs 2000** for increase of **at least 10 marks**
7. Rs 1000	○ ○	**Rs 2000** for increase of **at least 12 marks**
8. Rs 1000	○ ○	**Rs 2000** for increase of **at least 14 marks**
9. Rs 1000	○ ○	**Rs 2000** for increase of **at least 16 marks**
10. Rs 1000	○ ○	**Rs 2000** for increase of **at least 50 marks**

then you will get nothing. On the left side, you will always receive Rs 1000 irrespective of student scores.

In the second row: If the student scores increase by 2 points, then you will get 2000 rupees. If the student's scores do not increase by 2 points then you will get zero. On the left side, you can get Rs 1000.

In the third row: If the student's scores increase by 4 points, you will get Rs 2000. On the left side, you can get a fixed amount of Rs 1000. And so on.

Please make a choice between the left and right options, for each row.

Endline 1. *Good Morning! Remember we visited your school a few months back and did section 7. Today we will do that task again. Now let's review all instructions once again and conduct this task.*

«Same instructions as earlier involving two examples, a guided question, and comprehension checks. Before the final task, the following is additionally shared.»

Next is the question involving real payments. We came to you 2-3 months back with this plan, and you made your choices. Now you have the opportunity to make a different selection. Payments will be made at the end of the academic year as per what you decide now.

Endline 3. *Good Morning! Remember we visited your school earlier and you filled out some choices regarding bonus payment. As promised, we will be awarding payments today. The payments will be awarded in accordance with the payment rules shared with you.*

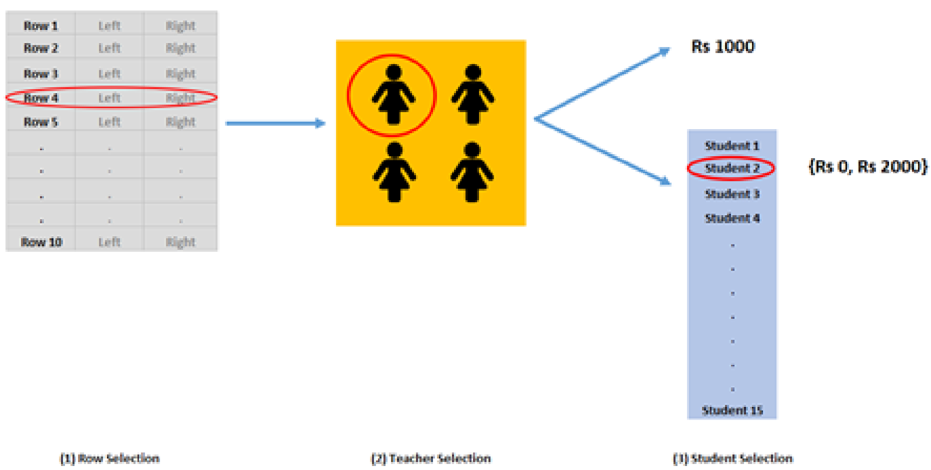
«Surveyor uses a customized spin-the-wheel app which has three wheels for choosing rows, teachers and students linked to each teacher. More details are in the next section.

Payment of Rs 1000, Rs 2000, or nothing is awarded to one teacher per school. After the payments are made, teachers complete another price list that would be relevant for the next year. »

You just saw how payments are made based on this task. If you were given an option to receive similar bonus payments for next year, how would you make your choices? You will now make choices for next year. «A recorded script with instructions is played.»

Please proceed to fill the table on your screen.

Payment Process



Before the school visit: End-of-year math scores on final exams were obtained from schools and classroom-level averages were computed. For each school, surveyors were handed a list with names of teachers, their choices in MPL task at endline 1, names and scores of all below-average students linked to that teacher.

At the school visit: All study teachers were asked to sit in a room. A standardized recording was played explaining the process for payment. Surveyor answered any queries and then used a customized github app to conduct random selection of row, teacher and student to determine payment.

Process of random selection:

1. Example: First, the surveyor asked a teacher to spin a demo wheel. This was done so that teachers trust that the process was truly random.

2. Row Selection: The surveyor then spun a wheel with numbers 1-10. The row selection was announced to all teachers. Teachers were then asked to refer to the relevant row in their tables.

3. Teacher Selection: Next, the surveyor spun a wheel that had names of study teachers for that school. The name of the selected teacher was announced.

Payment process:

- If the teacher had a left-hand-side selection for the corresponding row, an amount of Rs 1000 was awarded. The process then concluded.
- If the teacher had a right-hand-side selection for the corresponding row, then the surveyor spun a third wheel with names of below-average students for the relevant teacher. Teacher was first asked to confirm that the names of students were correct. After the student was selected, the surveyor checked the scores of the student across the final and mid-term exams. If the condition for bonus was met, the teacher received Rs 2000. If the condition for bonus was not met, the teacher received nothing. The process concluded.

Appendix C: Intervention Details

Table C1 provides details of content delivered to both treatment and control groups over the course of five weeks. Both groups received two sessions per week. The format for each week was as follows:

- Session One: Instructor-driven interactive lecture covering the topical matter, and leaving teachers with an assignment of the week. The assignment was due before the debrief session.
- Session Two (Debrief): Group discussion of the assignment responses, led by the instructor.

Implementation Partners

WorldBeing. *WorldBeing* is a non-profit organization based out of the US that develops evidence-based programs from positive psychology and has a record of serving disadvantaged communities in low-income countries around the world. Their toolkit is drawn from the latest research, integrating evidence from the fields of positive psychology, social-emotional healing, youth development, attitudinal healing, and restorative practice.

Sukrit. *Sukrit* is an NGO based in Punjab, India which works in the fields of education, youth empowerment and mental health. They conduct a range of activities including career counseling and mentorship for college-bound youth, skill development workshops for high-school students, and motivational workshops for teacherss. The NGO led the training sessions for the purposes of the study.

Table C1: *Intervention Content for Treatment and Control*

Week	Treatment Group	Control Group
1.	**Character Strengths** *Discussion:* Talent, skills, and character strengths *Activity:* Identify your character strengths, and those of people around you.	**Beekeeping** • What is beekeeping? • Types of bee species • Life span and interesting facts about honeybees
2.	**Self-Efficacy** *Discussion:* They are able who think they are able. Story of a visibly weak soldier who defeated a drunken elephant. *Activity:* Share a problem you recently encountered. Distinguish between factors that were inside/outside your control.	**Beautiful World of Fishes** • Types of fishes • Life span and interesting facts about fishes
3.	**Resilience** *Discussion:* Bouncing back from adversity by turning problems into possibilities. Toolbox for resilience. *Activity:* Share a personal story where you exhibited resilience. Share a story of resilience from someone in your life.	**Save Soil** • Types of soil • Importance for agriculture • Dangers of soil erosion
4.	**Goal-Setting and Planning** *Discussion:* Link between goal-setting and realizing our potential. How to set achievable goals and make progress? *Activity:* Write down your goals related to personal and professional life. What steps will you take to reach these?	**Wonderful Life of Plants** • Lifecycle of plant growth • Plants of the region • Importance of light, water, and soil for plants
5.	**Problem-Solving** *Discussion:* Skills to deal with things outside our comfort zone - emotion management, benefit-finding, sharing with someone, devising an action plan. *Activity:* Identify situations where you overcame tough problems. Create your problem toolbox for future.	**Small Bite, Big Threat** • Mosquitoes that lead to dengue and malaria • Methods for mosquito control

Notes: This table outlines the content of group meetings as part of the intervention. The treatment group was exposed to psychosocial content, while the control group was exposed to informational content about the local environment. There were five lectures, followed by five debrief sessions, for both treatment and control groups. Each group consisted of 29-30 teachers, and was led by a trained facilitator. The assignment and reflection activities were discussed in debrief sessions.

Appendix D: Survey Questions

D.1. Self-Reported Beliefs

Psychological Scales Teachers respond to survey questions related to locus of control and self-efficacy. For locus of control, I use a set of 10 items adapted from Rose and Medway (1981) teacher locus of control scale. Each item presents a forced-choice pair of statements with one internally oriented and another externally oriented. Details are in Table D.1.

To capture generalized self-efficacy, I use the Schwarzer and Jerusalem (1995) scale. Teachers are presented a set of ten statements and asked to indicate the degree to which they agree with the statements on a scale ranging from (1) Never true, (2) Rarely true, (3) Sometimes true, (4) Always true, and (5) Don't know.

1. I can always manage to solve difficult problems if I try hard enough.
2. If someone opposes me, I can find some way to get what I want.
3. I am certain that I can accomplish my goals.
4. I am confident that I could deal efficiently with unexpected events.
5. I can handle unforeseen situations by using my resourcefulness.
6. I can solve most problems if I invest the necessary effort.
7. I can remain calm when facing difficulties by relying on my coping abilities.
8. If I am in trouble, I can usually think of a solution.
9. When I am confronted with a problem, I can usually find several solutions.
10. I can handle whatever comes my way.

Beliefs about Teaching Inputs *Below is a list of factors that influence a student's learning. Rank the following elements from most important to least important, as per your opinion.*

1. Student's talent
2. Student's effort
3. Student's family background
4. Student's prior learning
5. Teacher's teaching method
6. Teacher's attention to students
7. Teacher's salary
8. School's management
9. School's teaching facilities

Vignette Questions *Below are three classroom scenarios. There are no right or wrong answers. Please answer to the best of your understanding.*

1. A is a student in your class who has scored 50 out of 100 in the mid-term examinations. If you put in extra 1 hour of effort every week on this child, what do you think will be his/her score on final exam? *Mention your numeric estimate of final score (between 0 and 100)*

2. B is a student in your class who has scored 80 out of 100 in the mid-term examinations. If you put in extra 1 hour of effort every week on this child, what do you think will be his/her score on final exam? *Mention your numeric estimate of final score (between 0 and 100)*

3. C is a student in your class who has scored 35 out of 100 in the mid-term examinations. If you put in extra 1 hour of effort every week on this child, what do you think will be his/her score on final exam? *Mention your numeric estimate of final score (between 0 and 100)*

Table D.1. Locus of Control

Item	Domain
1. When the grades of your students improve, it is more likely: a. because you found ways to motivate the students, or b. because the students were trying harder to do well.	Grade improvement
2. Suppose you had difficulties with online teaching for students in your classroom. Would this probably happen: a. because you lacked the appropriate materials, or b. because you didn't spend enough time in developing activities suited for online learning?	Quality of preparation
3. When some of your students fail a test, it is more likely: a. because they weren't attentive in class, or b. because you explained the concept with fewer examples	Grade deterioration
4. If you couldn't keep your class quiet, it would probably be: a. because the students came to school more noisy than usual, or b. because you couldn't settle them down.	Class management
Think about a high performing student in your class. Keeping this student in mind, choose the statement that you agree with the most.	
5. Suppose this student performed better when he was assigned to your class compared to the previous year. This would most likely happen: a. because you taught concepts well, or b. because the student was trying harder.	Grade improvement
6. Suppose you are teaching this student a particular concept and the student has trouble learning it. Would this happen: a. because the student didn't pay attention, or b. because the explanation of the concept was not clear	Quality of preparation
7. When this student scores a low score on a test, it is more likely: a. because the student wasn't paying attention to the lessons in class, or b. because you didn't simplify concepts enough to ensure understanding	Grade deterioration
Now think about a low performing student in your class. Keeping this student in mind, choose the statement that you agree with the most.	
8. When this student pulls up grade from a "C" to a "B," it is more likely: a. because you came up with an idea to motivate the student, or b. because the student was trying harder to do well.	Grade improvement
9. Suppose you are teaching this student a particular concept and the student does not get it. Would this happen: a. because the student was very careless, or b. because you couldn't explain it very well?	Quality of preparation
10. When the performance of this student appears to be slowly deteriorating, it is usually: a. because you weren't trying hard enough to motivate him or her, or b. because the student was putting less effort into his or her schoolwork.	Grade deterioration

Notes: This table presents items used to capture teacher locus of control. These items were adapted from Rose and Medway (1981) scale.

D.2. Classroom Visits

Classroom Observations To conduct classroom observations, surveyors were trained to score the classroom observation rubric through in-class field practice by visiting pilot schools.

The survey instrument was created to be bilingual with English and Punjabi text for better surveyor comprehension. For terms that could have multiple connotations, examples were provided in the tool to minimize surveyor subjectivity in filling responses. For example, the phrase 'teaching aide' was accompanied by examples of things that could be included as teaching aides like charts, objects, and counting techniques.

In each school visit, surveyors observed one class per teacher for all study teachers. Surveyors were instructed to sit at the back of the class for the entire duration, keep the questionnaire hidden, and minimize any disturbance to ongoing class activities. To prevent any feelings of fear or anxiety among teachers due to the presence of an outsider, surveyors were instructed to compliment every teacher at the end of the class by saying that they teach very well. The questionnaire was not shared with any teacher or school official before the end of the study.

Details of items used in the classroom observation tool along with their source and pre-specified domains are in Table D.2. The list of items was finalized after intensive rounds of piloting and fieldwork. Items were selected based on their contextual relevance, objective nature (requiring binary or numeric responses), and ease of understanding for the observer. Items exhibiting ceiling or floor effects in the pilot data, as well as those that presented ambiguity regarding the construct being assessed, were removed following feedback from focus group discussions with observer teams.

Homework Notebook Review At the end of each classroom observation, surveyors collected notebooks of two random students in the classroom and responded to the below questions.

1. Was the notebook checked? (based on the presence of any tick marks by teacher) Yes/No

2. What was the degree of detail in feedback provided?
   - General feedback (e.g. overall comments like seen)
   - Specific feedback (question-wise comments)
   - No feedback

3. Was there any encouraging feedback provided? (e.g. Good/Excellent/Keep it up) Yes/No

Table D.2. Classroom Observation Items

Measure	Category	Source	Type
1. Textbook used	Use of materials and content	Stallings	Binary
2. Blackboard used	Use of materials and content	World Bank SDI	Binary
3. Teaching aid used	Use of materials and content	Stallings	Binary
4. Questions solved by teacher	Use of materials and content	Original	Numeric
5. Questions given for independent practice	Use of materials and content	Original	Numeric
6. Teacher gave homework	Use of materials and content	World Bank SDI	Binary
7. Number of students who asked questions	Classroom climate	World Bank SDI	Numeric
8. Teacher used an encouraging phrase	Classroom climate	World Bank TEACH	Binary
9. Teacher used a discouraging phrase	Classroom climate	World Bank TEACH	Binary
10. Teacher praised a specific student	Classroom climate	Muralidharan & Singh (2023)	Binary
11. Visited children individually	Effort at facilitating engagement	World Bank SDI	Binary
12. Made an attempt to engage backbenchers	Effort at facilitating engagement	Original	Binary
13. Number of students who were called by name	Effort at facilitating engagement	World Bank SDI	Numeric
14. Used local information to make learning relevant	Effort at making class accessible	World Bank SDI	Binary
15. Used local language for instructions	Effort at making class accessible	World Bank SDI	Binary
16. Teacher was audible	Effort at making class accessible	Original	Binary
17. Work on blackboard was visible	Effort at making class accessible	World Bank SDI	Binary
18. Pace was right (neither too fast nor too slow)	Effort at making class accessible	Original	Binary
19. Teacher was mostly standing	Demeanor	World Bank SDI	Binary
20. Teacher hit students	Demeanor	World Bank SDI	Binary
21. Teacher was rude	Demeanor	World Bank SDI	Binary
22. Teacher was smiling, laughing or joking with children	Demeanor	World Bank SDI	Binary
23. Started class with a stimulating question	Pedagogical practices	World Bank SDI (adapted)	Binary
24. Asked learners to demonstrate their understanding	Pedagogical practices	World Bank SDI	Binary
25. Asked questions that required students to recall information	Pedagogical practices	World Bank SDI	Binary
26. Asked questions that required students to apply information	Pedagogical practices	World Bank SDI	Binary
27. Asked questions that required students to use their creativity	Pedagogical practices	World Bank SDI	Binary
28. Summarized the lesson at the end of the class	Pedagogical practices	World Bank SDI	Binary
29. Gave feedback by correcting a mistake	Pedagogical practices	World Bank SDI	Binary

Notes: This table presents items used in the classroom observation tool used to capture intensive margin of effort by teachers. Trained observers observed teachers during a Math class and scored teachers on these items. Items were sourced using international instruments including World Bank TEACH, World Bank Service Delivery Indicators (SDI) and Stallings. Items marked as 'Original' were conceived by the author based on discussions with pedagogy experts and refined during field pilots prior to baseline data collection.

D.3. Additional Outcomes

Mental Health I administer the Center for Epidemiological Studies Depression Scale (Radloff, 1977) scale used in the mental health literature. The scale consists of 20 items that ask respondents to rate themselves on a 4-point Likert scale indicating the frequency with which they have experienced symptoms related to stress, burnout, and depression, including restless sleep, poor appetite, feeling lonely, etc. The score ranges on a scale of 0 to 60 and is categorized into no indication of a depressive disorder (< 15), mild to moderate depression (15-21), and severe/major depression (above 21).

Teachers' Expectations of their Students To capture teachers' expectations for different types of students in the classroom, teachers are asked to predict the scores of four students in class — one from each quartile of the score distribution.

Teachers' growth mindset To capture the growth mindset of teachers, I use the Implicit Theories of Intelligence scale (Blackwell et al., 2007). A score of 3 or less indicates a fixed mindset, a score of 4 or more is taken to indicate a growth mindset, while intermediate scores are considered borderline. I construct a binary indicator for whether the teacher had a growth mindset and use this as a measure of growth mindset. I also construct a standardized measure and use as a dependent variable.

Risk Preferences I measure risk preferences through the above multiple price list (MPL) task with 10 rows. Participants choose between a fixed payment and a lottery that yields 100,000 points with probability 0.5 and zero otherwise. The lottery remains the same in all rows while the safe payment increases. Earlier switching points indicate a lower certainty equivalent than later switching points. Note that this task was not real-stakes as it involved no real payment. It was used as part of the example tasks to train participants in working with the MPL format. This task was shown before the real-stakes belief-elicitation task and was conducted at both baseline and endline.

Table D.3. Multiple Price List for Risk Elicitation

Choice	**Option 1**	**Option 2**
Choice 1	Rs 5,000	50/50 chance between Rs 100,000 and Rs 0
Choice 2	Rs 10,000	50/50 chance between Rs 100,000 and Rs 0
Choice 3	Rs 20,000	50/50 chance between Rs 100,000 and Rs 0
Choice 4	Rs 30,000	50/50 chance between Rs 100,000 and Rs 0
Choice 5	Rs 40,000	50/50 chance between Rs 100,000 and Rs 0
Choice 6	Rs 50,000	50/50 chance between Rs 100,000 and Rs 0
Choice 7	Rs 60,000	50/50 chance between Rs 100,000 and Rs 0
Choice 8	Rs 70,000	50/50 chance between Rs 100,000 and Rs 0
Choice 9	Rs 80,000	50/50 chance between Rs 100,000 and Rs 0
Choice 10	Rs 90,000	50/50 chance between Rs 100,000 and Rs 0

Notes: This table presents hypothetical choices between a fixed payment and a lottery presented to teachers for elicitation of risk preferences, as part of a multiple price list procedure.

Appendix E: Online Appendix

E.1. Attrition

Attrition is defined as failure to complete any/all endlines conditional upon completing the baseline. 292 teachers completed the baseline out of which 263, 260, and 259 were observed across the first, second and third endline rounds respectively. Below, I investigate the magnitude and correlates of attrition and whether it confounds results in line with approach outlined in the pre-analysis plan.

Pooling across all endline rounds, attrition rates for the control and treatment groups are 12% and 9%, respectively. This difference is not statistically different. In table A2 in the main appendix, I also examine balance across indicators for missing at any round, and missing at all rounds. There is no differential attrition by treatment status.

Table A18: *Baseline characteristics of attritors and non-attritors*

	(1) Age	(2) Experience	(3) Female	(4) Education	(5) Certification	(6) Permanent
Missing for a given round	-2.350***	-0.890	0.084*	0.151***	-0.025	-0.603***
	(0.727)	(0.614)	(0.045)	(0.052)	(0.038)	(0.060)
R-squared	0.589	0.591	0.495	0.519	0.551	0.569
Observations	1168	1168	1168	1168	1168	1168

Notes: Column headers report the dependent variable in each specification, which is a teacher characteristic at baseline. Age and experience are reported in years, as of baseline in August 2022. Education is an indicator that takes the value one if the highest education is masters-level or above. Certification is an indicator that takes the value one if the teacher holds a teacher training diploma/certification. Permanent is an indicator for contractual status and takes the value one if the teacher is on a permanent contract. All specifications control for strata fixed effects and round fixed effects. Standard errors are clustered at the teacher level. Data is pooled across endline rounds and is at the teacher x endline level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A18 estimates whether attritors are different from non-attritors in terms of a vector of baseline characteristics – age, gender, education, experience, certification, and permanent (vs. contractual) status. Each column reports the coefficient from a regression of baseline characteristic on an indicator for attrition. I find that attritors are younger, more likely to hold a masters-level education, and less likely to be permanent teachers.

Table A19 estimates whether there is differential attrition by treatment status by regressing baseline characteristics on assignment indicator, restricting the sample to attritors only. I do not find evidence that baseline characteristics of attrited teachers in the treatment group are significantly different from those in the control group.[30]

Deviations from PAP: Due to a typo, the PAP mentioned that the specification investigating differential attrition will be restricted to attrition 'households', this was intended to mean attrition 'teachers'. The specifications in the PAP missed to include strata and round fixed effects, which I include in consonance with the design and analysis in the rest of the paper.

[30]Given that all teachers who attrit are contractual, the coefficient on the indicator for permanent status is zero, and the standard error and R-squared are undefined.

Table A19: *Baseline characteristics of attritors across treatment and control*

	(1) Age	(2) Experience	(3) Female	(4) Education	(5) Certification	(6) Permanent
Treat	-2.943	-2.367	-0.143	-0.155	-0.012	0.000
	(4.864)	(1.661)	(0.154)	(0.162)	(0.238)	(.)
R-squared	0.80	0.95	0.90	0.97	0.89	.
Observations	94	94	94	94	94	94

Notes: Column headers report the dependent variable in each specification, which is a teacher characteristic at baseline. Age and experience are reported in years, as of baseline in August 2022. Education is an indicator that takes the value one if the highest education is masters-level or above. Certification is an indicator that takes the value one if the teacher holds a teacher training diploma/certification. Permanent is an indicator for contractual status and takes the value one if the teacher is on a permanent contract. All specifications control for strata fixed effects and round fixed effects. Standard errors are clustered at the teacher level. Data is pooled across endline rounds and is at the teacher x endline level. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.