Changing Mindset to Foster Non-cognitive Skills and Academic Achievement: Evidence from a field experiment in Pakistan

December 17, 2021

Abstract

We conduct a randomized control trial to investigate the effect of a soft-skills intervention that highlights the role of beliefs about perseverance on non-cognitive skills, effort, and academic performance of college students. We have a sample of 366 undergraduate students from women-only and public colleges in a major metropolitan city in Pakistan. We find that a brief discussion stressing perseverance and a constructive interpretation of failures led to increased level of grit among treated students one year after the intervention. However, we find no significant increase in hours of study and college attendance after one year or test scores two and nine months after the intervention. Our findings suggest that soft-touch interventions may impact cognitive skills, but they can be insufficient for improving academic achievement.

JEL codes: D90, I21, I28, I24.
Keywords: Academic performance, Non-cognitive skills, Grit, Women, College students.

1 Introduction

Human capital accumulation is considered essential for improved later life outcomes and intergenerational welfare, thus improving student learning which has long been an important concern in development policies. An extensive literature on improving academic enrolment and learning documents positive but largely modest effects of supply and demand-side interventions that are implemented at considerable financial cost (Glewwe and Miguel, 2008; McEwan, 2015; Angrist et al., 2020). Costly programs are challenging to sustain in many developing countries with limited resources, highlighting the need for cost-effective solutions for improving learning outcomes. An emerging strand of behavioral and psychological literature suggests one group of interventions that aim to improve student perseverance and academic performance levels but can be implemented at a low cost. These interventions seek to change individual beliefs about the value of effort, inculcate determination and perseverance, and lead to improved learning (Heckman and Rubinstein, 2001; Paunesku et al., 2015; Humphries and Fabian, 2017; Alan et al., 2019; Bettinger et al., 2018). While evidence suggests non-cognitive and performance can be enhanced through discussion and self-regulated exercises (Alan et al., 2019; Durlak et al., 2011; Kautz et al., 2014), the implicit assumption of existing work is that underlying beliefs about the value of effort and perseverance are fixed in adults and can only be changed in childhood (Outes-Leon et al., 2020; Ganimian, 2020). However, in many developing countries, enrolment and learning gaps are persistent and larger for older students (Kaffenberger, 2021), making a policy relevant question to investigate the potential impact of interventions targeting non-cogntive skills and performance of adult students.

In this study, we use an individual-level randomized control trial to investigate the effect of a low-cost intervention that aims to foster grit, effort, and learning outcomes among adult college students in Lahore, Pakistan. We borrow from the literature on supportive psychology and use a soft-touch, easily scalable intervention tested with young children (Alan et al., 2019; Blackwell et al., 2007; Polley, 2018). The intervention consists of one-to-one discussions with students on the role of effort in enhancing skills and achieving goals, the importance of a constructive interpretation of failures, and perseverance to achieve their goals. In addition, we implement a 'placebo' discussion consisting of scientific facts about the human brain in the control group. Finally, we test if the treatment discussions led to improvements in psychometric measures of grit, growth mindset, self-efficacy, and measures of effort and test performance.

We find student engagement to be higher amongst the treated group immediately after the discussion. One year after the intervention is first delivered, treated students report 4.0 - 4.2 % higher level of 'grit' than the placebo group students. However, the treatment discussion does not lead to higher effort, as measured by increased hours of study or attendance, nor does it improve test scores measured two and nine months after the intervention. An analysis of results by student characteristics reveals encouraging insights for identifying sub-groups of students for which the intervention may be particularly effective. For instance, the treatment intervention was more effective in improving grit among students that were less likely to rank academic career goals as important, those who score lower on cognitive tests (Raven's matrices) and less competitive students. We also find that though grit improves after one year for treated students in their first year of college, it is not accompanied by an improvement in effort or test performance. Results suggest that an increase in grit without subsequent improvement in dedication to academic goals can fail to improve academic performance.

We contribute to the literature in three key ways. First, our study adds to the growing literature that tests the effect of soft-touch interventions on non-cognitive skills, such as grit and growth mindset, and academic performance.[1] Existing behavioral work focuses on changing the beliefs of young school children and show that it is possible to make children more resilient and improve their performance in school (Alan et al., 2019; Outes-Leon et al., 2020; Paunesku et al., 2015; Yeager et al., 2016, 2018). Related literature in psychology proposes that non-cognitive skills, such as perseverance, can predict individual success and that it may be possible to inculcate these skills not only in children but also among adults (Duckworth, 2007). In this study, we combine these two complementary strands of literature. We test a motivational intervention with adult, college students, and find that theory-based interventions to remedy motivational issues can indeed have a long-term change in mindset. However, soft-touch interventions may be insufficient to change the academic performance of adult students.

Second, we also contribute to a small group of studies investigating the potential role of interventions targetting student 'mindset' and performance in developing countries (Ganimian, 2020; Outes-Leon et al., 2020). Supply and demand-side constraints affecting school, teacher, and household inputs often result in poor academic performance and high dropout rates in these countries (Dillon et al., 2017; Banerjee et al., 2010). A recent review of 150 educational interventions designed to target teacher quality, school supply, child health, and attendance in 46 countries reveals that about half of the interventions had no impact on learning-adjusted years of schooling, while the total cost per child of interventions that had a positive impact ranged from USD 30 at the first cost quintile to USD 813 at the fifth cost quintile (Angrist et al., 2020). High costs put further strain on resources in developing world contexts, and even programs with positive impacts can be challenging to sustain. In contrast, interventions that aim to improve student non-cognitive skills and self-belief have the potential to be easily-scalable and cost-effective in developing world contexts. For instance, the intervention we tested was a 30-minute discussion implemented at a low cost of only USD 4 per respondent and was designed to be a soft-touch intervention that can be easily

[1] For a detailed review of recent studies, see (Paunesku et al., 2015; Yeager et al., 2016, 2018; Sriram, 2014; Yeager et al., 2014; Bettinger et al., 2018; Ganimian, 2020; Islam et al., 2020)

incorporated into the regular curriculum at colleges.

Third, to the best of our knowledge, this is the first study that focuses on women. As we discuss in Section 2.4, our decision to concentrate on women was partly driven by socio-cultural and logistical concerns. However, women are a relevant sample to explore the effects of interventions focusing on non-cognitive skills and performance for two reasons: one, women in the developing world often have to contend with unequal access to education and face challenging situations that may require a high degree of persistence to succeed. Two, existing inequalities perpetuate later in life, determining their decision-making power within the family, access to work, income levels later in life (The World Bank, 2018). Non-cognitive skills can, therefore, have potentially long-lasting effects on the well-being of women and their children (Lindqvist and Vestman, 2011; Duckworth, 2009). Note, these effects may exist for men and are relevant for individual welfare, but findings with younger students indicate that the effects are larger for females than for males (Polley, 2018; Islam et al., 2020; Bettinger et al., 2018).

Pakistan is a relevant setting for exploring the role of cost-effective interventions for improving student performance and well-being in a developing world context. Similar to other developing countries, enrollment rates at the primary level have increased dramatically to over 90% (UNESCO, 2009), however, enrolment rates are substantially lower at the higher education levels. For instance, the gross enrolment rate in higher education is 12% in Pakistan, 13% in Nepal, 16% in Bhutan, and 29% in India (The World Bank, 2020). In addition, degree completion rates are low, and there is substantial gender inequality in access and enrollment. For instance, in Punjab, the most populous province of Pakistan, only half of the students pass the qualifying exam and the gender parity index in enrollment stands at 0.85 (The University of Punjab, 2019). This highlights the need for innovative interventions to improve learning outcomes at higher levels of education in the developing world context, particularly for female students.

The rest of the paper is organized as follows. In Section 2, we present the experiment design and details of implementation. Section 3 summarizes the empirical strategy and Section 4 presents the results. Section 5 discusses various tests of robustness of the analysis and Section 6 concludes.

2 Experiment design and implementation

2.1 Conceptual framework

Psychologists hypothesize that beliefs about the results of effort can be an important determinant of performance in different domains (Dweck, 2006; Duckworth et al., 2007; Blackwell et al., 2007; Costrell, 1994; Stinebrickner, 2008; De, 2010). In their seminal paper, Blackwell et al. (2007) present a 'motivational model of achievement' where individuals focus on 'learning goals' rather than the quality of performance. These individuals believe ability can be further developed through learning and are more likely to persevere in challenging times. In contrast, those who view ability as innate and intelligence as fixed are less likely to persist in challenging tasks or in the face of setbacks. Blackwell et al. (2007) argue that such thinking can have important consequences for students who are facing 'a sustained challenge at critical points in their lives. They show via two studies conducted with secondary school students that an individual's theory of intelligence is malleable and can affect performance in real-world contexts. In this section, we borrow from the seminal work by Blackwell et al. (2007) and the simplified model adopted by Bettinger et al. (2018), to provide a stylized framework that informs our experiment design.

Consider a situation in which a student has to decide how much effort to exert. The student has certain beliefs, p, about her ability to learn and expects benefits from exerting effort, $p.B(e)$ (where $B' > 0$ and $B'' < 0$). The student also incurs an opportunity cost of effort, $C(e)$ for instance, the time she could have spent instead with her siblings or going out with friends (such that $C'(e) > 0, C''(e) < 0$). A utility maximising student exerts effort e^*, where $C'(e^*) = p.B'(e^*)$.

Under this framework, an individual's optimal level of effort is an increasing function of belief about the usefulness of effort. That is, $\frac{de*}{dp} > 0$. Individuals with a growth mindset believe "trying harder makes you smarter" (Blackwell et al., 2007). They exhibit grit in challenging circumstances and exert effort to achieve their goals. By contrast, those who believe that everyone is born with certain cognitive abilities, do not see the value of exerting effort. It follows that individuals who believe in the value of persistence and grit will exert more effort to achieve their goals than individuals who think outcomes only depend on innate, fixed ability. Further, findings from psychology studies indicate that beliefs about the usefulness of effort, p, can be influenced. It follows that if beliefs about effort are changed, this can change the optimal levels of effort, e^*, exerted by an individual (Blackwell et al., 2007). This simple framework provides the following main hypotheses of the study:

Hypothesis 1: Beliefs about the usefulness of perseverance and effort are malleable.

Hypothesis 2: A change in beliefs about the usefulness of effort, will alter a student's effort and academic performance.

Change in beliefs can be captured by a change in non-cognitive skills, such as the self-perceived ability to persevere, having a 'growth' mindset and believing one self to be capable of handling difficult tasks. We would also expect individuals who believe effort to be useful to exert more effort in academic tasks, e.g., by dedicating more time to studying, which may be reflected in improved academic performance (Alan and Ertac, 2015; Bandura, 1978; Duckworth, 2007; Roberts et al., 2007; Borghans et al., 2008). In what follows, we describe the design of the intervention to encourage student's belief in the ability to learn and grow; and the outcomes we measure to test the two hypotheses.

2.2 Intervention design

We adapt protocols and materials developed in seminal work in supportive psychology by Blackwell et al. (2007) and develop a soft touch, low-cost intervention for the local context. The intervention discussion is complimented by visual aids, discussions, and flyers to change beliefs about effort. The intervention materials have been adapted by Alan et al. (2019) for primary school students in Istanbul. We adapt this material to suit sample age. The treatment consists of guided discussions that highlight the human brain's plasticity against the notion of innate ability and stresses the role of effort in achieving goals. We emphasize a constructive interpretation of failures and attempt to motivate students to persevere in pursuing short and long-term goals. We reiterate our message by providing examples, such as how young children learn the language and basic skills and how everyone has the capacity to improve performance through sustained effort and practice.

The placebo content, also adapted from (Blackwell et al., 2007), comprised of discussions on the basic structure of the brain. Appendix A1 provides detailed information on treatment and placebo discussions. Both treatment and placebo discussions were conducted individually and face-to-face, between a student and an enumerator. The discussions lasted 30 minutes, on average.

After the treatment and placebo discussions, all students in our sample participate in a goal-setting activity to push all students to think about their academic and professional goals. Students were asked to outline their short and long-term goals. They were informed that 'SMART' - Specific, Measurable, Ambitious, Realistic, Time-bound, goals are more likely to be followed and achieved (Dobronyi and Petronijevic, 2019; Morisano et al., 2010).

The design by Blackwell et al. (2007) takes careful consideration of potential confounders such as experimenter demand effects: both treatment and placebo students participate in discussions of equal duration, conducted face-to-face by the same team of enumerators. The topic of the placebo discussion is similar to that of the treatment in that they both discuss the capabilities of the human brain. The difference is that the treatment discussion stresses the ability of the brain to grow, while the placebo provides a scientific summary of the functions of the brain. The treatment is designed to be more engaging than the placebo discussion. Indeed, in questions asked immediately after the intervention discussions, we find greater student engagement in the treatment than the placebo sample. Compared to the placebo group, treated students were 7 percentage points more likely to have found the discussion to be interesting, 16 percentage points more likely to report the discussion left an impact, 21 percentage points more likely to think that the information was 'helpful', 31 percentage points more likely to report that information will help in thinking carefully about the next steps for their goals, and 12 percentage points more willing to learn more about the topic in the future.[2] Table A1 in the Appendix provides a summary of the post-intervention 'engagement' questions. All differences in engagement measures for placebo and treatment samples are statistically significant ($p < 0.01$).

Next, literature recommends recording immediate change in the non-cognitive skills of treatment and placebo groups after the intervention, as another test of the treatment discussion content impacting students differently at the onset. If the treatment discussion had any impact, we would expect different measures post-intervention for the treatment and placebo group. We test for differences in non-cognitive skills using validated scales that have been shown to strongly predict individual performance in incentivized tasks to elicit the same measures (Burnette et al., 2013; Yeager et al., 2016; Alan et al., 2019). We discuss these measures and how they change immediately after the intervention in detail in Sections 3.1 and 4 respectively.

The intervention cost USD 4.08 per respondent to implement. This comprises of a fixed cost of content development USD 0.24 and field costs associated with implementing the intervention of USD 3.84 per respondent.[3] The development costs are low, primarily because of readily available content by Blackwell et al. (2007), that only has to be adapted to the local context. Part of these unit costs, such as those incurred in piloting and training, can be expected to be fixed and per unit costs will decrease for larger samples. Online Appendix OA1 provides details of costs incurred.

2.3 Implementation timeline and protocol

We recruit the study sample as follows: first, we request the administration of 10 colleges in our sample to provide a list of students. We randomly draw out a sample of students from each college, proportional to the size of the total student body, allowing for a surplus of 15% replacement sample per college to allow us to reach targeted amount of surveys for each college.

Due to logistical concerns, we conduct both the baseline and the treatment discussions in the same face-to-

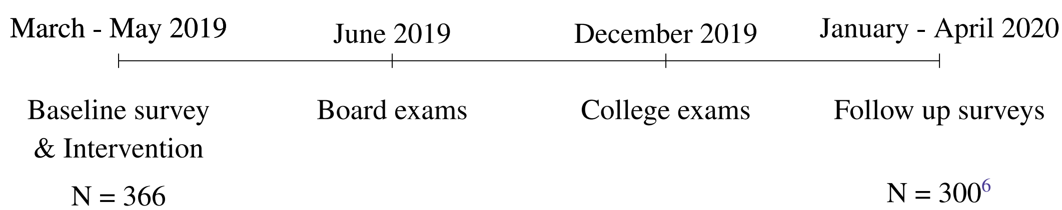
[2] We find good recall of treatment and placebo discussions in recall quizzes conducted on the phone with 40% (N=148) of the sample nearly three months after the initial intervention implementation. We found recall in treatment to be better than in the placebo group: students in the placebo group correctly answered 2.9 out of 5 quiz questions correctly, compared to 4.7 out of 5 questions asked of the treated group ($p = 0.000$).

[3] We use 140 PKR = 1 USD exchange rate as of March 31 2019, for conversion here and throughout the paper.

face interaction with student. We first conduct a short baseline student interview to collect information on individual and household level characteristics of the student. Once the baseline interview is complete, the survey software assigns each respondent to the treatment or placebo group, with equal probability.[4] The enumerator conducts the discussion assigned to the student. This is followed by the goal-setting exercise and a 'post-intervention' short survey, both of which are conducted with the treatment and placebo groups to measure student perceptions and non-cognitive skills immediately after the intervention. Note that to reduce experimenter bias from influencing student response on the baseline survey, enumerators do not know the treatment status of the student at the time of the baseline survey.

The timeline of our field activity is given in Figure 1. We collected baseline data and implemented the treatment and placebo interventions between March and May 2019. We collected administrative data on student performance in end-of-year, standardized board examinations held two months after our intervention, and internal college exams nine months after our intervention.[5] We conducted a follow-up survey later in January till April 2020.

Figure 1: Timeline of Field Activity



Follow-up surveys were interrupted when all higher education institutions were closed by the provincial government on March 13, 2020, in response to the global COVID-19. As a result, the follow-up data was collected using a mix of in-person and phone interviews. Out of 300 interviews at follow-up, 226 were conducted in-person before the lock-down, and 74 interviews were conducted on the phone between 14 March - 10 April 2020. Online Appendix section OA3 provides details on the regulatory framework surrounding the lock-down.

[4] All data was collected electronically. Treatment assignment is done at the individual level using the random number generator available in the survey software, SurveyCTO (`www.surveycto.com`). Selection to the treatment and placebo group is done with replacement.

[5] We planned to collect information on standardized board exams one year after the implementation. However, scheduled board exams were not administered due to country-wide closure in March 2019 due to the COVID-19 pandemic.

2.4 Study context

Our sample consists of students from women-only public colleges in Lahore, Punjab. This sample, which often has to struggle with continuing education due to socio-cultural norms, is a relevant sub-sample for an intervention designed to encourage effort and persistence in their academic pursuits.

Enrolment rates in Punjab at the undergraduate level average at 12%, with enrolment rates for females particularly low at only 85% than that of the males. Four out of five students at this level are enrolled in public colleges (PES, 2016), where tuition costs are much lower.[7] Due to socio-cultural norms, public educational institutes for adults are usually separate for men and women.[8]

Lahore is the largest metropolitan city in Punjab, Pakistan. There are 63 public undergraduate colleges, 39 of which are women-only, catering to a population of 11 million people (Census, 2017; PDS, 2018). The pass rate at the undergraduate level remains consistently low, with only half of the female students enrolled in undergraduate public degree colleges in Lahore, completing their degree in 2019 (PDS, 2017, 2018). Of the women who appear for their exams, less than a fifth score more than 60%, presenting a dismal picture of substantial learning gaps (The University of Punjab, 2019; PDS, 2018). These gaps have persisted despite recent government initiatives to encourage enrolment and academic performance (Fiszbein and Schady, 2009).

2.5 Description of the sample and treatment balance

Our sample consists of women enrolled in the first, second, and third year of 10 public colleges in Lahore, enrolled in Arts and Social Sciences degrees. Sample characteristics are summarized in Table A2 in the Appendix. The average respondent comes from households where parental education and labor force participation is unequal - fathers are both more educated and more likely to be working than the students' mothers in this sample. More than half the respondents come from families where the primary earner, the father, is salaried; one out of four come from families where the father has his own business and is self-employed. The average monthly income for the household was not reported in 108 instances; for those who respond, the average income per month is approximately USD 254, and is similar to the monthly household income of an average household in Punjab, Pakistan[9]

[7] Tuition fees in public colleges are approximately USD 22 per annum. The annual tuition in private colleges can range anywhere from USD 200 to USD 6000.

[8] Due to gender-segregated colleges, it was not possible for us to recruit male and female students enrolled in the same college. We concentrate on students from one kind of college to optimize statistical power - public and for women only.

[9] The average household income in Punjab, Pakistan, in 2018 was approximately PKR 43,000 (USD 307) (Pakistan Bureau of Statistics, 2019).

Respondents are 19 years of age and unmarried, on average. 91% plan to continue education after completing their current degree, and 92% want to join the labor force eventually. More than a third of the students report graduating from their previous degree in the 'first division' (more than 60% of total marks). A large proportion of the sample (71%) report discussing college assignments (homework, as it is locally known) and their education goals (93%) with their parents. Sample students can answer 4 out of 10 Raven's matrix questions correctly and were moderately satisfied with their current academic performance.

The study sample is balanced in all characteristics, other than age, the year of enrollment, and self-reported demand for paid work after graduation. It is worth noting that though the difference in average ages between the treated (19.2 years) and placebo group (19.5 years) is statistically significant at the 10% level, it is economically small. Furthermore, a test of joint significance shows that baseline characteristics are jointly insignificant in predicting treatment status ($p - value$ of joint significant F-test = 0.11). In Section 5 we test if our results are robust to the inclusion of individual characteristics.

3 Empirical strategy

3.1 Outcomes

We estimated the impact of our treatment discussion on two main sets of outcomes: (i) non-cognitive skills and (ii) academic effort and performance. To elicit non-cognitive skills, we make use of validated scales that have been found to be good predictors of behavior in incentivized tasks measuring the same outcome. Details of the three validated scales are available in Online Appendix OA2. We measure three different but related skills - perseverance, belief in the malleability of intelligence and self-efficacy. As discussed in section 2.2, we measure non-cognitive skills twice: first, we test if the treatment discussions successfully changed the students' initial non-cognitive skills immediately after the intervention. Second, we test if the intervention had a longer-term effect on non-cognitive skills, using data collected one year after the interventions were first implemented. We measure the following non-cognitive skills:

Grit: We measure perseverance using the 12-item scale for 'grit' developed by Duckworth et al. (2007) that elicits respondents' perceptions about their ability to overcome setbacks and persist in difficult tasks.

Growth mindset: We measure if the respondent believes that intelligence can grow by learning from your mistakes, using the 15-item scale 'Fixed vs. Growth mindset' scale developed by (Blackwell et al., 2007).

Self-efficacy: We measure respondents 'self-efficacy' using the 10-item scale by (Schwarzer and Jerusalem, 1995) to measure self-perceived ability to manage complex problems and deal with unexpected events.

We correct for multiple hypothesis testing in two ways: i) we create an Anderson index using variables (i) - (iii) for a *Mindset Index* out of all non-cognitive skill measures; and (ii) by calculating the sharpened $q - values$ that control for false discovery rate (FDR) (Anderson, 2008).

To understand if the treatment interventions led to changes in effort, we collect information on several proxies such as the time the student spends studying in college or at home. In addition, increased effort could manifest in improved academic performance, reflected in student test scores. Accordingly, we measure the following outcomes:

Hours of study at home: We collect data at baseline and end line on the number of hours per day a student reports studying or finishing homework at home in an average week.

Attendance in college: We collect data at baseline and end line on the number of days in the past week that the student attended college.

Test score percentage: Higher effort may lead to improved performance in tests and exams. We collected administrative data from sample colleges on two different exams after the baseline survey. First, on centralized board administered exams approximately two months after the intervention. We obtain marks scored by each respondent in the exam and then calculate *test score percentage*, which is a proportion of the total marks.[10] Board exams were not held the following year, in the summer of 2020 due to COVID-19 closure. Instead, students were promoted to the next year based on their performance in exams in December 2019. These were college administered exam, conducted a few months before the lock-down and approximately nine months after the treatment intervention. These serve as the second measure of academic performance for our study.

Table A3 in the Appendix provides the summary statistics of these variables for the overall sample. Students spent nearly three hours per day studying at home and four days in college at baseline, with little change over the study period. We see a slight improvement in the 'grit' scale between the first and final follow-up survey and a moderate decrease in the growth mindset scale. Students score an average of 48% in their board exams after two months, and 40% in college administered exams nine months after the intervention was implemented.[11] Next, we describe the empirical methodology used for testing if these outcomes differ by whether the respondent took part in the treatment or placebo discussions at baseline.

[10] The Board of Intermediate and Secondary Education (BISE) and the University of Punjab are responsible for conducting these standardized exams, at the same time, for all students. The total marks depend on the year the student is enrolled in: first-year students are marked out of a total of 550 marks, second-year students are marked out of 1100, and third and fourth (final) year students are marked out of 400.

[11] Out of the total sample, 36 students did not appear in the board exams at two months, colleges could not confirm board scores for another three students; and 34 did not appear in the college administered exams at nine months.

3.2 Estimation specification

We test the effect of our intervention on outcome variables specified in 3.1. For each outcome, we run the following basic specification:

$$y_{it} = \beta_0 + \beta_1 Treated_i + \beta_2 Post_{it} + \beta_3 Treated_i * Post_{it} + \epsilon_{it} \tag{1}$$

where y_{it} is an outcome variable, $Treated_i$ is a dummy variable capturing treatment status, $Post_t$ is a dummy variable for the follow-up time period, $Treated_i * Post_t$ is the average treatment effect in the longer run, and all errors are clustered at the individual level. For non-cognitive skills (grit, growth mindset, and self-efficacy scales) for which we have data only after the intervention, the coefficient on $Treated_i$ will show us the immediate impact of the intervention measured in the post-intervention survey, while the coefficient on $Treated_i * Post_t$ provides an estimate of treatment effects approximately one year later. We have data on test scores 2 and 9 months after the intervention, not at baseline. For these outcomes, we run a simplified estimation for each test period, without needing to include indicators for $Post_t$ and $Treated_i * Post_t$ since all scores are only collected at followup.

In results discussed in section 4, we show estimates from equation (1) both with and without controlling for college and enumerator fixed effects. We discuss the robustness of our results when controlling for individual level characteristics in Section 5.

3.3 Attrition

We were able to interview 300 out of 366 students from the baseline sample at follow-up – less than half (44%) of the attriters were from the treated group. We check for differential attrition between the treatment and placebo group by coding a dummy variable for whether individual i attrited for the follow-up survey and then test if attrition is significantly related to baseline covariates of interest and to treatment status.

As shown in Table A4, attrition does not vary by treatment status. Attriters vary by level of competitiveness and degree of satisfaction with their academic performance. In section 5, we show robustness of our results by (i) repeating all main analysis using controls selected by the 'post-double-selection' with LASSO (Belloni et al., 2014a,b) and (ii) estimating the upper and lower bounds of the estimates using (Lee, 2009) bounds.

4 Results

We test if the treatment discussion impacts non-cognitive skills, effort and test performance. Results are provided in Tables 1 - 3. As can be seen by the coefficient on $Treated$, the treatment intervention leads to a small and moderately significant decrease in the 'grit scale' measured immediately after the intervention (Table 1, columns 1 and 2). That is, respondents in our treatment group first record lower levels of grit than the placebo group immediately after the intervention discussion. We also see a small increase in the average levels of the 'growth mindset' among treated respondents immediately after the intervention discussion (column 3); however, this increase is not robust to the inclusion of college and enumerator fixed effects (column 4). Combined with higher levels of engagement with the treatment than the placebo discussion reported in Section 2.2, these results are suggestive evidence of the treatment discussion being able to affect student non-cognitive skills at the outset.

Table 1: Long term impact of treatment on non-cognitive skills

	Grit (1)	Grit (2)	Growth mindset (3)	Growth mindset (4)	Self efficacy (5)	Self efficacy (6)	Mindset index (7)	Mindset index (8)
Treated	-1.027	-1.112	1.192	0.987	0.058	0.163	0.029	0.011
	(0.654)	(0.654)*	(0.692)*	(0.678)	(0.546)	(0.536)	(0.107)	(0.109)
Post	-0.577	-0.670	-2.750^{AAA}	-2.798^{AAA}	0.074	0.104	-0.323	-0.333
	(0.528)	(0.521)	(0.714)***	(0.710)***	(0.472)	(0.479)	(0.090)***	(0.092)***
Treated*Post	1.648^{A}	1.727^{A}	0.244	0.303	-0.138	-0.221	0.165	0.170
	(0.747)**	(0.744)**	(0.976)	(0.977)	(0.684)	(0.687)	(0.133)	(0.134)
Mean	41.125	41.125	48.660	48.660	32.938	32.938	-0.143	-0.143
N	666	666	666	666	666	666	666	666
R^2	0.005	0.034	0.043	0.092	0.000	0.048	0.018	0.050
FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. Treated*Post is the average treatment effect of our treatment intervention. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. Values in squared brackets represent sharpened $q - values$ that control for false discovery rate (FDR) (Anderson, 2008). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Adjusting critical values following the approach by Benjamini and Hochberg, 1995: ^{AAA}Significance at 1% level, ^{AA}Significance at 5% level, ^{A}Significance at 10% level.

Next, we look at the longer run impacts of the treatment and observe a significant and positive impact on grit levels after one year. Grit levels are 4.0 - 4.2 % higher for the treated sample than the average scores for the placebo sample over the same period. We do not see similar impacts on growth mindset or self-efficacy levels. This is suggestive evidence of a longer term impact of the treatment discussions on certain non-cognitive skills reflecting persistence and grit.

Table 2: Impact of treatment on measures of effort

	Hours study (1)	Hours study (2)	Days attend (3)	Days attend (4)
Treated	0.077	-0.007	-0.029	-0.082
	(0.188)	(0.175)	(0.188)	(0.180)
Post	0.039	-0.002	0.179	0.184
	(0.149)	(0.148)	(0.188)	(0.189)
Treated*Post	-0.053	-0.007	0.119	0.110
	(0.209)	(0.208)	(0.251)	(0.251)
Mean	3.366	3.366	4.766	4.766
N	666	666	666	666
R^2	0.000	0.105	0.005	0.064
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. 'Hours study' is the number of hours students report studying at home. 'Days attend' are the number of days the student attended college in the most recent full week. 'Treated*Post' is the average treatment effect of our treatment intervention. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Adjusting critical values following the approach by Benjamini and Hochberg, 1995: ^{AAA}Significance at 1% level, ^{AA}Significance at 5% level, ^{A}Significance at 10% level.

We test if the treatment discussion impacted the effort exerted by students. The treatment discussions do not lead to students' spending more time studying at home per day or attending college more regularly (Table 2). In line with this result, we do not detect any significant improvement in student performance in either board or college exams (Table 3). On average, students in the placebo sample score 49% after two months in the board exams and 40% at nine months in the college exams. Treated students scored approximately 0.01 - 0.02 percentage points lower than placebo students two months after the intervention (Table 3, columns 1 and 2). However, this decrease is statistically insignificant. At nine months, much of

Table 3: Impact of treatment on exam performance

Time after baseline	2 months		9 months	
	Score %	Score %	Score %	Score %
	(1)	(2)	(3)	(4)
Treated	-0.006	-0.015	-0.003	0.001
	(0.016)	(0.015)	(0.025)	(0.023)
Mean	0.488	0.488	0.403	0.403
N	327	327	332	332
R^2	0.000	0.136	0.000	0.188
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. 'Score percent' are the marks scored in examination end-of-year and after nine months of our intervention as a percentage of the maximum (full) marks allowed for that enrollment year. These were measured 2 and 9 months after the baseline activities and are present after our intervention. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Adjusting critical values following the approach by Benjamini and Hochberg, 1995: ^{AAA}Significance at 1% level, ^{AA}Significance at 5% level, ^{A}Significance at 10% level.

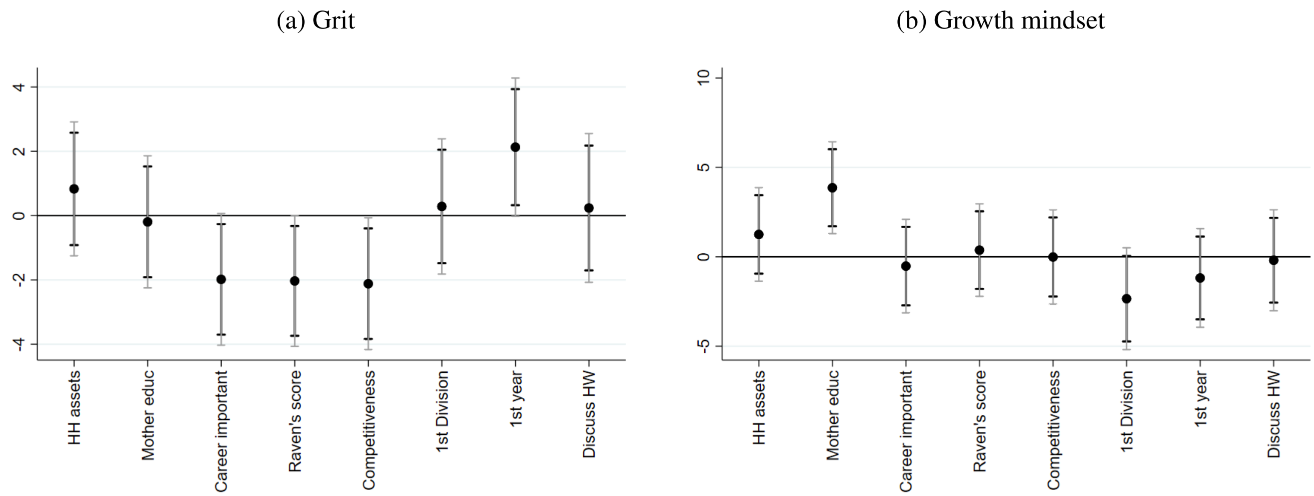
the variation in test scores observed at two months by treatment status disappears: the coefficient is close to 0 as well as statistically insignificant (Table 3, columns 3 and 4).[12]

4.1 Heterogeneity in effects

To summarize, we find average treatment effects on students' grit level over the long run, but no improvement in student effort or performance measures. To shed light on possible mechanisms, we test if these results vary for sub-sample of students. We test heterogeneity in two steps: first, we run a saturated regression including several covariates by which we expect treatment effects to vary and their interaction with treatment status. Results are given in appendix tables D1- D2. We find certain covariates to have relatively large coefficients, such as whether the student is enrolled in the first year, if the student ranks career goal to be important, their degree of competitiveness, the education level of their mothers, and their cognitive skills as measured by Raven's score to have potentially large effects. In addition, due to a small sample, we may be underpowered to detect several large heterogeneous effects in a saturated regressions. Therefore, in the second step, we test for heterogeneity considering each covariate separately. Figures 2 and 3 summarize the observed heterogeneity in treatment effects.

[12] These effects are equivalent to a decrease of between 0.04 and 0.11 standard deviations at two months and 0.00 and 0.01 standard deviations at nine months. Standardized results are presented in Appendix Table A7.

Figure 2: Heterogeneity in treatment effects on grit and growth mindset



Note: These panels display heterogeneity in treatment effect on percentage scored in 'Grit' and 'Growth Mindset' scales. The point estimate is from an OLS regression of 'Score (%)' on baseline values of each dimension of heterogeneity, 'H', '$Treated$', '$Post$' and interactions of the three run separately for each dimension of heterogeneity. The panels show point estimates coefficients on $Treated * Post * `H'$ for the dimensions specified on the $x - axis$. Spiked lines provide the 90 and 95% confidence intervals, in black and grey, respectively. See Tables D4 and D5 in the Online Appendix for detailed results and definition of covariates presented in the graphs.

Results suggest the existence of important moderators of treatment effects on non-cognitive skills based on individual characteristics. Grit decreases with students' raven's score, the importance they give to careers after graduating, and the level of competitiveness (Figure 2a). We also see suggestive evidence of the average effects on grit being driven by students enrolled in the first year. In addition, treatment led to a large increase in the growth mindset scale of treated students with educated mothers.[13]

Treatment effects on measures of effort and test scores across sub-samples are generally insignificant (Figures 3a - 3d) with one notable exception. Compared to first-year students in the placebo sample, treated students report lower amounts of time spent studying at home (Figure 3a), with a corresponding marginal though statistically insignificant decrease in their scores at two months (Figure 3c). Note, this sub-sample had also reported higher levels of grit after one year (Figure 2a). While we cannot claim causality in this analysis, they indicate interesting avenues for future investigations. Literature provides several insights to understand the patterns in treatment impacts we observe, which we discuss next.

[13] Similarly, we see significant effects on self-efficacy and the mindset scale among students with educated mothers. Results are available in Online Appendix Section OA4.

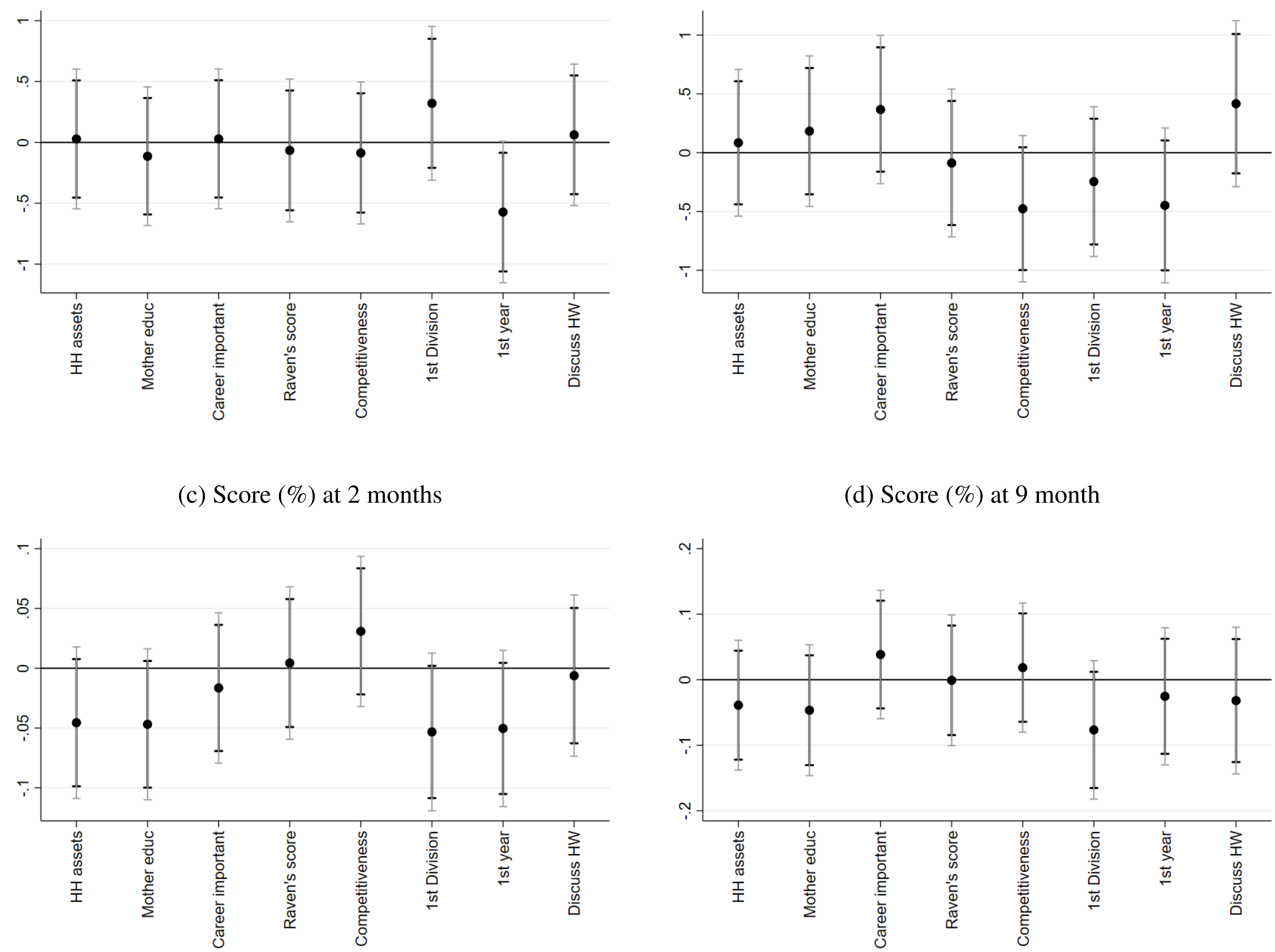
Figure 3: Heterogeneity in treatment effects on effort and test performance

(a) Hours of study (home) per day

(b) Days college attended per week

(c) Score (%) at 2 months

(d) Score (%) at 9 month



Note: These panels display heterogeneity in treatment effect on hours of study per day at home (Panel 3a), days in the last week that the student was present in college (Panel 3b), and percentage scored in exams 2 (Panel 3c) and 9 months (3d) from baseline. The point estimate in Panels 3a and 3b are from an OLS regression of each outcome on baseline values of each dimension of heterogeneity, 'H', indicators for time and period, and displays the coefficients on $Treated * Post *$ 'H' for the dimensions specified on the $x - axis$. Similarly, panels 3c and 3d show point estimates coefficients on $Treated *$ 'H' for the dimensions specified on the $x - axis$. Spiked lines provide the 90 and 95% confidence intervals, in black and grey, respectively. See Tables D8, D9, D10 and D11 in the Online Appendix for detailed results and definition of covariates presented in the graphs.

4.2 Discussion

Similar studies have been conducted with elementary school students in Istanbul (Alan et al., 2019), secondary school students in Peru (Outes-Leon et al., 2020) and Argentina Ganimian (2020), and high school students in Norway (Bettinger et al., 2018). In line with our findings, this literature finds impact on student non-cognitive skills. For instance, Alan et al. (2019) find grit increases by 0.28 - 0.35% among 4^{th} grade students in Istanbul, while effects on the growth mindset scales are insignificant. Students in Norway score 0.29 S.D more on an objective measure of determination to take on challenging tasks (Bettinger et al., 2018), which is very similar to the 0.25-0.27 S.D improvement in level of grit estimated in this study. However, contrary to our results, Alan et al. (2019) and Bettinger et al. (2018) find this change in non-cognitive skills is accompanied by an increase in objective test scores. It is also worth noting that in most studies that show an improvement in test scores, effort (e.g., hours of study or attendance) also increases (Alan et al., 2019; Polley, 2018). When efforts does not increase, test scores are unlikely to improve.[14] This is in line with the results of this study, where an improvement in grit levels is not accompanied by an increase in hours of study and college attendance, and we see no improvement in test scores. Recent meta analyses of psychological literature have explored why a change in non-cognitive skills, specifically grit, may not be accompanied by improvement in academic effort and performance (Jachimowicz et al., 2018; Credé et al., 2017). An emerging hypothesis is that while grit may emphasize the need to persevere, it does not capture an individual's intensity of focus, or 'passion' for a task (Dahl, 2016; Credé et al., 2017), when it is the latter the determines how important a long-term goal is to the individual, providing the impetus essential for changing behavior (e.g, increase effort). According to this theory, it may take both grit and dedication to academic goals to increase in order to improve performance in academic tasks.

This may explain why in this study the treatment only prompted higher levels of grit without accompanying increase in study effort or test performance. This hypothesis is supported by data at available at follow-up: students in our sample who report academic goals to be 'important or extremely important' to them after one year, not only score 2.93 points higher on the grit scale (p = 0.000), they also score 4.1 percentage points higher in their board exams at 2 months (p=0.011) and spend 0.31 additional hours studying at home (p = 0.090).[15] While the treatment discussion increased perseverance or grit in our sample, it did not lead to an increased focus on academic goals: treated students are just as likely to report academic goals

[14] More recent work suggests that even when effort increases, it may not increase sufficiently to lead to an improvement in scores, at least in the short run (Islam et al., 2020). Note, Alan et al. (2019) and Outes-Leon et al. (2020) find performance on objective tests (e.g. maths) increase, but there are no effects on subjective exams (e.g., language). Therefore, student performance on different subjects may have been differently impacted by treatment but we do not have test scores disaggregated by subject and comment if there is heterogeneity in effects by test subject.

[15] Note, the correlations between the importance of goal with days per week college is attended and scores at 9 months are also positive but they are insignificant (p = 0.258 and 0.71, respectively).

are important to them as the placebo students (p = 0.998), which may explain why find an increase in grit but no improvement in study time or test scores due to the treatment.

The heterogeneity analysis provides further corroboration of this hypothesis. Recall that student enrolled in the first year were both more likely to report higher levels of grit and had reduced hours of study at home at the end of the first year. When we restrict our analysis to the subsample of first year students, treated students are 6.2 percentage points less likely to have ranked academic goals as important after one year. In line with this pattern, we observe that in questions asked immediately after the intervention was administered, treated first-year students are also less likely to modify their academic goals to SMART goals than more senior students.[16] Overall, these results suggest that with decreased focus on academic goals, in favor of other non-academic goals, treated students in their first year of study spent fewer hours studying at home and experienced a decrease in scores in their board exams.

Finally, heterogeneity analysis points towards other important potential mediators of treatment effects on non-cognitive skills that are in line with existing findings from other studies. For instance, educated mothers can foster aspirations and may facilitate the treatment in engendering a growth mindset (Roy and Bhattacharya, 2018). In addition, the intervention may have been particularly effective in improving grit and persistence in particular groups of disadvantaged students with lower levels of cognitive ability, competitiveness and importance for career goals. Altogether these represent interesting sub-groups of students for whom soft touch interventions about malleability of intelligence may be effective for bringing positive changes in non-cognitive skills.

5 Robustness tests

We test the robustness of our main results in three main ways. First, though our sample is balanced on characteristics measured at baseline and attrition is unrelated to treatment status, we test if results are robust when we control for individual characteristics. We do this by re-estimating main regressions with individual level controls selected using ‘post-double-selection’ with LASSO (Belloni et al., 2014a,b).[17] Results are robust to the inclusion of these controls. Tables E13 - E14 in the Online Appendix provide these results. We also estimate the upper and lower bounds of treatment effects using the Lee (2009) procedure to account for any differential attrition. Results are robust and provided in Online Appendix

[16] These results are available in the Online Appendix Table D12.

[17] Note, we run an ANCOVA estimation for all analysis, controlling for baseline values of outcome variables where available, as advised in literature (McKenzie, 2012). In this robustness check we additionally control for individual characteristics at baseline.

Tables E15 - E16.

Second, we test if the results are robust to the method of survey administration. At endline, 74 surveys were conducted over the phone due to college closure in March 2020 in response to the Covid-19 pandemic. We find no statistically significant differences in characteristics of respondents interviewed in the person and those interviewed via phone ($p - value$ = 0.86, shown in Online Appendix Table E17). As an additional check against difference in response by method of survey, we re-estimate the main analysis for outcomes based on survey measures using only data from the sub-sample of students who participated in the face-to-face survey at both baseline and follow-up. Results are shown in Online Appendix Table E18 and remain qualitatively similar for measures from the survey - the longer run treatment effect on respondent grit is positive, albeit smaller in size than the results summarized in Section 4.

Third, we test if potential spillover of information between the treated and placebo sample students may have lead to biased estimates of treatment effects. We adapt the methodology developed by (Banerjee et al., 2010) to match participants with their friends and if there was indirect impact of the treatment discussion on their friends in our sample. At baseline, we ask respondents to identify two friends with whom they spend most time and identify if these friends are also in the treated or placebo samples. In case of a positive spillover of the treatment, we expect friend's outcomes to be similar to that of the treated sample. In particular, friends of treated students who are in the treated sample themselves may experience reinforced effects of the treatment, potentially leading to over-estimating the effects shown in the main analysis. However, similar to (Banerjee et al., 2010), we find no indication of such spillover on non-cognitive skills, effort in terms of study at home or college, or on test scores at 2 and 9 months. Tables E19 - E21 in the Online Appendix provide these results.

6 Conclusion

Improving learning outcomes have long been an important concern for policy makers in many developing countries. Studies suggest that improving school infrastructure, health and nutrition interventions, such as providing free lunch or deworming supplements can improve learning and retention. However, the impact of these interventions are modest when compared to the high costs of implementation (McEwan, 2015). A related literature on human psychology suggests that non-cognitive skills, such as grit, determination and perseverance, can also be important predictor of academic performance. These behavioral interventions have the advantage of being potentially cost-effective, which is an attractive feature in resource constrained settings.

We use a randomized control trial testing if a soft-touch, low-cost behavioral intervention can foster non-cognitive skills, effort and academic performance in a developing country. The intervention involves a brief discussion highlighting the role of effort and persistence in enhancing skills and performance that can be incorporated into school or college curriculum at low marginal cost. We find that the intervention succeeds in increasing grit amongst female college students in our sample, but it is not accompanied by an increase in student effort and academic performance. Our results support findings from other settings where behavioral interventions have been able to change non-cognitive skills but not resulted in any impact on behavior (Islam et al., 2020; Ganimian, 2020). To our knowledge, our study is the first to test an easily-scaleable intervention for adult, college-going students, whose non-cognitive skills are considered to be largely fixed. We show that it is possible to change non-cognitive skills with the help of soft-touch behavioral interventions. We exploit a rich data to investigate potential moderators of these effects, and find the impact can vary by year of enrollment, parental education, inherent levels of cognitive ability and competitiveness.

The study sample is relatively small, and this could mean that we are underpowered to detect certain effects. Nevertheless, the exploratory nature of our results call for a more direct identification of factors that not only affect how students think but can potentially change how they plan and act to achieve academic targets. Due to logistical and budgetary constraints, our sample consists only of women, who face substantially higher socio-cultural barriers to education than men in the study setting. An interesting question to test in future research would be if the effects of a similar intervention would differ for male students.

Our findings are important for policy focusing on education and improving student persistence in low-income countries, where a large fraction of students in different levels of education face a learning crisis. While the role of mindset in determining academic performance is rarely studied in economics, non-cognitive skills such as grit and persistence have been hypothesized to be important determinants of welfare in psychological theory. This study adds to a growing strand of literature trying to connect the two complimentary fields of study.

References

ALAN, S., T. BONEVA, AND S. ERTAC (2019): "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit," The Quarterly Journal of Economics, 134, 1121–1162.

ALAN, S. AND S. ERTAC (2015): "Patience, self-control and the demand for commitment: Evidence from a large-scale field experiment," Journal of Economic Behaviour and Organization, 115, 111–122.

ANDERSON, M. (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.

ANGRIST, N., D. K. EVANS, D. FILMER, R. GLENNERSTER, F. H. ROGERS, AND S. SABARWAL (2020): "How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric," Policy Research Working Paper No. 9450.

BANDURA, A. (1978): "Self-efficacy: Toward a unifying theory of behavioral change," Advances in Behaviour Research and Therapy, 1, 139–161.

BANERJEE, A. V., R. BANERJI, E. DUFLO, R. GLENNERSTER, AND S. KHEMANI (2010): "Pitfalls of participatory programs: Evidece from a randomized evaluation in education in India," American Economics Journal: Economic Policy, 2, 1–30.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2013): "Inference on Treatment Effects after Selection among High-Dimensional Controls," The Review of Economic Studies, 81, 608–650.

——— (2014a): "High-Dimensional Methods and Inference on Structural and Treatment Effects," The Journal of Economic Perspectives, 28, 29–50.

——— (2014b): "Inference on Treatment Effects After Selection Among High-dimensional Controls," The Review of Economic Studies, 81, 608–650.

BETTINGER, E., S. LUDVIGSEN, M. REGE, I. F. SOLLI, AND D. YEAGER (2018): "Increasing perseverance in math: Evidence from a field experiment in Norway," Journal of Economic Behavior Organization, 146, 1–15.

BLACKWELL, L. S., K. H. TRZESNIEWSKI, AND C. S. DWECK (2007): "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention," Child Development, 78, 246–263.

BORGHANS, L., A. L. DUCKWORTH, J. HECKMAN, AND B. TER WEEL (2008): "The Economics and Psychology of Personality Traits," Journal of Human Resources, 43.

BURNETTE, J. L., E. H. O'BOYLE, E. M. VANEPPS, J. M. POLLACK, AND E. J. FINKEL (2013): "Mind-sets matter: a meta-analytic review of implicit theories and self-regulation," Psychological bulletin, 139(3), 655–701.

CENSUS (2017): Population Census of Pakistan, Pakistan Bureau of Statistics.

COSTRELL, R. M. (1994): "A simple model of educational standards," The American Economic Review, 956–971.

CREDÉ, M., M. C. TYNAN, AND P. D. HARMS (2017): "Much ado about grit: A meta-analytic synthesis of the grit literature," Journal of Personality and Social Psychology, 113, 492–511.

DAHL, M. (2016): "Don't believe the hype about grit, pleads the scientist behind the concept," New York Magazine, Accessed 2020-02-20.

DE, FRAJA. G., O. T. . Z. L. (2010): "Must try harder: Evaluating the role of effort in educational attainment," The Review of Economics and Statistics, 92(3), 577–597.

DILLON, M. R., H. KANNAN, J. T. DEAN, E. S. SPELKE, AND E. DUFLO (2017): "Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics," Science, 375, 47–55.

DOBRONYI, C. R., O. P. AND U. PETRONIJEVIC (2019): "Goal setting, academic reminders, and college success: A large-scale field experiment," Journal of Research on Educational Effectiveness, 12, 38–66.

DUCKWORTH, ANGELA L., . Q. P. D. (2009): "Development and validation of the Short Grit Scale (GRIT-S)," Journal of personality assessment, 91, 166–174.

DUCKWORTH, A. L., C. PETERSON, M. D. MATTHEWS, AND D. R. KELLY (2007): "Grit: Perseverance and passion for long-term goals," Journal of Personality and Social Psychology, 92, 1087–1101.

DUCKWORTH, A. L., P. C. M. M. D. . K. D. R. (2007): "Grit: Perseverance and passion for long-term goals." Journal of Personality and Social Psychology, 92, 1087–1101.

DURLAK, J. A., R. P. WEISSBERG, A. B. DYMNICKI, R. D. TAYLOR, AND K. B. SCHELLINGER (2011): "The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions," Child Dvelopment, 82, 405–432.

DWECK, C. (2006): Mindset: The new psycology of sucess, New York, NY: Random House 2006.

FISZBEIN, A. AND N. SCHADY (2009): "Conditional Cash Transfers: Reducing Present and Future Poverty," Unpublished manuscript.

GANIMIAN, A. J. (2020): "Growth mindset interventions at scale: Experimental evidence from Argentina," Educational Evaluation and Policy Analysis, 42, 417–438.

GLEWWE, P. AND E. MIGUEL (2008): " The Impact of Child Health and Nutrition on Education in Less Developed Countries," Handbook of Development Economics, 4, 3561–3606.

HECKMAN, J. J. AND Y. RUBINSTEIN (2001): "The importance of noncognitive skills: Lessons from the GED testing program," American Economic Review, 91, 145–149.

HUMPHRIES, J. E. AND K. FABIAN (2017): "On the intrepretation of non-cognitive skills - What is being measures and why it matters," Journal of Economic Behaviour and Organization, 136, 174–185.

ISLAM, A., S. KWON, E. MASOOD, N. PRAKASH, S. SABARWAL, AND D. SARASWAT (2020): "When Goal-Setting Forges Ahead but Stops Short," IZA Discussion Papers.

JACHIMOWICZ, J. M., A. WIHLER, E. R. BAILEY, AND A. D. GALINSKY (2018): "Why grit requires perseverance and passion to positively predict performance," Proceedings of the National Academy of Sciences, 115, 9980–9985.

KAFFENBERGER, M., P. L. (2021): "A structured model of the dynamics of student learning in developing countries, with applications to policy," International Journal of Educational Development, 82, 102371.

KAUTZ, T., J. J. HECKMAN, R. DIRIS, B. TER WEEL, AND L. BORGHANS (2014): "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success," National Bureau of Economic Research, w20749.

LEE, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," The Review of Economic Studies, 76, 1071–1102.

LINDQVIST, E. AND R. VESTMAN (2011): "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish Enlistment," American Economic Journal: Applied Economics, 3, 101–128.

MCEWAN, P. (2015): "Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments," Review of Educational Research, 85, 353–394.

McKenzie, D. (2012): "Beyond baseline and follow-up: The case for more T in experiments," Journal of Development Economics, 99, 210–221.

Morisano, D., J. B. Hirsh, J. B. Peterson, R. O. Pihl, and B. M. Shore (2010): " Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance," Journal of Applied Psychology, 95, 255–264.

Outes-Leon, I., A. Sanchez, and R. Vakis (2020): "The Power of Believing You Can Get Smarter: The Impact of a Growth-Mindset Intervention on Academic Achievement in Peru," Policy Research Working Paper 9141, World Bank, Washington.

Pakistan Bureau of Statistics (2019): "Pakistan Social and Living Standards Measurement Survey PSLM/HIES 2018-19 Report," .

Paunesku, D., G. M. Walton, C. Romero, E. N. Smith, D. S. Yeager, and C. S. Dweck (2015): "Mind-set interventions are a scalable treatment for academic underachievement," Psychological science, 26, 784–793.

PDS (2017): "Punjab Development Statistics Report," .

——— (2018): Punjab Development Statistics 2018, Bureau of Statistics Punjab.

PES (2016): Pakistan Education Statistics 2016, Academy of Educational Planning and Management.

Polley, T. (2018): "Believing is Achieving: Targeting Beliefs about Intelligence to Increase Learning in Bangladeshi Secondary Schools," Unpublished manuscript.

Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg (2007): "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes," Perspectives on Psychological Science, 2, 313–345.

Roy, Sanchari, M. M. and S. Bhattacharya (2018): "Hidden human capital: Self-efficacy, aspirations and achievements of adolescent and young women in India," World Development, 111, 161–180.

Ryckman, R. M., M. Hammer, L. M. Kaczor, and J. A. Gold (1996): "Construction of a Personal Development Competitive Attitude Scale," Journal of Personality Assessment, 66, 374–385.

Schwarzer, R. and M. Jerusalem (1995): "Generalized Self-Efficacy scale," in Measures in health psychology: A user's portfolio. Causal and control beliefs, ed. by J. Weinman, S. Wright, and M. Johnston, Windsor, England: NFER-Nelson, 35–37.

SRIRAM, R. (2014): "Rethinking intelligence: The role of mindset in promoting success for academically high-risk students," Journal of College Student Rentention, 15, 515–536.

STINEBRICKNER, R., . S. T. R. (2008): "The causal effect of studying on academic performance," The BE Journal of Economic Analysis & Policy, 8.

THE UNIVERSITY OF PUNJAB (2019): "Merit Lists," Data retrieved from http://pu.edu.pk/page/downloads.

THE WORLD BANK (2018): The many faces of learning crisis, World Development Report (Washington, DC).

——— (2020): "Enrollment, tertiary," Data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.TER.ENRR.

UNESCO (2009): "EFA Global Monitoring Report: Overcoming Inequality: Why Governance Matters," Data retrieved from www.unesco.org/new/en/education/themes/leading-the-international-agenda/efareport/reports/2009-governance.

YEAGER, D. S., P. HANSELMAN, D. PAUNESKU, C. HULLEMAN, C. DWECK, C. MULLER, AND A. DUCKWORTH (2018): "Where and for whom can a brief, scalable mindset intervention improve adolescents' educational trajectories," PsyArXiv.

YEAGER, D. S., R. JOHNSON, B. J. SPITZER, K. H. TRZESNIEWSKI, J. POWERS, AND C. S. DWECK (2014): "The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health and achievement during adolescence," Journal of Personality and Social Psychology, 106, 867–884.

YEAGER, D. S., C. ROMERO, D. PAUNESKU, C. S. HULLEMAN, B. SCHNEIDER, C. HINOJOSA, H. Y. LEE, AND J. E. A. O'BRIEN (2016): "Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school," Journal of Educational Psychology, 108, 374–391.

Appendix

A1 Details of the intervention discussions

The treatment and placebo discussions were based on the materials designed by Blackwell et al. (2007), which we translate into the local language (*Urdu)* and adapt for the local context.[18] We conducted extensive piloting of the material to ensure the conversation is sufficiently brief and of comparable length to discussions in other studies, to effectively communicate the main messages without losing student's interest and attention. Our discussions lasted 30 minutes, on average. All discussions were aided by illustrative slides and images to keep the student engaged. The treatment and placebo discussions took place face-to-face, between a trained enumerator and the student.

A1.1 Treatment discussion

The discussion is based on the concept that it is possible to grow your intelligence. For instance, we discuss:

- The concept that brain is a muscle that changes and grows stronger the more we use it
- The role of nerve cells that have branches connecting them to other cells in a network that grows when we learn new things
- The scientific findings that showed measurable physical change in the brain due to increased mental activity
- The importance of practice to grow 'smarter' and help the brain grow
- The idea that failure is not a setback but an opportunity to learn, and that sustained effort can eventually make you succeed

The discussion ended with the enumerator leading the student through guided examples where the student may have encountered a setback, e.g. when they received a lower grade than they expected, and how they can use a growth mindset to achieve their goals. As an illustration of how to apply the key concepts just discussed, the enumerator and student went over the key things to remember for achieving their goals - for instance, asking questions from their peers, teachers and adults to learn; analyzing mistakes and situations where they had done well; maintaining consistent effort; and having a specific plan or objective for study sessions, etc.

[18] An excellent resource of free and paid resources based on this seminal work is available at: https://www.mindsetworks.com/

A1.2 Placebo discussion

The placebo discussion was also based on a similar discussion in Blackwell et al. (2007). The discussion provides an overview of the structure of the brain and its functions. For instance, we discuss:

- The size of the brain and it's proportion of human body weight
- The estimated number of nerves in the brain that communicate, in a complicated network, and the way information passes via a series of electrical signals
- The three parts of the brain: the cerebrum, cerebellum and the brainstem, and a discussion of their functions.
- An overview of the frontal, parietal, temporal and occipital lobes, and their responsibilities
- An overview of the two broad group of cells in the brain.

Each sub-heading were discussed in detail, providing information on the composition and responsibilities of relevant component, and ending with a note on how their roles are inter-linked for the brain to conduct basic functions.

A2 Additional Tables

Table A1: Response to treatment and placebo discussions

Information provided …	Means		Treated(mean) - Placebo(mean)
	Placebo	Treated	
	(1)	(2)	(3)
… is interesting	0.922	0.995	0.073***
	(0.269)	(0.073)	(0.020)
… left an impact on me	0.838	0.995	0.157***
	(0.369)	(0.073)	(0.028)
… is useful for the way I think	0.765	0.979	0.213***
	(0.425)	(0.145)	(0.033)
… prompts me to think about goals and next steps	0.665	0.973	0.308***
	(0.473)	(0.162)	(0.037)
… is something I may be interested in receiving in future	0.872	0.995	0.123***
	(0.336)	(0.073)	(0.025)
N	179	187	366

Note: the table summarizes response to binary questions about the information provided to the students, coded as 1 if the respondent said 'Yes' and 0 if 'No'. Columns (1) and (2) present mean values for the placebo and treatment group for each question presented in the row, respectively. Column (3) provides the difference in means between the treatment and placebo mean values. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, ***.

Table A2: Full sample description and balance at baseline

	Sample mean			*p-value*
	Full	Placebo (P)	Treatment (T)	*P vs T*
	(1)	(2)	(3)	(4)
Household characteristics:				
Father's education (years)	8.43	8.17	8.69	0.31
Father is a business owner	0.25	0.25	0.26	0.91
Father is a salaried worker	0.57	0.56	0.57	0.79
Mother's education (years)	6.64	6.49	6.79	0.58
Mother is a business owner	0.02	0.01	0.02	0.44
Mother is a salaried worker	0.05	0.05	0.07	0.41
Avg. monthly household income (USD)	254	267	242	0.40
Household assets (index)	0.00	-0.04	0.04	0.57
Individual characteristics:				
Age (years)	19.37	19.54	19.23	0.07*
Marital status: Single	0.96	0.96	0.96	0.73
Enrolled in first year	0.36	0.31	0.40	0.10*
Plans to continue education after graduating	0.91	0.92	0.91	0.81
Wants to pursue paid job	0.92	0.96	0.89	0.03**
Passed last degree in first division	0.72	0.75	0.70	0.29
Discusses homework with parents	0.71	0.73	0.690	0.45
Discusses goals with parents	0.93	0.94	0.90	0.11
Risk aversion (out of 6)	4.15	4.05	4.24	0.25
Raven's test score (out of 10)	4.08	4.07	4.09	0.93
Competitiveness scale (out of 75)	57.62	57.86	57.40	0.61
Satisfaction with academic performance (out of 5)	2.37	2.36	2.37	0.95
P-value of F-statistic of joint significance				0.11
Observations	366	179	187	

Note: Column (1) shows the mean values for the specified variable for the overall sample, while column (2) shows the control group mean values, and column (3) shows treatment group mean values. Column (4) shows p-values from a balance test obtained by regressing the variable specified in each row on treatment status. Binary variables are 1 when the response is 'Yes'; 0 otherwise. Risk aversion can range from 1-6, higher values represent higher levels of aversion. Raven's score can range from 0 - 10, with higher scores representing higher cognitive ability. Competitiveness scale (Ryckman et al., 1996) provides a score from 0 - 75 and is increasing in level of competitivness. Satisfaction with academic performance is measured using a likert scale for 1 for strongly disagreeing to 5 for strongly agreeing. There are 15, 4, 3 and 108 missing values for father's education, mother's education, division scored in last degree, and monthly household income, respectively. F-statistic explains the overall significance of the model- the null hypothesis is that all coefficients on the independent variables are equal to zero. $***p < 0.01, **\, p < 0.05, *p < 0.1$.

Table A3: Summary statistics of outcomes

Outcomes	N (1)	Mean (2)	Median (3)	S.Dev. (4)
Panel (a): Measures at baseline				
Grit (out of 60)	366	40.86	40.00	6.28
Growth mindset (out of 90)	366	50.49	51.00	6.64
Self- efficacy (out of 40)	366	32.93	33.50	5.24
Mindset index	366	0.02	0.03	1.02
Hours of study at home per day	366	3.39	3.00	1.80
Days college attended past week	366	4.67	5.00	1.80
Panel (b): Measures of effort at followup				
Grit (out of 60)	300	41.13	40.00	6.18
Growth mindset (out of 90)	300	47.88	48.00	7.13
Self-efficacy (out of 40)	300	32.94	34.00	5.11
Psychological index	300	-0.22	-0.16	0.98
Hours of study at home per day	300	3.40	3.00	1.62
Days college attended past week	300	4.91	6.00	1.56
Test Scores (after 2 months)	327	0.48	0.50	0.15
Test Scores (after 9 months)	332	0.40	0.42	0.23

Note: Column (1) shows the number of observations; column (2) shows the value of outcome; column (3) shows the median value and column (4) shows standard deviation. Hours of study and days present are measured during the baseline survey, and during the final followup (endline) survey. Grit, Growth, self-efficacy, psychological index are measured using validated scales in the the same first interaction with the students, immediately after the baseline survey and intervention is administered, and once again in the final followup survey. Test scores are only collected at two points in time, 2 and 9 months after the baseline and intervention is administered.

Table A4: Balance test for Attrition

	Mean (1)	Median (2)	S.Dev. (3)	$p-value$ (4)
Dummy: Treatment	0.51	1.00	0.51	0.20
Household characteristics:				
Father's education (years)	8.43	10.00	4.80	0.24
Father is a business owner	0.25	0.00	0.44	0.31
Father is a salaried worker	0.57	1.00	0.50	0.14
Mother's education (years)	6.64	8.00	5.15	0.73
Mother is a business owner	0.02	0.00	0.13	0.25
Mother is a salaried worker	0.05	0.00	0.23	0.72
Avg. monthly household income (USD)	254	179	236	0.50
Household assets (index)	0.00	0.30	1.39	0.48
Individual characteristics:				
Age (years)	19.37	19.00	1.62	0.13
Marital status: Single	0.96	1.00	0.20	0.24
Enrolled in first year	0.36	0.00	0.48	0.33
Plans to continue education after graduating	0.91	1.00	0.28	0.91
Wants to pursue paid job	0.92	1.00	0.27	0.32
Passed last degree in first division	0.72	1.00	0.45	0.11
Discusses homework with parents	0.71	1.00	0.46	0.335
Discusses goals with parents	0.93	1.00	0.27	0.26
Risk aversion (out of 6)	4.15	4.00	1.59	0.88
Raven's test score (out of 10)	4.08	4.00	2.11	0.15
Competitiveness scale (out of 75)	57.62	58.00	8.33	0.07*
Satisfaction with academic performance (out of 5)	2.37	2.00	0.91	0.09*
P-value of F-statistic of joint significance				0.11

Note: Column (4) shows p-values from the balance test on individual and household characteristics on attrition. The cells show the coefficient on treatment assignment when the variable in the row is regressed on the treatment assignment. The total number of observations in each regression are equal to 366 except for father's education (351), mother's education (362), monthly household income (258) and passed last degree in the first division (363). F-statistic explains the overall significance of the model- the null hypothesis is that all coefficients on the independent variables are equal to zero. If the null is rejected, there is a well balanced sample. ***p < 0.01, **p < 0.05,*p < 0.1.

Table A5: Long term impact of treatment on (standardized) non-cognitive skills

	Grit	Grit	Growth mindset	Growth mindset	Self efficacy	Self efficacy	Mindset index	Mindset index
	SD	SD	SD	SD	SD	SD	SD	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-0.159	-0.172	0.170	0.141	0.011	0.032	0.029	0.011
	(0.101)	(0.101)*	(0.099)*	(0.097)	(0.107)	(0.105)	(0.107)	(0.109)
Post	-0.089	-0.104	-0.393	-0.400	0.015	0.020	-0.323	-0.333
	(0.082)	(0.081)	(0.102)***	(0.101)***	(0.093)	(0.094)	(0.090)***	(0.092)***
Treated*Post	0.255	0.267	0.035	0.043	-0.027	-0.043	0.165	0.170
	(0.115)**	(0.115)**	(0.140)	(0.140)	(0.135)	(0.135)	(0.133)	(0.134)
N	666	666	666	666	666	666	666	666
R^2	0.005	0.034	0.043	0.092	0.000	0.048	0.018	0.050
FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. All outcomes are standardized. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. Treated*Post is the average treatment effect of our treatment intervention. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Impact of treatment on (standardized) measures of effort

	Hours study SD	Hours study SD	Days attend SD	Days attend SD
	(1)	(2)	(3)	(4)
Treated	0.045	-0.004	-0.016	-0.046
	(0.110)	(0.102)	(0.104)	(0.100)
Post	0.023	-0.001	0.099	0.102
	(0.087)	(0.086)	(0.104)	(0.105)
Treated*Post	-0.031	-0.004	0.066	0.061
	(0.122)	(0.121)	(0.139)	(0.139)
N	666	666	666	666
R^2	0.000	0.105	0.005	0.064
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. All outcomes are standardized. 'Hours study' are the number of hours students report studying at home. 'Days attend' are number of days the student attended college in the most recent full week. 'Treated*Post' is the average treatment effect of our treatment intervention. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Impact of treatment on (standardized) exam performance

Time after baseline	2 months		9 months	
	Score SD	Score SD	Score SD	Score SD
	(1)	(2)	(3)	(4)
Treated	-0.044	-0.105	-0.013	0.003
	(0.109)	(0.102)	(0.114)	(0.104)
N	327	327	332	332
R^2	0.000	0.136	0.000	0.188
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. 'Score SD' are the standardized proportion scored in end-of-year and after nine months of our intervention examination out of the maximum (full) marks allowed for that enrollment year. These were measured 2 and 9 months after the baseline activities and are present after our intervention. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: For Online Publication

OA1 Cost of the intervention

We employ the intervention designed by psychology literature (Blackwell et al., 2007) and adapt it to the local context. The translated instruments were piloted extensively in a college not belonging to our sample, and include costs of translation and salary paid to enumerators. Note, that we do not include cost of researcher time input into the development of the study design. The development of these instruments are estimated at USD 89.29.[19]

The final intervention was implemented by a team of enumerators, who also collected baseline data in the same visit. We estimate that half of the time and resources of the field team at baseline were spent on intervention implementation, for a cost of USD 1,405. Included in field team costs are the costs of training, and salaries of enumerators and field supervisors. We assume that the total time spent with treated and placebo participants are not meaningfully different and divide the total costs of implementation by the total number of participants contacted for implementation (N = 366). We estimate the per participant costs amount to USD 4.08. The per unit cost is expected to fall with larger sample. As such, we assume these estimates to provide an upper limit of the costs that can be incurred with a larger group of students.

[19] We use 140 PKR = 1 USD exchange rate as of March 31 2019.

OA2 Scales used to measure non-cognitive skills

Grit is measured using a validated 12-item scale, each item is based on a likert scale ranging from 1 (Not like me at all) to 5 (very much like me), developed by Duckworth et al. (2007).

- I have overcome setbacks to conquer an important challenge.
- New ideas and projects sometimes distract me from previous ones.
- My interests change from year to year.
- Setbacks don't discourage me.
- I have been obsessed with a certain idea or project for a short time but later lost interest.
- I am a hard worker.
- I often set a goal but later choose to pursue a different one.
- I have difficulty maintaining my focus on projects that take more than a few months to complete.
- I finish whatever I begin.
- I have achieved a goal that took years of work.
- I become interested in new pursuits every few months.
- I am diligent.

Growth mindset is constructed from the sum of 15 items, each scored on a scale from 1 (agree a lot) to 6 (disagree a lot) point Likert scale. There are six components of theory of intelligence (includes three entity theory and three increment theory items); 9 components of effort beliefs and goals, developed by Blackwell et al. (2007).

- Your intelligence is something you can't change very much.
- You have a certain amount of intelligence, and you really can't do much to change it.
- You can learn new things, but you can't really change your basic intelligence.
- No matter who you are, you can change your intelligence a lot.
- You can always greatly change how intelligent you are.
- No matter how much intelligence you have, you can always change it a good amount.

- To tell the truth, when I work hard at my college work, it makes me feel I am not very smart.
- It doesn't matter how hard you work— if you're not smart, you won't do well.
- When something is hard, it just makes me want to work more on it, not less.
- If you're not good at a subject, working hard won't make you good at it.
- If a subject is hard for me, it means I probably won't be able to do really well at it.
- If you're not doing well at something, it's better to try something easier.
- If you don't work hard and put in a lot of effort, you probably won't do well.
- The harder you work at something, the better you will be at it.
- If an assignment is hard, it means I'll probably learn a lot doing it.

Self- efficacy is 10-item scale developed by Schwarzer and Jerusalem (1995). Each item is scored on a 1 (Not at all true) to 4 (exactly true) point Likert scale:

- I can always manage to solve difficult problems if I try hard enough.
- If someone opposes me, I can find the means and ways to get what I want.
- It is easy for me to stick to my aims and accomplish my goals.
- I am confident that I could deal efficiently with unexpected events.
- Thanks to my resourcefulness, I know how to handle unforeseen situations.
- I can solve most problems if I invest the necessary effort.
- I can remain calm when facing difficulties because I can rely on my coping abilities.
- When I am confronted with a problem, I can usually find several solutions.
- If I am in trouble, I can usually think of a solution.
- I can usually handle whatever comes my way.

OA3 The Covid-19 lockdown for educational institutions in Punjab, Pakistan

On 7 March, 2020, sample colleges closed for spring break. Before they could re-open, on 14 March, the Higher Education Commission implemented a complete closure of educational institutes across the country till April, 5 in response to the global Covid-19 pandemic. All institutions will remain closed till 31 May, 2020. On 23 March, the Government of Punjab announced a complete lockdown due to the covid-19 pandemic. Finally, on 26 March, the federal government announced all educational institutes will remain closed till 31 May. At the time of registration of this document, the lock-down will be in place till 30 April, 2020.[20] Before the lockdown, we had completed 61% of the surveys face-to-face. The in-person interaction involved repeat interventions with the treated and placebo participants.

[20] All schools and colleges were closed down under the Government of Punjab's Notification No. SO(A-II)1-1/2001(P). The date for reopening schools and colleges was extended to May 31 by order of the Government of Pakistan, Notification No.F.1-1/2020-FEPT. The general lockdown in Punjab was implemented by a special order (NO(IS-III)1-1/2004) under section 144 of the national Code of Criminal Procedure 1898.

OA4 Heterogenity tables

Table D1: Heterogeneity in mindset scales by baseline characteristics (saturated regression)

Dependent variable:	(1) Grit	(2) Growth mindset	(3) Self efficacy	(4) Mindset Index
T*Post	3.425	0.309	0.671	0.414
	(1.598)**	(1.991)	(1.670)	(0.302)
Household assets*Treated*Post	0.509	1.185	-0.947	0.077
	(1.072)	(1.401)	(1.085)	(0.216)
Mother education*Treated*Post	-1.119	4.815	1.840	0.553
	(1.178)	(1.387)***	(1.130)	(0.224)**
Career goal important*Treated*Post	-1.817	1.003	-0.722	-0.137
	(1.124)	(1.407)	(1.046)	(0.213)
Raven's score*Treated*Post	-1.632	-0.178	1.316	-0.045
	(1.128)	(1.237)	(0.982)	(0.184)
Competitiveness*Treated*Post	-1.709	-0.288	-0.827	-0.266
	(1.024)*	(1.320)	(0.967)	(0.186)
First division*Treated*Post	0.263	-2.903	-1.530	-0.411
	(1.107)	(1.670)*	(1.031)	(0.224)*
First year*Treated*Post	1.684	-1.661	0.947	0.080
	(1.265)	(1.644)	(1.194)	(0.229)
Discuss HW*Treated*Post	0.129	-0.889	-0.398	-0.115
	(1.188)	(1.469)	(1.212)	(0.227)
Mean	41.125	48.660	32.938	-0.143
N	662	662	662	662
R^2	0.129	0.105	0.081	0.104

Note: Standard errors in parentheses. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. Treated*Post is the average treatment effect of our treatment intervention. For each dimension of heterogeneity, 'H', we show here the coefficient of interaction of H*Treated*Post. All regressions control for base levels of H and H*Treated. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D2: Heterogeneity in effort measures by baseline characteristics (saturated regression)

	(1) Hours study at home	(2) Days college attended
Treated*Post	-0.061	-0.035
	(0.390)	(0.443)
Household assets*Treated*Post	0.178	0.062
	(0.294)	(0.310)
Mother education*Treated*Post	-0.047	0.186
	(0.316)	(0.333)
Career goal important*Treated*Post	0.002	0.376
	(0.291)	(0.322)
Raven's score*Treated*Post	-0.269	-0.318
	(0.306)	(0.333)
Competitiveness*Treated*Post	-0.169	-0.327
	(0.297)	(0.327)
First division*Treated*Post	0.606	-0.041
	(0.355)*	(0.357)
First year*Treated*Post	-0.840	-0.777
	(0.335)**	(0.347)**
Discuss HW*Treated*Post	0.082	0.605
	(0.293)	(0.358)*
Mean	3.366	4.766
N	662	662
R^2	0.107	0.046

Note: Standard errors in parentheses. 'Hours study at home' is the hours per day spent studying, and 'Days college attended' are the days (out of 6) that the student is present in college. Treated*Post is the average treatment effect of our treatment intervention. For each dimension of heterogeneity, 'H', we show here the coefficient of interaction of H*Treated*Post. All regressions control for base levels of H and H*Treated. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D3: Heterogeneity in test scores by baseline characteristics (saturated regression)

	(1)	(2)
Dependent variable:	Test score % 2month	Test score % 9 month
Treated	0.065	0.064
	(0.044)	(0.072)
Household assets*Treated	-0.046	-0.017
	(0.032)	(0.054)
Mother education*Treated	-0.047	-0.033
	(0.033)	(0.057)
Career goal important*Treated	-0.003	0.045
	(0.034)	(0.055)
Raven's score*Treated	-0.000	0.004
	(0.032)	(0.052)
Competitiveness*Treated	0.041	0.025
	(0.031)	(0.051)
First division*Treated	-0.047	-0.082
	(0.037)	(0.057)
First year*Treated	-0.034	0.004
	(0.035)	(0.060)
Discuss HW*Treated	0.002	-0.021
	(0.034)	(0.058)
Mean	0.488	0.403
N	326	329
R^2	0.085	0.025

Note: Standard errors in parentheses. Test score percentage' are the marks scored in end-of-year examination as a percentage of the maximum (full) marks allowed for that enrollment year. These were measured 2 and 9 months, after the baseline activities. Treated is the average treatment effect of our treatment intervention. For each dimension of heterogeneity, 'H', we show here the coefficient of interaction of H and Treated. All regressions control for base levels of 'H'. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D4: Heterogeneity in treatment impact on Grit

Panel A: Dependent variable: Grit

H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	-0.496	-0.674	-0.386	-0.945
	(0.909)	(0.834)	(0.791)	(0.857)
Post	-0.600	-0.611	-0.632	-0.631
	(0.530)	(0.529)	(0.521)	(0.525)
H	-0.662	2.020	3.236	1.082
	(0.823)	(0.811)**	(0.777)***	(0.810)
H*Treated	-0.937	-0.755	-1.248	-0.125
	(1.207)	(1.193)	(1.170)	(1.199)
Treated*Post	1.227	1.783	2.668	2.520
	(0.993)	(0.930)*	(0.899)***	(0.884)***
H*treated*Post	0.830	-0.195	-1.983	-2.032
	(1.063)	(1.047)	(1.045)*	(1.038)*
Mean	41.125	41.125	41.125	41.125
N	666	666	666	666
R^2	0.012	0.023	0.045	0.012

Panel B: Dependent variable: Grit

H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	-0.643	0.371	-0.457	-1.124
	(0.729)	(1.115)	(0.752)	(1.146)
Post	-0.622	-0.658	-0.604	-0.571
	(0.517)	(0.528)	(0.527)	(0.527)
H	3.191	1.795	2.412	0.281
	(0.789)***	(0.850)**	(0.926)***	(0.926)
H*Treated	-0.624	-1.843	-1.945	0.155
	(1.185)	(1.311)	(1.301)	(1.326)
Treated*Post	2.650	1.485	0.808	1.478
	(0.902)***	(0.992)	(0.838)*	(1.135)
H*Treated*Post	-2.119	0.284	2.129	0.237
	(1.046)**	(1.073)	(1.097)*	(1.180)
Mean	41.125	41.125	41.125	41.125
N	666	662	666	666
R^2	0.049	0.012	0.031	0.006

Note: Standard errors in parentheses. All errors are clustered at the individual level. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale (Duckworth et al., 2007). The H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D5: Heterogeneity in treatment impact on Growth Mindset

Dependent variable: Growth mindset				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	0.697	2.317	1.050	0.927
	(0.915)	(0.916)**	(0.930)	(0.890)
Post	-2.764	-2.764	-2.742	-2.645
	(0.718)***	(0.715)***	(0.714)***	(0.712)***
H	-0.410	0.844	-0.503	-2.103
	(0.900)	(0.865)	(0.869)	(0.853)**
H*Treated	0.993	-2.392	0.282	0.501
	(1.281)	(1.250)*	(1.258)	(1.245)
Treated*Post	-0.376	-1.539	0.480	-0.023
	(1.207)	(1.140)	(1.157)	(1.155)
H*treated*Post	1.251	3.862	-0.521	0.375
	(1.334)	(1.310)***	(1.335)	(1.319)
Mean	48.660	48.660	48.660	48.660
N	666	666	666	666
R^2	0.048	0.055	0.045	0.059

Dependent variable: Growth mindset				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	-0.151	0.189	0.882	0.710
	(0.921)	(1.247)	(0.832)	(1.135)
Post	-2.716	-2.721	-2.756	-2.759
	(0.709)***	(0.716)***	(0.713)***	(0.712)***
H	-2.460	0.321	0.504	-0.379
	(0.850)***	(0.997)	(0.924)	(0.904)
H*Treated	2.785	1.614	0.678	0.680
	(1.246)**	(1.429)	(1.300)	(1.343)
Treated*Post	0.211	1.808	0.718	0.386
	(1.121)	(1.410)	(1.063)	(1.392)
H*Treated*Post	-0.012	-2.344	-1.183	-0.194
	(1.343)	(1.453)	(1.407)	(1.440)
Mean	48.660	48.660	48.660	48.660
N	666	662	666	666
R^2	0.058	0.049	0.046	0.044

Note: Standard errors in parentheses. All errors are clustered at the individual level. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale (Blackwell et al., 2007). H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D6: Heterogeneity in treatment impact in Self Efficacy

Independent variable: Self Efficacy				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	-0.015	0.387	-0.101	0.548
	(0.741)	(0.716)	(0.790)	(0.720)
Post	0.086	0.066	0.057	0.046
	(0.465)	(0.473)	(0.471)	(0.474)
H	0.368	0.493	1.010	0.564
	(0.666)	(0.673)	(0.665)	(0.663)
H*Treated	0.097	-0.701	0.348	-1.166
	(1.033)	(1.034)	(1.024)	(1.037)
Treated*Post	-0.039	-0.938	0.509	-0.537
	(0.908)	(0.862)	(0.897)	(0.808)
H*treated*Post	-0.206	1.746	-1.291	1.061
	(1.000)	(0.978)*	(0.981)	(0.996)
Mean	32.938	32.938	32.938	32.938
N	666	666	666	666
R^2	0.001	0.006	0.009	0.003

Independent variable: Self Efficacy				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss work (8)
Treated	0.037	0.190	0.242	-1.367
	(0.703)	(0.920)	(0.611)	(1.079)
Post	0.041	0.075	0.074	0.053
	(0.467)	(0.475)	(0.473)	(0.473)
H	2.341	-0.336	-0.010	-0.944
	(0.647)***	(0.713)	(0.761)	(0.725)
H*Treated	0.208	-0.184	-0.465	2.015
	(1.004)	(1.111)	(1.137)	(1.199)*
Treated*Post	0.250	0.526	-0.384	-0.196
	(0.860)	(0.983)	(0.770)	(1.191)
H*treated*Post	-0.833	-1.001	0.619	0.114
	(0.980)	(1.052)	(1.038)	(1.213)
Mean	32.938	32.938	32.938	32.938
N	666	662	666	666
R^2	0.048	0.005	0.001	0.009

Note: Standard errors in parentheses. All errors are clustered at the individual level. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale (Schwarzer and Jerusalem, 1995). H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D7: Heterogeneity in treatment impact in Mindset Index

Independent variable: Mindset Index

H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	0.022	0.206	0.060	0.057
	(0.147)	(0.136)	(0.145)	(0.137)
Post	-0.325	-0.328	-0.329	-0.320
	(0.090)***	(0.090)***	(0.090)***	(0.090)***
H	-0.068	0.320	0.347	-0.057
	(0.125)	(0.121)***	(0.119)***	(0.123)
H*Treated	0.021	-0.377	-0.055	-0.072
	(0.198)	(0.195)*	(0.193)	(0.196)
Treated*Post	0.073	-0.077	0.345	0.182
	(0.178)	(0.160)	(0.172)**	(0.163)
H*Treated*Post	0.183	0.534	-0.359	-0.051
	(0.197)	(0.192)***	(0.192)*	(0.191)
Mean	-0.143	-0.143	-0.143	-0.143
N	666	666	666	666
R^2	0.020	0.043	0.038	0.021

Independent variable: Mindset Index

H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss work (8)
Treated	-0.072	0.071	0.069	-0.163
	(0.133)	(0.186)	(0.123)	(0.171)
Post	-0.327	-0.327	-0.326	-0.325
	(0.090)***	(0.090)***	(0.090)***	(0.090)***
H	0.272	0.168	0.274	-0.101
	(0.121)**	(0.124)	(0.142)*	(0.121)
H*Treated	0.241	-0.027	-0.157	0.273
	(0.192)	(0.214)	(0.212)	(0.205)
Treated*Post	0.292	0.369	0.111	0.158
	(0.171)*	(0.194)*	(0.149)	(0.200)
H*treated*Post	-0.278	-0.303	0.138	0.013
	(0.192)	(0.208)	(0.205)	(0.213)
Mean	-0.143	-0.143	-0.143	-0.143
N	666	662	666	666
R^2	0.048	0.023	0.030	0.022

Note: Standard errors in parentheses. All errors are clustered at the individual level.Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D8: Heterogeneity in treatment impact on hours of study at home

Panel A: Dependent variable: Hours study at home				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	-0.120	0.268	0.128	0.098
	(0.248)	(0.225)	(0.262)	(0.244)
Post	0.037	0.027	0.033	0.051
	(0.148)	(0.149)	(0.149)	(0.151)
H	-0.063	0.687	0.340	-0.245
	(0.208)	(0.210)***	(0.212)	(0.210)
H*Treated	0.382	-0.409	-0.099	-0.069
	(0.344)	(0.347)	(0.347)	(0.349)
Treated*Post	-0.061	0.013	-0.058	-0.041
	(0.260)	(0.253)	(0.248)	(0.240)
H*Treated*Post	0.028	-0.114	0.029	-0.066
	(0.293)	(0.291)	(0.293)	(0.299)
Mean	3.366	3.366	3.366	3.366
N	666	666	666	666
R^2	0.005	0.022	0.008	0.008

Panel B: Dependent variable: Hours of study at home				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	0.004	0.314	-0.199	-0.097
	(0.203)	(0.308)	(0.193)	(0.260)
Post	0.034	0.019	0.032	0.043
	(0.149)	(0.149)	(0.148)	(0.150)
H	0.325	0.687	0.596	0.184
	(0.214)	(0.201)***	(0.232)**	(0.221)
H*Treated	0.183	-0.286	0.572	0.261
	(0.353)	(0.367)	(0.374)	(0.333)
Treated*Post	-0.015	-0.249	0.175	-0.100
	(0.241)	(0.311)	(0.237)	(0.277)
H*Treated*Post	-0.087	0.321	-0.573	0.062
	(0.298)	(0.322)	(0.296)*	(0.296)
Mean	3.366	3.366	3.366	3.366
N	666	662	666	666
R^2	0.014	0.026	0.051	0.010

Note: Standard errors in parentheses. All errors are clustered at the individual level. The dependent variable, 'Hours study of at home' is the hours per day that the student reports spent studying. The H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D9: Heterogeneity in treatment impact on hours of study at home

Panel A: Dependent variable: Days college attended				
H:	HH assets	Mother education	Career goal importance	Raven's score
	(1)	(2)	(3)	(4)
Treated	0.220	0.125	-0.070	0.116
	(0.232)	(0.231)	(0.235)	(0.239)
Post	0.413	0.413	0.413	0.413
	(0.174)**	(0.174)**	(0.174)**	(0.174)**
H	0.002	-0.040	-0.132	0.082
	(0.261)	(0.265)	(0.282)	(0.280)
H*Treated	0.164	0.009	-0.415	0.211
	(0.178)	(0.178)	(0.175)**	(0.176)
Treated*Post	-0.496	-0.329	0.075	-0.344
	(0.316)	(0.320)	(0.316)	(0.316)
H*Treated*Post	0.084	0.183	0.367	-0.088
	(0.318)	(0.326)	(0.321)	(0.321)
Mean	4.894	4.894	4.894	4.894
N	732	732	732	732
R^2	0.023	0.020	0.028	0.021

Panel B: Dependent variable: Days college attended				
H:	Compet-itiveness	First division	First year	Discuss HW
	(5)	(6)	(7)	(8)
Treated	-0.184	-0.064	-0.131	0.245
	(0.255)	(0.301)	(0.220)	(0.313)
Post	0.413	0.399	0.413	0.413
	(0.174)**	(0.174)**	(0.174)**	(0.174)**
H	0.263	0.173	0.224	-0.241
	(0.289)	(0.313)	(0.262)	(0.353)
H*Treated	0.128	-0.109	-0.287	0.121
	(0.177)	(0.215)	(0.205)	(0.214)
Treated*Post	0.348	0.100	0.317	-0.392
	(0.312)	(0.342)	(0.342)	(0.353)
H*Treated*Post	-0.476	-0.246	-0.448	0.417
	(0.317)	(0.325)	(0.336)	(0.360)
Mean	4.894	4.894	4.894	4.894
N	732	726	732	732
R^2	0.023	0.016	0.024	0.020

Note: Standard errors in parentheses. All errors are clustered at the individual level. The dependent variable, 'Days college attended' are the days (out of 6) in the last (open) week that the student is present in college. The H*Treated*Post is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D10: Heterogeneity in treatment impact in percentage scored in exams two months after the intervention

Dependent variable: Exam percent scored at two months				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	0.015	0.015	0.002	-0.008
	(0.023)	(0.021)	(0.023)	(0.021)
H	0.024	0.035	-0.001	0.015
	(0.023)	(0.023)	(0.023)	(0.023)
H*Treated	-0.046	-0.047	-0.017	0.004
	(0.032)	(0.032))	(0.032)	(0.033)
Mean	0.488	0.488	0.488	0.488
N	327	327	327	327
R^2	0.007	0.008	0.002	0.004

Dependent variable: Exam percent scored at two months				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	-0.021	0.036	0.007	-0.002
	(0.022)	(0.028)	(0.020)	(0.028)
H	0.007	0.089	0.061	-0.001
	(0.023)	(0.024)***	(0.025)**	(0.025)
H*Treated	0.031	-0.053	-0.050	-0.006
	(0.032)	(0.034)	(0.033)	(0.034)
Mean	0.488	0.488	0.488	0.488
N	327	326	327	327
R^2	0.009	0.041	0.020	0.001

Note: Standard errors in parentheses. All errors are clustered at the individual level. Percent scored refer to the marks secured by respondent in end-of-year examinations (2 months after the intervention) as a percentage of maximum (full) marks allowed for respective year of study. H*Treated is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D11: Heterogeneity in treatment impact in percentage scored in exams nine months after intervention

Dependent variable: Exam percent scored at nine months				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	0.019	0.019	-0.020	-0.002
	(0.033)	(0.032)	(0.036)	(0.033)
H	-0.014	0.011	0.003	0.023
	(0.036)	(0.036)	(0.035)	(0.035)
H*Treated	-0.039	-0.047	0.038	-0.001
	(0.051)	(0.051)	(0.050)	(0.051)
Mean	0.403	0.403	0.403	0.403
N	332	332	332	332
R^2	0.007	0.003	0.004	0.002

Dependent variable: Exam percent scored at nine months				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	-0.011	0.054	0.008	0.020
	(0.035)	(0.045)	(0.031)	(0.049)
H	0.003	0.068	0.004	0.016
	(0.035)	(0.037)*	(0.038)	(0.038)
H*Treated	0.018	-0.077	-0.025	-0.032
	(0.050)	(0.054)	(0.053)	(0.057)
Mean	0.403	0.403	0.403	0.403
N	332	329	332	332
R^2	0.001	0.008	0.001	0.001

Note: Standard errors in parentheses. All errors are clustered at the individual level. Percent scored refer to the marks secured by respondent in end-of-year examinations (9 months after the intervention) as a percentage of maximum (full) marks allowed for respective year of study. H*Treated is the long-term heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D12: Heterogeneity in treatment effect on willingness to modify academic goals to SMART academic goals

Dependent variable: Willing to modify goals				
H:	HH assets (1)	Mother education (2)	Career goal importance (3)	Raven's score (4)
Treated	-0.066	-0.028	-0.049	-0.090
	(0.068)	(0.067)	(0.063)	(0.065)
H	-0.091	-0.016	-0.182	-0.036
	(0.070)	(0.069)	(0.068)***	(0.070)
H*Treated	0.132	0.054	0.090	0.213
	(0.097)	(0.097)	(0.096)	(0.096)**
Mean	0.687	0.687	0.687	0.687
N	366	366	366	366
R^2	0.006	0.001	0.024	0.019

Dependent variable: Willing to modify goals				
H:	Compet-itiveness (5)	First division (6)	First year (7)	Discuss HW (8)
Treated	-0.016	0.101	0.058	-0.012
	(0.068)	(0.090)	(0.061)	(0.094)
H	0.050	0.033	0.117	0.075
	(0.069)	(0.081)	(0.071)*	(0.080)
H*Treated	0.033	-0.157	-0.177	0.017
	(0.097)	(0.107)	(0.100)*	(0.109)
Mean	0.687	0.687	0.687	0.687
N	366	363	366	366
R^2	0.005	0.008	0.009	0.007

Note: Standard errors in parentheses. All errors are clustered at the individual level. Modified academic goal is a binary variable equal to 1 if the respondent was willing to modify their goals into Specific, Measureable, Achievable, Realistic and Time based (SMART) goals with the help of the enumerators. H*Treated is the immediate post-intervention heterogeneous treatment effect. N refers to the sample size. H*Treated is the post-intervention heterogeneous treatment effect. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

OA5 Robustness checks

OA5.1 Controlling for individual characteristics (Post-double-selection with LASSO)

Table E13: Long term impact of treatment, using PDS Lasso

	Grit	Growth mindset	Self efficacy	Hours study	Days attend
	(1)	(2)	(3)	(4)	(5)
Treated	-0.954	0.897	0.191	-0.037	0.103
	(0.574)*	(0.658)	(0.500)	(0.075)	(0.211)
Post	-0.689	-2.753	0.047	0.048	0.311
	(0.516)	(0.707)***	(0.473)	(0.168)	(0.228)
T*Post	1.753	0.291	-0.219	0.087	-0.041
	(0.739)**	(0.973)	(0.685)	(0.236)	(0.287)
Mean	41.125	48.660	32.938	3.366	4.766
N	666	666	666	461	461

Note: Standard errors in parentheses. All errors are clustered at the college level. Fixed effects include college and enumerator fixed effects. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. T*Post is the average treatment effect of our treatment intervention. Controls are selected using PDS Lasso (Belloni et al., 2013). N refers to the sample size.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E14: Impact of treatment on end-of-year test performance, using PDS Lasso

Time after baseline	2 months	9 months
	Score %	Score
	(1)	(2)
Treated	-0.019	0.005
	(0.018)	(0.029)
Mean	0.488	0.403
N	230	227

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects.Test score percentage' are the marks scored in end-of-year examination as a percentage of the maximum (full) marks allowed for that enrollment year. These were measured 2 months after the baseline activities and are present for only one round of the data. N refers to the sample size. Controls are selected using PDS Lasso (Belloni et al., 2014a,b). N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

OA5.2 Upper and Lower Bound of treatment estimates using Lee (2009)

Table E15: Lee bounds estimate of primary outcomes in the immediate and longer term

Longer run effect	Grit (1)	Growth mindset (2)	Self Efficacy (3)	Mindset Index (4)	Days attend (5)	Hours study (6)
Treated*Post	1.354	0.451	-0.281	0.145	-.029	-0.136
Lower bound	0.363	-0.588	-1.328	-0.025	-0.317	-0.382
	(0.929)	(1.252)	(0.838)	(0.159)	(0.274)	(0.259)
Upper bound	2.166	1.726	0.607	0.292	0.325	0.125
	(0.803)***	(1.314)	(0.808)	(0.174)*	(0.302)	(0.234)
Trimming prop.						6.1%
N	366	366	336	366	366	366

Note: This provides the bounds on change in the outcome variable over the longer run, at the time of the follow-up, when 300 respondents could be surveyed out of the 366 respondents interviewed at baseline. Standard errors in parentheses. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale measured at the time of follow-up. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale measured at the time of follow-up. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale measured at the time of follow-up. Mindset Index is an Anderson index constructed from grit, growth mindset and Self-efficacy measured at the time of follow-up. Coefficients reported in the 'Treated' row are from a simple OLS regression on change in the outcome variable since baseline on treatment status. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table E16: Lee bounds estimates of test score outcomes

Time after baseline	2 months Score %	9 months Score %
Treated	-0.006	-0.003
	(0.016)	(0.025)
Lower bound	-0.011	-0.004
	(0.017)	0.029
Upper bound	0.002	0.013
	(0.018)	(0.029)
Trimming prop.	2.2%	1.8%
N	357	341

Note: Standard errors in parentheses. Test score percentage' are the marks scored in end-of-year examination as a percentage of the maximum (full) marks allowed for that enrollment year. These were measured 2 and 9 months after the baseline activities and are present for only one round of the data. Coefficients reported in the 'Treated' row are from a simple OLS regression on change in the outcome variable since baseline on treatment status. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

OA5.3 Restricting analysis to the sample that participated in face-to-face surveys at endline

Table E17: Sample description of face-to-face sample

	N	Mean	Median	S.Dev.	$p - value$
	(1)	(2)	(3)	(4)	(5)
Age (years)	226	19.37	19.00	1.61	0.03^{**}
Dummy: Enrolled in first year	226	0.38	0.00	0.49	0.06^{*}
Dummy: Single	226	0.95	1.00	0.22	0.42
Father's education (years)	217	8.61	10.00	4.71	0.38
Dummy: Father is a business owner	226	0.23	0.00	0.42	0.96
Dummy: Father is a salaried worker	226	0.58	1.00	0.58	0.61
Mother's education (years)	223	6.36	8.00	5.25	0.64
Dummy: Mother is a business owner	226	0.01	0.00	0.09	0.18
Dummy: Mother is a salaried worker	226	0.07	0.00	0.26	0.79
Average monthly household income (000's)	163	35.94	25.00	38.17	0.64
Index: Household assets	226	-0.06	0.00	1.35	0.67
Risk preferences (higher is more risk averse)	226	4.25	4.00	1.55	0.18
Ravens test score (out of 10)	226	4.22	4.00	2.19	0.78
Scale: Competitiveness (out of 75)	226	58.23	58.00	8.39	0.69
Dummy: Continue education after graduating	226	0.91	1.00	0.29	0.95
Dummy: Passed last degree in first division	226	0.75	1.00	0.44	0.25
Satisfaction with academic performance	226	2.42	2.00	0.93	0.78
Daily hours studying at home in average week	226	2.92	3.00	2.12	0.46
Daily hours doing household chores in average week	226	6.56	7.00	6.19	0.49
Dummy: Discusses homework with parents	226	0.71	1.00	0.45	0.89
Dummy: Discusses goals with parents	226	0.91	1.00	0.28	0.52
Dummy: Pursue paid job	226	0.94	1.00	0.24	0.05^{**}
Dummy: Have a role model	226	0.35	0.00	0.48	0.59
P-value of F-statistic					0.54

Note: Column (5) shows $p - values$ from the balance test specified in equation 1 for 266 individuals who received the repeat treatment. The cells show the coefficient on treatment assignment when the variable in the row is regressed on the treatment assignment. F-statistic explains the overall significance of the model- the null hypothesis is that all coefficients on the independent variables are equal to zero. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table E18: Long term impact of treatment on self-reported measures, restricted to sample participating in face-to-face surveys

	Grit	Grit	Growth mindset	Growth mindset	Self efficacy	Self efficacy
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.362	-0.269	0.502	0.151	0.620	0.690
	(0.555)	(0.531)	(0.644)	(0.623)	(0.371)*	(0.360)*
Post	-0.226	-0.226	-2.755	-2.755	0.066	0.066
	(0.278)	(0.278)	(0.774)***	(0.774)***	(0.583)	(0.583)
T*Post	1.126	1.126	0.013	0.013	-0.533	-0.533
	(0.568)**	(0.568)**	(0.674)	(0.674)	(0.724)	(0.724)
Mean	40.991	40.991	49.179	49.179	32.722	32.722
N	452	452	452	452	452	452
R^2	0.003	0.004	0.042	0.043	0.002	0.003
FE	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale measured at the time of follow-up. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale measured at the time of follow-up. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale measured at the time of follow-up. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. T*Post is the average treatment effect of our intervention. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

OA5.4 Spillovers results

Table E19: Spillover effects: Indirect impact of treatment on primary outcomes on peers

	(1) Grit	(2) Grit	(3) Growth mindset	(4) Growth mindset	(5) Self efficacy	(6) Self efficacy	(7) Mindset index	(8) Mindset index
Friends with treated*Post	-1.437	0.756	-0.584	1.789	0.995	-0.845	-0.098	0.170
	(1.093)	(1.359)	(1.457)	(1.708)	(1.021)	(1.204)	(0.181)	(0.232)
Friends with treated*	1.567	1.196	0.489	-1.781	-0.061	0.866	0.189	0.015
Treated*Post	(1.016)	(1.389)	(1.160)	(1.571)	(0.867)	(1.146)	(0.172)	(0.240)
Mean	40.609	40.609	49.173	49.173	32.645	32.645	-0.167	-0.167
N	443	443	443	443	443	443	443	443
R^2	0.006	0.047	0.024	0.083	0.007	0.066	0.019	0.054
FE	No	Yes	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. All regressions control for Treatment status and 'post' time period. Coefficient on 'Friends with Treated*Post' is the indirect average long term treatment effect of treatment of participant i on j outcomes. Coefficient on 'Friends with treated *Treated*Post' is the average indirect long term effect treatment of participant i on j outcomes when j is also treated. Grit is measured using 12-item scale ranging from a 1 to 5-point Likert scale. Growth mindset is a scale constructed from the sum of 15 items on 1 to 6-point Likert scale. Self-efficacy is measured using 10-item scales on 1 to 4 Likert scale. Mindset Index is an Anderson index constructed from grit, growth mindset and self-efficacy. Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E20: Spillover effects: Indirect impact of treatment on measures of effort on peers

	Hours study (1)	Hours study (2)	Days attend (3)	Days attend (4)
Friends with treated*Post	-0.139	-0.098	-0.215	-0.221
	(0.305)	(0.300)	(0.378)	(0.375)
Friends with treated* Treated*Post	-0.009	-0.060	-0.238	-0.261
	(0.289)	(0.276)	(0.242)	(0.246)
Mean	3.256	3.256	4.809	4.809
N	443	443	443	443
R^2	0.001	0.128	0.026	0.075
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects. All regressions control for Treatment status and 'post' time period. Coefficient on 'Friends with Treated*Post' is the indirect average long term treatment effect of treatment of participant i on j outcomes. Coefficient on 'Friends with treated *Treated*Post' is the average indirect long term effect treatment of participant i on j outcomes when j is also treated. 'Hours study' is the number of hours students report studying at home. 'Days attend' are the number of days the student attended college in the most recent full week. T*Post is the indirect average treatment effect of our treatment of participant i on j outcomes. 'Mean' is the average value of the outcome for the control group. N refers to the sample size.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E21: Spillover effects: Indirect impact of treatment on exam performance of peers after 2 and 9 months months

Time after baseline	2 months		9 months	
	Score	Score	Score	Score
	(1)	(2)	(3)	(4)
Friends with treated	-0.014	-0.026	-0.034	-0.033
	(0.025)	(0.023)	(0.040)	(0.039)
Friends with treated*Treated	-0.020	-0.022	0.040	0.037
	(0.025)	(0.024)	(0.037)	(0.035)
Mean	0.495	0.495	0.401	0.401
N	214	214	217	217
R^2	0.009	0.129	0.006	0.191
FE	No	Yes	No	Yes

Note: Standard errors in parentheses. Fixed effects include college and enumerator fixed effects.Test score percentage' are the marks scored after 2 and months of the intervention. Treated is the indirect average treatment effect of our treatment of participant i on j outcomes. 'Mean' is the average value of the outcome for the control group. N refers to the sample size. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.