Changing Beliefs and Attitudes of Primary School Children: Growth Mindset in Rural China

Li Zhou[1] Sonja Fagernäs[2] Iftikhar Hussain[3]
Lingzhi Li[4] Panu Pelkonen[5] Zhen Huang[6]

06 February 2023

Abstract

We conduct a randomized controlled trial to test the impact of a non-cognitive skills intervention on grade 5 pupils in rural China. Treated students, half of whom live in school dorms, follow an intensive growth mindset program. We study the effects of the training on stated attitudes, choices and performance in classroom effort task experiments, and academic results over three terms. Treated pupils improve their mindset orientation and are more likely to choose more ambitious tasks in experiments, despite a higher risk of failure. We find little evidence of significant effects on test scores, or of within-dorm spillover effects.

JEL:I20; I28; C93. Keywords: Noncognitive skills; Mindset; Schooling; Field experiments; Rural China

1-Professor, College of Economics and Management, Nanjing Agricultural University (zhouli@njau.edu.cn); 2-Senior Lecturer, Department of Economics, University of Sussex Business School (s.a.e.fagernas@sussex.ac.uk), 3-Lecturer, Department of Economics, University of Sussex Business School (i.hussain@sussex.ac.uk), 4-Researcher, College of Economics and Management, Nanjing Agricultural University (llz0405@qq.com) 5-Senior Lecturer, Department of Economics, University of Sussex Business School (p.o.pelkonen@sussex.ac.uk), 6-Academic Team Leader, Tsinghua University School of Sciences, Positive Psychology Research Center (huang_zhen@mail.tsinghua.edu.cn).

Acknowledgements: The authors acknowledge the financial support from the Chinese National Social Science Foundation (19ZDA117), the National Natural Science Foundation of China (No. 72141011), Office of Philosophy and Social Sciences of Guizhou Province (21GZGX05), the Priority Academic Program Development of Jiangsu Higher Education Institutions, China (PAPD), and the China Center for Food Security Studies in Nanjing Agricultural University, China.

I Introduction

The role of children's socioemotional and noncognitive skills in shaping subsequent schooling, labor market and related outcomes is well recognized (Bowles and Gintis 2001, Borghans et al 2008, Heckman et al. 2006, Almlund et al 2011, Deming 2017, Lindqvist and Vestman 2011). This literature has demonstrated, for example, that noncognitive skills and personality traits measured early in life have a payoff on a wide array of adult outcomes, over and above any impact of cognitive skills and years of schooling. The evidence also suggests that noncognitive skills may be more malleable than cognitive skills (Kautz et al., 2014). Thus, investments in noncognitive skills may well have the potential to reduce the often large gaps in outcomes arising from family background, race or gender. Some of these findings from industrial countries are also supported by more recent evidence emerging from low- and middle-income countries, although the evidence base is much thinner (Glewwe et al. 2017, 2022).

Experimental studies which specifically aim to manipulate or raise children's noncognitive skills have found mixed evidence on the efficacy of alternative strategies (for overviews, see Kautz et al., 2014, and Smithers et al., 2018). Recent experimental work in psychology and related fields, including economics, has focused on 'grit' and 'growth mindset (see e.g. Alan et al., 2019, Bettinger et al., 2018, Dobronyi et al., 2019, Foliano et al. 2019, Oreopoulos and Ptronijevic, 2019, Paunesku et al., 2015). Grit refers to the idea that individuals must exercise resilience and perseverance, overcome challenges and maintain interest despite failure, in the pursuit of medium- and long-term goals (Duckworth et al., 2009). Growth mindset on the other hand emphasizes the malleability of the brain. It challenges the ides of a fixed ability and tries to instill in individuals the belief that skills and ability can be improved with deliberate effort and sustained practice (Dweck, 2006). It provides students, for instance, with an optimistic view on their ability to improve, rather than holding a fixed view of 'their

place' in terms of educational achievement. Park et al. (2020) discuss the relationship between the two related, but distinct concepts of grit and growth mindset.

Our study contributes to the literature on noncognitive skills and childhood development and, more specifically, the literature focused on growth mindset and grit. We study the impact of a relatively intensive growth mindset program, implemented as a randomized controlled trial in rural China. The setting for our experiment is Majiang county, which lies in Guizhou, a land-locked, mountainous, and relatively poor province in southwest China. Our sample includes over 1,600 5^{th} grade students in state-funded primary schools. We investigate the impact of our mindset intervention on mindset attitudes and beliefs and other noncognitive skills such as grit, as well as test scores and a real effort task implemented in the classroom.

Findings from the growth mindset literature have been mixed (for reviews see e.g. Sisk et al., 2018, and Yeager and Dweck, 2020). A review by Sisk et al. (2018) for instance concludes that mindset interventions did not influence the academic achievement of typical students or those facing challenging situations. However, academically high-risk students and economically disadvantaged students might benefit from such interventions.

Many of the experimental studies on the impact of growth mindset rely on relatively short or light treatments, such as courses lasting up to an hour or two, or being delivered online (eg. Paunesku et al 2015, Bettinger et al., 2018). Bettinger et al. (2018) find positive effects on effort in mathematics and the performance of students who had poorer initial attitudes to learning. On the other hand, with an online intervention for university students, Dobronyi et al. (2019) find precisely estimated null effects for grade point average and course credits. Similar results for university students are found by Oreopoulos and Ptronijevic (2019).

In contrast to these light touch intervention studies, Alan et al. (2019) and Foliano et al. (2019) investigate more intensive programs undertaken by primary school classroom teachers.

The teachers are encouraged to adopt the program in their regular classroom curriculum. While Foliano et al. find little impact on academic achievement, in Alan et al. (2019) the intensive program yields strong positive results. The strengths of this study are a broad set of outcome variables, a carefully thought-out real effort task, as well as the focus on longer term outcomes.

Several features of our intervention and setting distinguish our study from previous studies. Firstly, we contribute to the literature by providing evidence from an emerging market setting, in a rural location which suffers from significant disadvantage. A significant share of the students are classified as low-income, and have absent parents, who work away from home as migrant wage laborers. As a consequence, around half of the students in our sample reside in state-run government-funded boarding schools. [1]

Secondly, in contrast to some of the previous studies we undertake a relatively intensive intervention, which involves the training of teachers by a professional psychologist with expertise in the growth mindset approach. Treated students are all in grade 5 and receive lessons in growth mindset from these trained teachers in five weekly two-hour sessions during the normal school day. Teachers implementing the student training sessions are from the same school, but teach in earlier grades (grades 2, 3 or 4). We view our treatment as lying somewhere in the middle of the spectrum of very light touch and highly intensive treatments.

Finally, a number of aspects of our research design allow us to gather a rich set of results. (i) In addition to students' beliefs, we measure parents' or carer's mindset beliefs at baseline, enabling us to better understand any interactions with parental beliefs or the home environment, an underexplored dimension of growth mindset studies. By incorporating variables on family circumstances and parental mindset, we contribute to the literature on inequality in children's socioemotional and noncognitive skills and the role of family circumstances (Attanasio et al.

[1] Lee and Park (2010) discuss the socioemotional consequences for children of fathers' migration in rural China. The impact of growth mindset training has been studied less in the Chinese context. Some existing studies find a positive association between growth mindset and academic performance or motivation in upper elementary school or junior high school in a few Chinese cities (e.g., Huang et al., 2022; Zhao et al., 2018).

2020). (ii) We investigate spillover effects within school dormitories, where peer interaction may be especially intense. There is little or no research in the prior literature on spillover effects on other students. (iii) We investigate fadeout in treatment impact on test scores as well as the real effort task. (iv) We assess for the importance of intrinsic versus extrinsic motivation by implementing the real effort task under incentivized as well as unincentivized conditions.[2]

In our study, students in each school are randomized into one of three groups: a treatment group who receive the mindset intervention; a placebo group who watch a nature documentary for the equivalent time; and a pure control group who are free to use their time as they please. The treatment takes place at the beginning of grade 5 in the fall. Outcomes are collected both soon after the intervention is complete (the 'midline') and also six months later (the 'endline'). In both the midline and endline, students take part in an effort task where they opt for easier or harder puzzles, and try to solve them both with and without incentives. In our student survey we also collect information on their noncognitive skills (mindset, grit and goal setting). In addition, we assess the impact of the treatment on test scores in official exams administered by the county in mathematics and Chinese.[3]

We document substantial gaps in noncognitive skills and mindset beliefs at baseline by both poverty status, parental attitudes and presence at home, boarding status as well as by gender. Furthermore, we also uncover large gaps by parental involvement in education (as measured by a question asking students whether parents help them with homework); these gaps by parental involvement remain even after we control for poverty status as well as test scores.

[2] Studies have demonstrated that motivation and effort during tests can be affected by test performance incentives. See the literature on this issue extensively surveyed in Almlund et al., 2011.

[3] A random subset of parents and carers of the treated students were also offered a growth mindset treatment. The treatment took the form of an information package, followed by the option to discuss the approach with volunteers. However, take-up of the parental component of the treatment was very limited, and did not lead to significant impact. Hence, we view our treatment primarily as one targeting students in a school setting. The growth mindset literature on the potential relevance of treating parents is limited, with Andersen & Nielsen (2016) being one exception in a developed country context.

With respect to our experimental results, our first core finding is that the treatment leads to a positive impact on mindset attitudes and beliefs, measured six months after the end of the treatment. Relative to the control group, treated students are more likely to agree with statements such as the following: "No matter what your intelligence, you can always improve," and "It's possible for anyone, including me, to improve their grades."

We do not find any significant impact on measures of grit or goal orientation, measured separately from mindset attitudes. However, we do find significant positive effects on the likelihood of opting for more challenging tasks in the real effort task. This is the case even when these choices increase the likelihood of a perceived failure in a task. The effect persists in the endline, although the magnitude is smaller. Furthermore, students opt for challenging tasks with and without incentives, providing suggestive evidence that the results are driven by improvements in intrinsic motivation.

We do not find statistically significant effects on test scores – Chinese and Mathematics – one, two or three semesters after the intervention. We find that there is heterogeneity in the treatment effect for the growth mindset outcome, and to some extent, for the real-effort task. In particular, when analyzing sub-samples, we see greater improvements in growth mindset orientation among students who themselves or whose parents had higher growth mindset scores at baseline and who had initially better academic achievement. We also find that the gains are relatively larger for children who live with their parents or are not boarding in the school. On the other hand, the treatment raises growth mindset scores more for those from low-income families. We test for behavioral spillovers within dorms but find no evidence for spillovers in growth mindset orientation or the effort task.

Finally, we use non-experimental methods (value added models) to assess the impact that the teachers trained by this experiment have on the students in their own (non-experimental)

classrooms.[4] This allows for an assessment of whether the teachers trained by this study 'internalize' the mindset training and whether this leads to changes in students' test scores. However, once again we find no evidence of any impact on test scores.

Our findings demonstrate that noncognitive skills, as captured by growth mindset attitudes and beliefs, are malleable and can be boosted via a teacher-led intervention. This manipulation of growth mindset appears to have consequences as measured in the real effort task, but we do not find any significant impact on students' exam performance. The finding on exam performance is in accordance with findings from several other studies, but contrasts with the findings of Alan et al. (2019), which influenced our design.

Our findings on heterogeneous impacts also contrast with the view that mindset interventions would be particularly helpful for academically high-risk students (e.g. Sisk et al., 2018). We speculate that weaker students, those with lower initial mindset scores and boarders could be less responsive to the intervention possibly due to lower aspirations. Overall, our results suggest that the treatment did not reduce student gaps by family background, but may even have increased them.

It is possible that due to the lack of awareness of the growth mindset approach in our setting, we are unlikely to suffer from the well-known problem facing randomized controlled trials regarding close substitutes for the treatment (Heckman et al., 2000, and Kline and Walters, 2016). In the typical growth mindset evaluation undertaken in education settings in high income countries, there is widespread knowledge and training in the growth mindset approach. For example, Foliano et al. (2019) speculate that this could explain why they find no impact in their growth mindset study with a UK sample.

[4] For details of the value-added model approach, see for example, Andrabi et al. (2011), Jacob et al. (2010) and Todd and Wolpin (2003).

The rest of the paper is organized as follows. Section II describes the context, the data and descriptive evidence. Section III explains the growth mindset training and Sections IV and V describe the experimental design and the real effort task respectively. Sections VI and VII present the results and Section VIII concludes.

II Context and Descriptive Statistics

The intervention took place in Majiang county which lies within the Qiandongnan Miao and Dong prefecture, in Guizhou province. Guizhou province is an economically disadvantaged area in southwest China with average per capita disposable income of RMB 21,795 (US$ 3,410) in 2020, equivalent to two thirds of the national average (China Statistical Yearbook, 2021). Majiang has a population of about 168,000. There are 6 towns and two subdistricts in Majiang and the county is served by 21 primary schools.

Our study covers children in all 21 primary schools of Majiang county, which are all state-funded. Less developed rural areas in China commonly have large numbers of 'left-behind' children, with at least one parent who has migrated to an urban area for work. The benefits of China's unprecedented growth have been concentrated in urban areas, resulting in rural-to-urban migration, but China's strict domestic migration policies prevent rural workers from migrating to urban areas with their children (Liu and Villa, 2020). With 166 million rural migrants in 2013, China has an estimated 61 million left-behind children, which amounts to 38% of China's population of rural children (All-China Women's Federation, 2013, National Bureau of Statistics of China, 2014). Left-behind children have substantially worse outcomes than children without migrating parents on numerous dimensions, including education and socio-emotional well-being (Chen et al., 2009; He et al., 2012; Wang and Mao, 2018). There were 0.57 million left-behind children in Guizhou in 2017.

Since 2004, the Chinese government has undertaken a massive construction of rural boarding schools as a way to meet its education objectives in rural areas. As of 2018, 9.35 million students in rural China were boarding in school (Ministry of Education of China, 2018). As we document below, in our sample 21% of students are left-behind children where both parents work in cities and almost half of the sample students board in school.

II.A Data and Survey Instruments

We conducted the baseline survey in early September 2020 before the intervention to gather information on students' (i) beliefs related to their 'mindset', (ii) attitudes and behaviors related to grit, and (iii) attitudes related to goals. The survey also asks for information on attitudes to school work and parental help with homework.[5] We match this survey information with school registry data. The school registry data includes information on the age and gender of the students; their performance on Chinese and mathematics tests taken in January 2021, July 2021 and January 2022; whether the family is officially classified as low income; whether the child's parents work away from home (and therefore are not normally at home in the week) and whether the child is a boarder at the school. The Chinese and mathematics test results are com

The questions that we use to measure growth mindset orientation and grit are largely similar to those used by Alan et al. (2019). These are adaptations from the growth mindset scale by Dweck (2006) and grit scale by Duckworth and Quinn (2009). Our indicators for goal orientation are adaptations based on Mingley et al. (2000). The questionnaires are based on a set of statements and the respondents are asked to agree or disagree on a scale of 1-4 (completely agree; agree; disagree; completely disagree). The scores from these questions are then added together (with scores reversed for questions where agreement implies worse

[5] The full survey questionnaire is laid out in the Online Appendix.

mindset, grit or goal orientation skills and attitudes) to construct a raw composite score, which is then normalized to have a mean of zero and a standard deviation of one. The full set of questions used to measure mindset, grit and goal orientation can be found in Appendix 1.

Although schools in Majiang were largely unaffected by the Covid-19 pandemic from mid-2020 to mid-2021, nevertheless as a precaution we used online questionnaires for the students at the baseline. Students filled out the questionnaire on school computers, and in some cases used teachers' smart phone. Parental questionnaires relating to mindset and attitudes were sent to the parents by headteachers via the WeChat platform.

II.B Descriptive Statistics

Table 1 shows summary statistics at baseline. Column 1 shows the mean value for the control group (the balance tests reported in columns 2 and 3 are discussed below). The school registry data in panel A of Table 1 shows that 30 percent of students come from a low-income family background, almost half of the students board in the school, and parents of around 20 percent of the students work away from home. Chinese and mathematics scores from the prior grade (grade 4) are also reported in panel A.

Panel B reports summary statistics from the student survey, including the normalized growth mindset, grit and goal orientation indices. In addition, panel B also reports detailed information provided by students on questions relating to time spent on homework, parental and family relations, including who the main carer is, who usually helps them with homework, how often parents check on their work and whether either parent discuss the child's school, friendships or worries with them.

The variable definitions can be found in Appendix Table 1. Figure 3 presents the distribution of raw scores for the baseline orientations of the students; for growth mindset, grit and goal orientation. Overwhelmingly, the responses are above the midpoint of the scale

suggesting that on average the students' attitudes were geared towards growth mindset, grit and goal orientation, albeit to a varying degree. A recent PISA study found that growth mindset attitudes among 15-year old Chinese students in large cities fall within the global average and are not far from the OECD average (OECD, 2021).

II.C Correlates of Noncognitive Skills

Appendix Table 2 presents correlates of the noncognitive skill measures. Column 1 shows that males, low-income students, boarders and those left behind all suffer from large disadvantage in growth mindset. The connection between growth mindset and poverty has been discussed previously for instance by Claro et al. (2016). In column 2 we add a measure of parental involvement with the student's work, parental growth mindset and academic test scores. We use the question on whether anyone helps the student with their homework. Students who report that no-one helps them with homework report substantially lower growth mindset scores relative to the omitted category reporting that someone other than a parent helps them with homework. Perhaps unsurprisingly, parental mindset, test scores and growth mindset are positively correlated. We repeat this analysis for grit and goal orientation in columns 3 to 6. Similar patterns are observed, although the significance of some of these variables is somewhat weaker for the goal orientation outcome.

Online appendix Table 4 presents correlations between the three noncognitive skill measures, growth mindset, grit and goal orientation. This shows that there is quite a strong correlation (0.54) between mindset growth and grit, but the correlation between goal orientation and the other two measures is somewhat weaker (0.3).

III Growth Mindset Intervention

Our intervention was supported by the education bureau in Majiang, which recognized the relevance of the growth mindset curriculum and encouraged all primary schools to participate in the intervention.

III.A Training of School Teachers

The growth mindset program was conducted by regular teachers in each primary school, but who did not teach Grade 5 students, the focus group for our intervention. In June 2020, two to four volunteer teachers who were willing to participate in the program and normally taught Grade 2 to 4 students were selected from each school. The 30 volunteer teachers participated in a growth mindset training program[6]. The first stage of the program was carried out over the course of two days (July 20 to 21, 2020). The instructor, a psychologist trained in the Growth Mindset approach, and one of the co-authors of our study (Huang) first introduced the concepts and their importance for student development. The training program was based primarily on a teacher's handbook on the growth mindset by Brock and Hundley (2016). Teachers then received detailed instructions on how to cover the curriculum, which was shaped by the instructor. The program took the form of a 10-hour course delivered over five weeks. It consisted of lectures, videos and activities with interactive and reflective content (such as essays, posters, drawings or discussion). The students were taught about the structure of the brain and how training and exercise improves skills by creating new connections in the brain. Fixed ability was contrasted with personal improvement, or 'growth mindset'. Learning was presented as a process, that rewards perseverance, can incorporate failures and difficulty and how to approach these constructively. The program ends by emphasizing goal setting and habit formation around the growth mindset and self-discipline.

[6] Among the 30 trained teachers, 18 teachers taught Grade 2, 6 taught Grade 3 and 6 taught Grade 4.

Appendix 1 provides the outline of the course content. The seminar was structured in an interactive manner. The instructor guided the teachers through the materials and suggested classroom activities, aiming to actively engage the teachers to illustrate the relevant concepts in different activities.

One month later (August 25 to 26, 2020), the second stage of the program was carried out by the same instructor with the purpose of ensuring that each teacher was familiar with the mindset approach and was ready to implement the mindset curriculum.

At the end of the academic year, we conducted an anonymous survey among trained teachers and asked for their views on the mindset curriculum. Out of the 30 volunteers, 13 filled in the survey. Around 85% found the training course useful a similar proportion agreed that it could be popularized in the general curriculum.

III.B Implementation of the Intervention

Before the intervention began, Grade 5 students within each school were randomly allocated to one of three groups. Students in the treatment group took part in growth mindset sessions with the trained teachers. Simultaneously, students in the placebo group would watch the documentary series *Planet Earth* produced by the BBC. Students in the control group were free to learn by themselves or play in the school.

As explained above, relative to previous mindset interventions, we view this intervention to be intense in terms of duration and content. With respect to the treatment group, trained teachers spent five one-and-half-hour sessions covering and discussing the materials. Most trained teachers selected the last two class sessions on Wednesday afternoon for the program.

This would otherwise be free time for students in their schedule. The training took place in the fall semester, finishing by November 2020.

The intervention also contained a component for students' parents or carers: they were encouraged to adopt the growth mindset idea as part of a parenting philosophy. To do this, 60 undergraduate volunteers were recruited and trained. From mid-November to mid-December 2020, undergraduate volunteers delivered various guidelines in the form of videos, audios, and essays to the parents of treated students via WeChat once a day. The videos were made by the instructor responsible for the teacher training. The audios were obtained from a learning platform in China. The material encouraged parents to highlight the role of effort in success and encouraging perseverance and positive attitudes towards learning. However, the take-up of this parental component of the treatment was very low. It is therefore not surprising that this aspect of the treatment yielded no impact. We subsume the results of this exercise in online Appendix Tables 6 and 7.

IV Experimental Design and Empirical Strategy

Figure 1 provides the timeline for the experiment. As explained above, teacher training in the mindset approach was undertaken in July 2020. The baseline survey of students took place in September 2020 and the mindset intervention was implemented from October through December 2020. The mid-line survey instrument, including a real effort task, was implemented in January 2021. The end-line survey, which also included the real effort task, was undertaken in June 2021. Official school assessments took place in January 2021 and July 2021. The end-line survey gathered information on students' beliefs about mindset, attitudes and behaviors related to grit, and attitudes related to goal orientation.

Of the total 1,993 grade 5 students in the 21 schools, 1,680 completed the baseline survey, 1,647 completed the midline and 1,611 the end-line surveys. In total, we received 1,445 valid questionnaires for the parent survey. The relatively larger non-response is due to difficulty of reaching parents as opposed to reaching students in schools.

IV.A Balance

As explained in section IV, within each school, students were randomized to the treatment or two control groups in approximately similar proportions. The treatment group included 572 pupils, and the placebo and control groups 558 and 550 students respectively. In the main control group students have free time, and in the 'placebo' group the students watch a nature documentary for an equivalent time of the mindset training. Table 1, columns 2 and 3 assess balance between treatment and control groups at baseline. For this exercise, the 'placebo' treatment group is included in the control group.[7] Column 2 of Table 1 reports the difference between the treatment and control means and column 3 reports the p-value for the null hypothesis that the difference is equal to zero. Column 2 reports coefficients from a regression of the outcome in question on the treatment dummy, including school fixed effects; standard errors are clustered at the school level.

The results in columns 2 and 3 suggest that differences between treatment and control groups are generally economically small and statistically insignificant. The difference in the Chinese exam result is marginally significant, though the combination of Chinese and mathematics is not. In the analysis below we report results with and without baseline characteristics, in order to account for these differences between the control and treatment groups. These specification differences make little difference to our conclusions in practice.

[7] Balance between all three groups separately is reported in online appendix Table 5. These results are in line with those reported in Table 1.

Attrition rates are very low (overall 2.0% at mid-line and 4.1% at end-line) and are similar for the treated and control groups: 1.2% for treatment group and 2.3% for control group in the midline and 3.1% versus 4.6% for the treatment and control groups at the endline, respectively.

IV.B Empirical Strategy

We evaluate the impact of the experiment using the following model:

$$y_{is} = \beta_0 + \beta_1 D_i + \beta_2 P_i + X_i\theta + \lambda_s + \varepsilon_{is}\,, \qquad (1)$$

where y_{is} is the outcome of interest (for example, growth mindset score, test score, choice in effort task.) for student i enrolled in school s. D_i is the student's treatment status, switched on for those students enrolled in the mindset intervention, and set to zero otherwise. P_i is the placebo treatment, and vector X_i stands for detailed baseline student and family characteristics, including gender, low family income status, whether the student boards at school and details of the main carer. λ_s represent school fixed effects. We also include the lagged test score in the controls.

The main parameter of interest is β_1, which measures the impact of the treatment on the outcome. The placebo treatment β_2 and how it differs from β_1 may be of interest, if we are concerned that part of the treatment effect is due to the treatment restricting the activities of students from free play or work. The effect of such artificial restricting of time use is therefore directly measured by β_2 since the students are asked to watch nature documentaries for the equivalent time of the mindset treatment.

V Real Effort Task

The real-effort task is designed to elicit core aspects of grit: challenge seeking and perseverance through setbacks. Specifically, we elicit students' choices between a challenging and an easy task. The design of the real effort task is similar in nature to the one in Alan et al. (2019), with the exception that most of our tasks are unincentivized.

In these tasks, students are presented with a sequence of tasks. We conduct such tasks in both the midline and endline. An example of the actual task can be found in Appendix 2, and a detailed description of the tasks in the Online Appendix. In the midline, both easier and harder tasks get more difficult in each round, whereas in the endline the level of difficulty remains the same for each type of task in all rounds. All rounds are unincentivized in the midline, whereas the final round in the endline is incentivized. The reason for this variation in provision of incentives was to test whether there is any interaction between the mindset treatment on the one hand and intrinsic versus extrinsic motivation on the other.

The mid-line experiments were carried out in January 2021, shortly after the teachers had covered the 5-week curriculum. We visited all the primary schools in Majiang and conducted in-class experiments designed to measure effort or grit for all the Grade 5 students. More specifically, students participated in three rounds of a numerical real-effort task**.**

In the first task, students are presented with a grid which contains different numbers where the goal is to find several pairs of numbers that add up to 100 within 2 minutes. Half of the students are randomly presented with a large grid and the target is to find three pairs (a hard task). The other half are presented with a small grid and the target is to find two pairs (an easy task). The first round allows us to obtain a measure of the difficulty of the tasks without student choice.

In the second and third rounds, there are two grids on the booklet, and before each round starts, subjects have the chance to choose between a hard task and an easy task, i.e. from the second round onwards, students are completely free to make their own choices.

Before the three rounds start, the instructors show a grid to children and demonstrate how to find as many pairs of numbers that add to 100 as possible. This is intended to familiarize the children with the task before they make decisions and measure task-specific ability.

In the main three-round part of the experiment, subjects are handed a four-page booklet. The first page contains student details and there is one task per page on the following pages. Before the second or third round starts, subjects are instructed to circle their game of choice for the upcoming round on the booklet. It was repeatedly emphasized that there was no right or wrong decision in these games, everyone was different, and each student was free to choose as they pleased. The students are given 2 minutes to find as many matching number pairs as they can. All students are instructed to fold their arms once the 2 minutes are over. During this time, experimenters go around the class and circle either "Succeeded" or "Failed" on the students' booklet for that round, based on whether the required pairs were correctly found. Students have the opportunity to switch back and forth between the hard and easy tasks as the rounds progress.

In the experiment, the main outcome of interest is whether the pupils opt for the harder task voluntarily, even as they quickly learn that they are less likely to find the pairs in the given time, and have an examiner tick the box saying 'fail' in their booklet. Since most tasks carry no reward from solving the harder puzzle, such a choice can be interpreted as challenge-seeking behavior. We also study the effect on the likelihood of success, which would be expected to be lower with a hard choice, unless the experiment changes the students' problem-solving capacity.

The endline experiment was conducted at the same time as the endline survey, i.e. in June 2021, towards the end of the academic year. In the endline experiments, children again have 2 minutes to find pairs of numbers that add up to 100. This time the game is played for four (rather than three) rounds. In the first round, children are randomly allocated to a hard or easy task. In the following three rounds, children are required to make a choice between a hard and

an easy task. In both the hard and easy tasks, the goal is to find three pairs of numbers. The grid in the hard task is much larger than that in the easy task.

One difference relative to the midline real-effort task is that the additional fourth round is incentivized, and subjects are rewarded based on their performance in that round. The rewards include a box of markers and a lamp. These items were carefully selected and were of value to children of this age group. The markers were about 4 times lower in value than the lamp, and the children were aware of this. The hard task yields two gifts (a box of markers and a lamp) in the case of success and zero in the case of failure, and the easy task yields one gift (a box of markers) in the case of success and zero in the case of failure. Students are informed about the payoffs before the fourth round starts.

VI Main Results

In this section, we report the results for all outcomes of interest, starting with the impact on noncognitive skills. Next, we focus on the impact of the intervention on the effort task and then report the effect on test scores. The estimates for the treatment effects are presented with and without controls variables. This controls for instance for the fact that the initial grade 4 Chinese were higher for the treatment group at the 10 percent level (Table 1). In all cases, we also report results for a set of sub-samples to identify whether the impact is heterogeneous across groups.

VI.A Noncognitive Skills

Table 2 reports estimates of the impact of the treatment on the composite measures for growth mindset, grit and goal orientation at the endline (no information was collected on these outcomes at the midline).

The results in column 2 or Table 2 show that at the endline, six months after the intervention, the treated group report 0.13 standard deviations higher growth mindset orientation. This corresponds to approximately a 0.3 point increase in the score between 6-24 (sum of 6 questions). The placebo treatment has no impact. We find no evidence of any impact on grit or goal orientation (columns 3 to 6).

Table 3 reports evidence on heterogeneous treatment effects for the growth mindset outcome. We do so by reporting results for sub-samples as follows: girls versus boys; academically weak versus strong students (based on year 4 test scores); those with low versus high initial growth mindset scores; and those whose parents have low versus high growth mindset scores at baseline. We also estimate treatment effects by boarding status, low-income status, and being a 'left behind' child, whose parent(s) are migrant workers. The table only reports the coefficient for the treatment dummy for each sub-sample.

The results in Table 3 suggest that growth mindset orientation improves more among students who can be characterized as relatively advantaged. The treatment effects are larger among those with higher year 4 results and those with higher baseline growth mindset or parental growth mindset orientation. They are also larger for students who are not boarding and are not left behind. On the other hand, the treatment raises growth mindset scores more for those from low-income families.

These results are contrary to the view that mindset interventions would be particularly helpful for academically weaker students (e.g. Sisk aet al., 2018). We speculate that weaker students, those with lower initial mindset scores and boarders could be less responsive to the intervention due to lower aspirations. Boarding status for instance correlates with lower incomes, lower mindset attitudes, lower baseline performance and when asked, lower expectations of going to college in the future by both the children and their parents. It is possible that growth mindset interventions might have to be complemented with other approaches, or

be more intensive in nature to influence the attitudes of more disadvantaged students, with initially weaker growth mindset attitudes, grit and goal orientation and academic performance.

VI.B Real Effort Task

In the first round of the real effort task – when students are randomly assigned an easy or hard task, i.e. there is no choice – at midline, 196 out of 835 of students (24%) fail in the easy task, i.e. fail to find the pairs of numbers which add up to 100. Among those given the hard task, 665 out of 808 (82%) fail. Out of those given the easy task, 325 or 39% go on to choose a hard task in the second round, whereas for those assigned the hard first task, 217, or 27% go on to opt for a hard task.

Tables 4 to 6 lay out the main results for the impact of treatment on the real effort tasks in the midline and endline. Table 4 shows the number of 'hard' choices selected by students in the midline and endline. Columns 1 and 3 describe results without controls, and 2 and 4 with additional student characteristics included.

In both the midline and endline, we find significantly positive effects of the treatment on the number of hard choices. In the midline, the control mean is 0.55 hard choices selected (out of a total of 2) and the treatment effect is 0.09 (column 2), significant at the 5% level and implying a 16% increase in the likelihood of taking the more challenging option. In the endline the effect is smaller, since the control mean is 1.09 (out of a total of 3) and the treatment increases by .106. This implies a 10% increase.[8] Furthermore, in the endline, the result is only significant at the 10% significance level. The results are consistent with and without controls,

[8] We do not find that the Growth Mindset treatment would have made the students better at solving the puzzles (see Table 12).

and the effect of the placebo treatment is not significantly different from zero. The only individual characteristics that predict a higher likelihood of hard choices consistently are male gender and better performance in mathematics.

Overall, while the effort tasks were somewhat different in the mid- and endline, there appears to be a weakening treatment effect over time. However, if the treatment effects are compared against the placebo treatment – a valid comparison group in itself – the treatment increased the likelihood of a hard choice by 0.17-0.18 at the endline, which would suggest a similar magnitude to the midline results.

From the perspective of the students in the effort task, taking on the more challenging 'hard' option should lead to a lower rate of success, unless the treatment itself has improved the problem-solving capacity of students or made them try harder than they otherwise would. We do not find that treatment would have made the pupils better problem solvers; we test for this using the first task where difficulty is random (Results on Online Appendix Table 8). In Table 5 we regress the likelihood of success on the treatment, and find negative effects throughout, although these are only significant (at the 10% level) at the midline (column 3). Nevertheless, these results suggest that the main behavioral change arising from the treatment is to encourage the students to choose the more challenging option, despite a greater likelihood of failing.

In order to study whether the students are more likely to choose hard tasks with or without explicit incentives, they were rewarded with a gift for a correctly solved puzzle in the final round of the endline experiment, as described in Section VI. The results in Table 6 suggest that students opted for harder tasks even without rewards. The effect is significant for the final task 4, which is the incentivized one. However, the difference in the coefficients between the results for the incentivized and non-incentivized results is not statistically significant.

Finally, we study heterogeneity in the treatment effect along a number of dimensions. The results are reported in Table 7, both for the midline and endline for the number of hard choices. In the midline, we find that boys and those with lower academic results have somewhat larger treatment effects, although this finding doesn't persist fully in the endline. Surprisingly, we find that the initial growth mindset attitudes of both students and parents are positively associated with larger treatment effects. This finding goes against the view that mindset training can reduce initial gaps in growth mindset orientation. In contrast, the treatment effects in both midline and endline are larger for students who scored higher on growth mindset orientation or who had parents with such beliefs. Similarly, to the results for growth mindset, this suggests that either more intensive interventions might be required, or they might need to complemented with other approaches.

VI.C Test Scores

In Tables 8 and 9, we show the results of the mindset treatment on Chinese and mathematics test scores, respectively. Students are tested in official centrally graded exams in both January and June, after the fall and spring semesters. We report results for both of these in the academic year 2020-21, referring to the January tests as 'midline' (soon after the treatment) and June tests as the 'endline' (over 6 months after the treatment). Both tables contain midline and endline results with and without control variables.

Overall, we find that without controls, the treatment effects for both Chinese and mathematics are positive (0.04-0.06 SD), but not statistically significant. Once controls are added, standard errors of the estimates reduce substantially, but so do the point estimates, leading to effects which are fairly closely centered around zero. This applies to both midline and endline estimates. Finally, appendix Table 3 reports longer-term impact on test scores. The outcomes

employed are test scores from a year later, in grade 6. These results also show that there is no evidence of any impact of the treatment on test scores.

Table 10 shows heterogeneous treatment effects for Chinese and mathematics test scores. Using the same sub-samples as for the other outcomes, we find virtually no persistent treatment effects; the endline treatment effects are statistically insignificant at usual levels for all sub-samples for both Chinese and Mathematics. In the midline we do find that the initial growth mindset values may be related to the treatment effects, but not in quite the same way as with the real effort tasks. We estimate more negative treatment effects for students with high growth mindset beliefs, or parents with low growth mindset beliefs. Overall, with regards to test scores, there is no consistent message emerging regarding the benefits of the growth mind-set treatment.

VI.D Spillovers

The children who live in dormitories spend most of their time together and in continuous interaction with their roommates. It is therefore of interest to investigate potential spillover effects. In Table 11 we investigate spillover effects within dorms for growth mindset, the number of hard choices in the real effort task, and test scores. We do so by adding a variable to the regression model measuring the proportion of students in the dorm that were treated across the 176 dorms in the sample. The randomization to treatment took place at the school level, so the number of treated pupils within dorms will vary. The mean share of dorm peers treated is 0.36, with a SD of 0.32.[9] The sample in the estimation includes only students whose dorm room is known for the Spring term of year 5. Standard errors are clustered at the dorm level.

[9] We could in principle also estimate spillovers across classrooms. However, there are much fewer classrooms (49), and the variance of the treated peers is much smaller than across dorms (Mean = 0.34, SD = 0.06), resulting in imprecise estimates.

The parameter of interest is the coefficient on the share of students in the dorm room who are treated by the program (apart from the person themselves). Across all outcomes of interest, the estimates for the spillovers are insignificant, suggesting limited role for spillover effects.

VII Impact of Trained Teachers: Value Added Models

We now investigate for potential impact on the students of the volunteer teachers who undertook the mindset training. Recall that these teachers normally teach students in one of grades 2 to 4. Their students, on the other hand, received no mindset training via the intervention. However, the trained teachers may have internalized the mindset training and used it to improve their pedagogical practice which may in turn lead to improvements in test scores. We now test this hypothesis.

As there was no randomization at the teacher level, we employ non-experimental methods to undertake this evaluation. In particular, we make use of official registry data on students, linking pre- and post-training test score information in order to implement 'value added models' (Todd and Wolpin, 2003). For this exercise, we have access to test scores, linked to teachers, from June/July 2020 (the 'pre' period) and from December 2020 / January 2021 (the 'post' period).

Of the 30 teachers trained in the growth mindset approach, 18 teach grade 2 students, 6 each teach grade 3 and grade 4 students. We use test score data for all students in these three grades, comparing outcomes for students taught by trained versus non-trained teachers. In order to account for learning dynamics, we implement both lagged test score as well as gain (changes in test scores) value added models (see, for example, Andrabi et al., 2011, and Jacob et al., 2010).

Table 12 provides descriptive statistics for outcomes and characteristics for students of the volunteer teachers who take up the growth mindset training ('trained') and those who do not ('non-trained'). The first two rows demonstrate that the volunteer teachers appear to be positively selected, with differences of 0.15 and 0.05 of one S.D. in prior Chinese and mathematics test scores, respectively. The last two rows of the table also suggest that teachers trained in growth mindset also teach somewhat more advantaged students (with students less likely to be both boarders and 'left behind,' i.e. parents residing and working away from home).

Results from this exercise are reported in Table 13. The first row ('trained') reports the impact of the teacher trained in the growth mindset program on test scores. None of the estimates, whether we use the lagged test score value added model (columns 1 and 2) or the gains version (columns 3 and 4), show any statistically significant impact on for Chinese or mathematics test scores. There is however, one caveat we should note, which is that the standard errors are quite large and therefore we are unable to rule out economically significant effects of the training. Nevertheless, in three out of four cases, point estimates are small or even negative, suggesting limited impact.

VIII Conclusion

In this study, we report on the results of an intensive growth mindset randomized controlled trial, targeting noncognitive skills. The targeted population of students are from a disadvantaged rural county in China, where a substantial fraction of children is left behind or reside in school dormitories because parents have migrated to cities for work, a common situation in rural China.

The treatment we provide consists of a 10-hour course taught over five weeks at the beginning of the school year for the 5th grade of primary school. The treatment includes lectures, videos and hands-on exercises and is taught to the students by trained teachers from the same

school. The training emphasizes personal skill development as opposed to a 'fixed ability' worldview, learning from feedback, and resilience. In total we have a sample of roughly 1600 students who are allocated to either the treatment, free time, or a placebo treatment where the equivalent time is spent on watching a natural history documentary.

We test the response of students to the treatment in three ways. First, we measure their response to questions designed to elicit their growth mindset attitudes and beliefs. Second, we run classroom-based games, where students solve puzzles and make choices regarding harder and easier tasks. Third, we measure the effect of treatment on academic test performance. The latter two outcomes are measured in the middle of the school year (January) and at the end (May/June). In addition, since our randomization is at the student level, we are able to test for spillover effects within dormitories, where peer interaction may be especially intense. Such spillovers have not typically been evaluated in the existing literature.

The results show marked changes in growth mindset attitudes among students. In classroom experiments, we find that the treated students are more likely to opt for harder tasks, even when there are no incentives to take them up. We find that there is heterogeneity in the treatment effect for the growth mindset outcome, and to some extent, for the real-effort task. The treatment did not reduce student gaps in outcomes by family background, but may even have increased them. The treatment seemed more effective for students who initially had higher growth mindset scores or whose parents had higher growth mindset scores. These results suggest that if the goal is to reduce gaps by family background, initial attitudes or performances, more targeted approaches may be required. Furthermore, we find no evidence of spillover effects for either growth mindset or for the real-effort task within dorms.

With respect to academic performance, we find that there is no significant impact of training students in the mindset approach for either Chinese or mathematics test scores. In general, the students displayed a growth mindset orientation to begin with, with a share scoring at

the high end of the scale, which might possibly dampen the overall impact. Furthermore, we exploit the fact that the teachers trained to implement the mindset intervention for the grade 5 students usually teach younger cohorts of students. We assess whether training these teachers has any impact on the test scores of their regular students, in grades 2 to 4. The question of interest is whether trained teachers 'internalize' the mindset approach, leading to an impact on test scores. We use a non-experimental approach for this part of the analysis, making use of administrative matched student-teacher data to implement test score value added models. Once again, we find no evidence of any impact on academic performance.

Overall, the experiment shows that a five-week mindset intervention has substantial effects on both attitudes and beliefs, as well as behavior on a real effort task. Arguably, the null effects for test scores are not surprising given that the intervention did not focus on specific skills that are examined by standardized tests, but rather, targeted attitudes to learning, which may manifest themselves in other ways. It is also possible that the practical relevance for learning measured with test scores requires a longer time horizon. Our findings that a relatively intensive noncognitive treatment in rural China can change students' attitudes and beliefs contributes to the growing evidence base demonstrating the malleability of noncognitive skills (Kautz et al., 2014). Whether such changes translate into improved choices and later outcomes are important questions which are left for future research.

References

Alan, Sule, Teodora Boneva and Seda Ertac (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics*, 134(3): 1121–1162.

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1-181). Elsevier.

Andersen, Simon Calmar and Helena Skyt Nielsen (2016). Reading intervention with a growth mindset approach improves children's skills. *PNAS* October 25, 2016 113 (43) 12111-12113

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, 3(3): 29-54.

Attanasio, O., Blundell, R., Conti, G., & Mason, G. (2020). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 191, 104171.

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F. and Yeager, D. (2018). Increasing Perseverance in Math: Evidence from a Field Experiment in Norway. *Journal of Economic Behavior and Organization*, 146: 1–15.

Borghans, L., Duckworth, A. L., Heckman, J. J. and ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43 (4): 972–1059.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, 39(4): 1137-1176.

Brock, A. and Hundley, H. (2016). The Growth Mindset Coach. A Teacher's Month-by-Month Handbook for Empowering Students to Achieve. Ulysses Press, Berkeley.

Burgoyne, A.P., Hambrick, D.Z., Macnamara, B.N. (2020). How firm are the foundations of mind-set theory? The claims appear stronger than the evidence. *Psychological Science*, 31 (3) (2020), pp. 258-267

China Women's Federation (2013). Report on rural left-behind children and urban and rural migrant children in China. Beijing, 2013 (In Chinese).

Chen, X., Huang, Q., Rozelle, S., Shi, Y., & Zhang, L. (2009). Effect of migration on children's educational performance in rural China. *Comparative Economic Studies*, 51(3), 323–343.

Claro, S., Paunesku, D. and Dweck, C.S. (2016). Growth mindset tempers the effects of poverty on academic achievement. PNAS, 113 (31), 8664-8668.

Cunha, Flavio, James J. Heckman, Lance J. Lochner and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics*

of Education, ed. Eric A. Hanushek and Frank Welch, 697-812. Amsterdam: North-Holland.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4): 1593-1640.

Dobronyi, C.R., P. Oreopoulos and U. Petronijevic (2019). Goal Setting, Academic Reminders, and College Success: A Large-Scale Field Experiment. *Journal of Research on Educational Effectiveness*, 12(1): 38-66.

Duckworth, Angela L., and Patrick D. Quinn, "Development and Validation of the Short Grit Scale (Grit-S)," *Journal of Personality Assessment*, 91 (2009): 166–174.

Dweck, Carol S. (2006). *Mindset: The new psychology of success*. Random House, New York (2006)

Glewwe, P., Huang, Q., & Park, A. (2017). Cognitive skills, noncognitive skills, and school-to-work transitions in rural China. *Journal of Economic Behavior & Organization*, 134: 141-164.

Glewwe, P., Song, Y., & Zou, X. (2022). Labor market outcomes, cognitive skills, and noncognitive skills in rural China. *Journal of Economic Behavior & Organization*, 193: 294-311.

Foliano F., H. Rolfe, J. Buzzeo, J. Runge, and D. Wilkinson (2019). Changing Mindsets: Effectiveness Trial. Educational Endowment Foundation. (https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Changing_Mindsets.pdf [Accessed 1.12.2022]).

He, B., Fan, J., Liu, N., Li, H., Wang, Y., Williams, J., & Wong, K. (2012). Depression risk of "left-behind children" in rural China. *Psychiatry Research*, 200(2–3): 306–312.

Heckman, J., Hohmann, N., Smith, J. and Khoo, M. (2000). Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment. *The Quarterly Journal of Economics*, 115(2), 651–694.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3): 411-482.

Huang, Zhen, Xiangdong Wei, Runhao Lu and Jiannong Shi (2022). Whether and how can a growth mindset intervention help students in a non-western culture? Evidence from a field experiment in China. Educational Psychology, 42(7): 913-929.

Humphries, John Eric, and Fabian Kosse (2017). On the interpretation of noncognitive skills - What is being measured and why it matters. *Journal of Economic Behavior & Organization*, vol. 136, April 2017, 174-185.

Jacob, B. A., Lefgren, L., & Sims, D. P. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, 45(4): 915-943.

Kautz T., J.J. Heckman, R. Diris, B. Ter Weel, L. Borghans (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. NBER Working Paper 20749.

Kline, P. and C. R. Walters (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start. *The Quarterly Journal of Economics*, 131(4): 1795-1848.

Lee, L., and A. Park (2010) "Parental migration and child development in rural China", Working Paper, Gansu Survey of Children and Families, WP 24.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 3(1): 101-28.

Liu, M., & Villa, K. M. (2020). Solution or isolation: Is boarding school a good solution for left-behind children in rural China? *China Economic Review*, 61, 101456.

Midgley, C., Maehr, M.L., Hruda, L.Z., Anderman, E., Anderman, L., Freeman, K.E, Gheen, M., Kaplan, A., Kumar, R., Middleton, M.J., Nelson, J., Roeser, R. and Urdan, T. (2000). Manual for the Patterns of Adaptive Learning Scales. University of Michigan.

OECD (2021). Sky's the Limit: Growth mindset, students, and schools in PISA. OECD Directorate for Education and Skills.

Oreopoulos, P. and Petronijevic, U. (2019) The Remarkable Unresponsiveness of College Students to Nudging and What We Can Learn from it, IZA Discussion Paper No. 12460. Park, D., Tsukayama, E., Yu, A., & Duckworth, A. L. (2020). The development of grit and growth mindset during adolescence. *Journal of Experimental Child Psychology*, 198, 104889.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S. and Dweck, C. S. (2015). Mind-set Interventions Are a Scalable Treatment for Academic Under-achievement. *Psychological Science*, 26(6): 784–793.

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L. and Macnamara, B. N. (2018). To What Extent and Under Which Circumstances Are Growth Mind-Sets Important to Academic Achievement? Two Meta-Analyses. *Psychological Science*, 29(4): 549–571.

Smithers, L.G., Sawyer, A.C.P., Chittleborough, C.R. *et al.* A systematic review and meta-analysis of effects of early life noncognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat Hum Behav* **2,** 867–880 (2018).

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485), F3-F33.

Wang, S., & Mao, Y. (2018). The effect of boarding on campus on left-behind children's sense of school belonging and academic achievement: Chinese evidence from propensity score matching analysis. *Asia Pacific Journal of Education*, 38(3): 378-393.

World Bank. 2018. World Development Report 2018: Learning to Realize Education's Promise. Washington, DC: World Bank. doi:10.1596/978-1-4648-1096-1.

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269–1284.

Zhao, Y., Niu, G., Hou, H., Zeng, G., Xu, L., Peng, K., et al. (2018). From Growth Mindset to Grit in Chinese Schools: The Mediating Roles of Learning Motivations. *Front. Psychol.* 9:2007. doi: 10.3389/fpsyg.2018.02007

Zhang, L., Yi, H., Luo, R., Liu, C., and Rozelle, S. (2013). The Human Capital Roots of the Middle Income Trap: The Case of China. *Agricultural Economics*, 44, 151–62.

Figure 1 Outline of the Trial

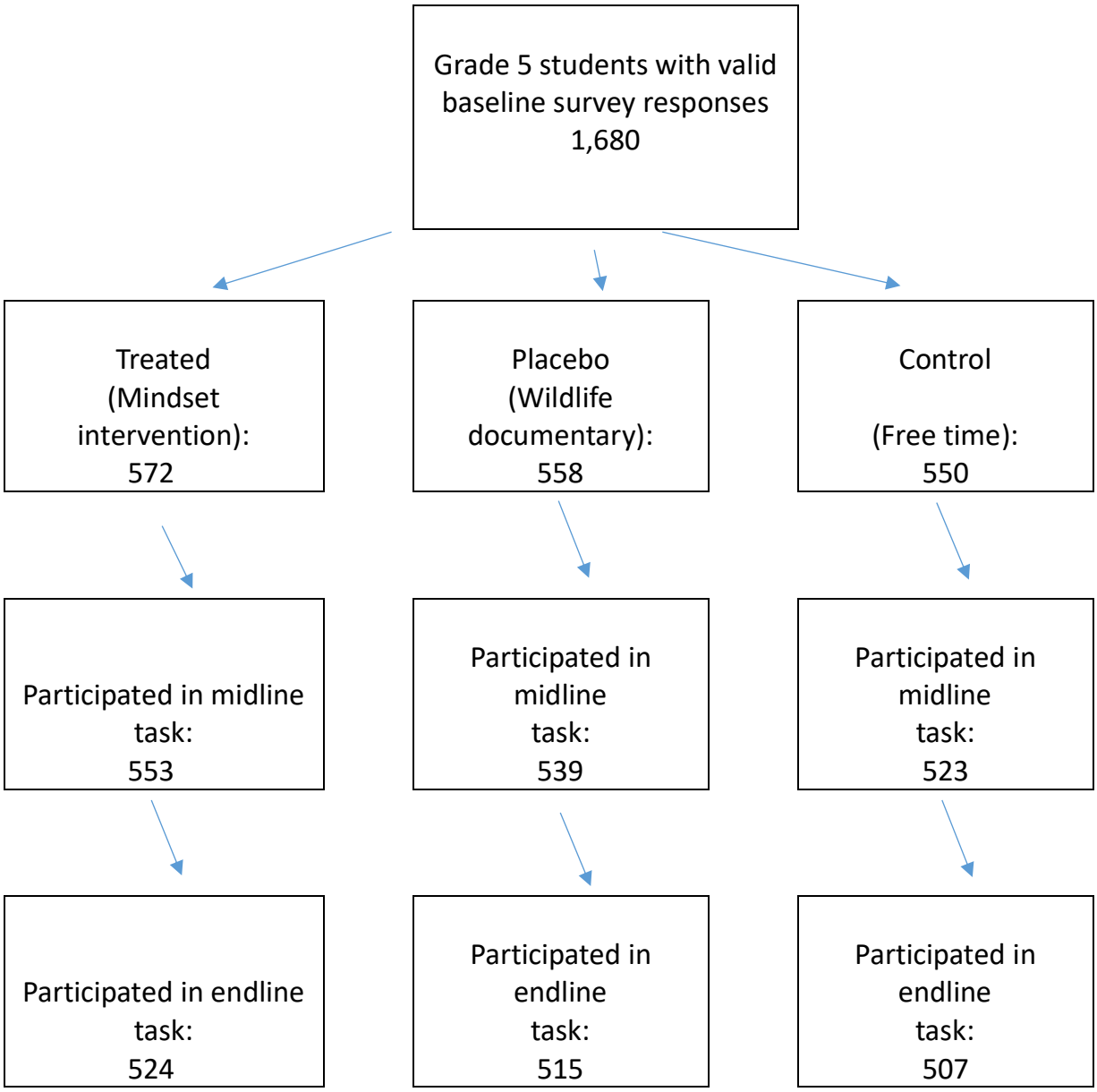

Figure 2 Timeline

Survey dates, teacher training, intervention

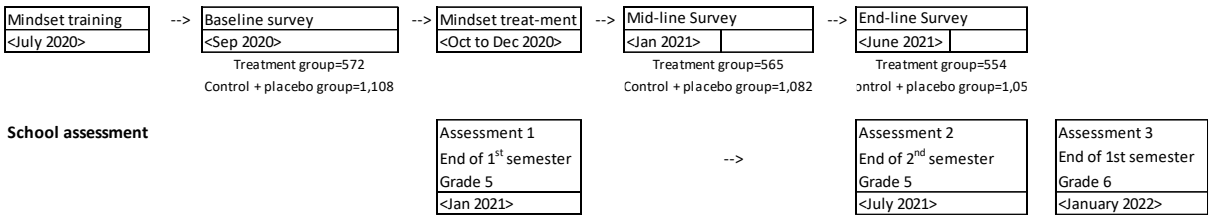

Figure 3 Distributions of pupil baseline orientations, raw scores

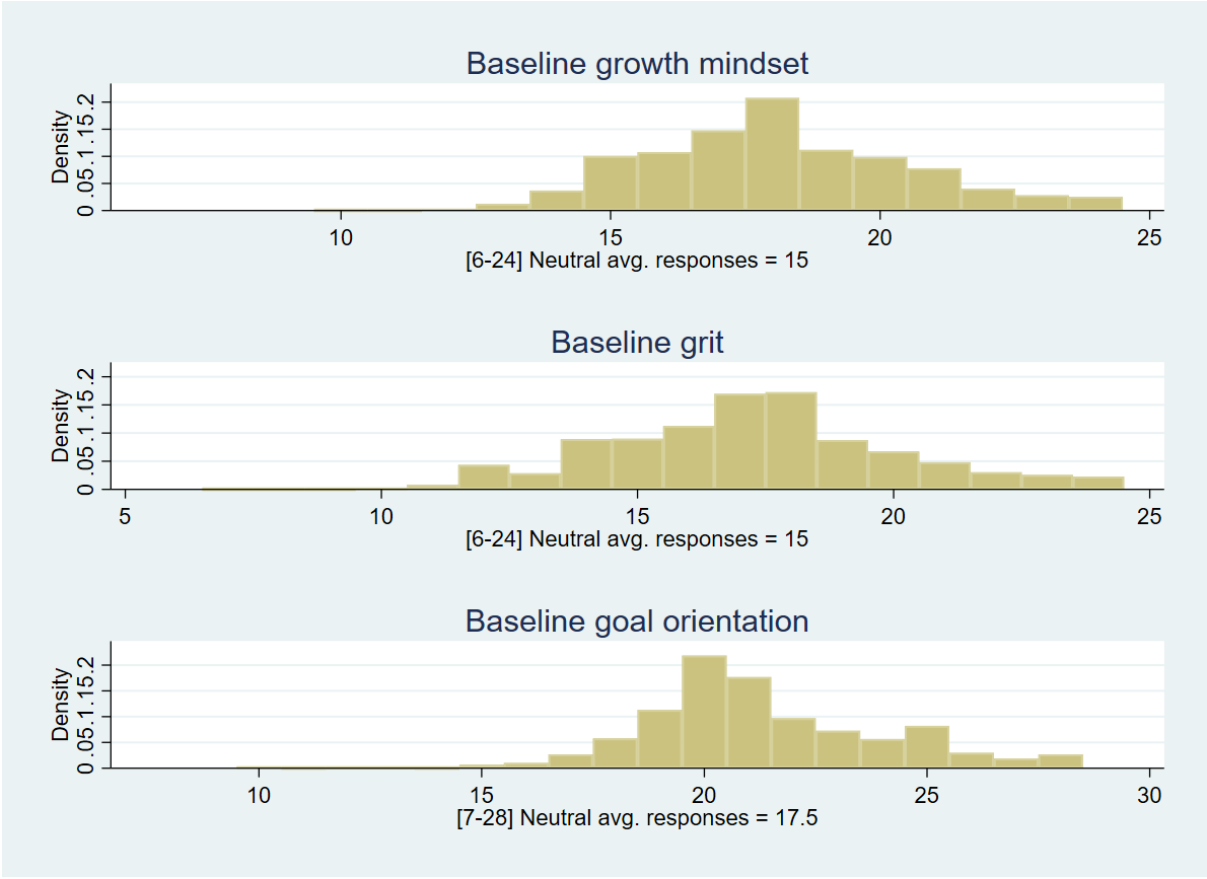

Table 1 Summary statistics and balance test

	(1) Control mean	(2) Treatment-Control	(3) P-value	(4) Obs.
Panel A - School registry data				
Male	0.531	0.006	0.823	1680
	[0.499]	(0.026)		
Low income	0.311	-0.005	0.828	1645
	[0.463]	(0.021)		
Boarding	0.470	0.009	0.663	1645
	[0.499]	(0.020)		
Left behind (parents work away)	0.208	0.007	0.729	1627
	[0.406]	(0.020)		
Chinese (grade 4)	61.983	1.778*	0.074	1680
	[20.013]	(0.942)		
Mathematics (grade 4)	63.950	1.120	0.333	1680
	[20.800]	(1.129)		
Chinese & Mathematics (grade 4)	125.933	2.899	0.131	1680
	[37.092]	(1.839)		
Panel B - Student survey data				
Growth mindset (normalized)	-0.005	0.015	0.839	1680
	[1.010]	(0.074)		
Grit (normalized)	-0.011	0.034	0.377	1680
	[0.995]	(0.038)		
Goal orientation (normalized)	-0.012	0.032	0.387	1680
	[1.023]	(0.036)		
Aiming for college	3.941	0.048	0.553	1680
	[1.248]	(0.080)		
Time spent on homework	2.518	0.035	0.437	1680
	[0.879]	(0.044)		
Mother helps with homework	0.289	0.050*	0.032	1680
	[0.453]	(0.022)		
Father helps with homework	0.195	-0.022	0.205	1680
	[0.396]	(0.016)		
Nobody helps with homework	0.261	-0.027	0.298	1680
	[0.440]	(0.026)		
Extra class for school courses	0.126	-0.004	0.657	1680
	[0.332]	(0.008)		
Extra class for arts	0.139	0.015	0.417	1680
	[0.346]	(0.018)		
Member of school cadre	0.248	0.004	0.831	1680
	[0.432]	(0.017)		
Mother as main carer	0.618	0.032	0.282	1680
	[0.486]	(0.029)		
Father as main carer	0.586	0.021	0.517	1680

	[0.4923]	(0.032)		
Grandparents as main carer	0.498	-0.006	0.734	1680
	[0.500]	(0.016)		
Mother works away	0.074	-0.0004	0.977	1680
	[0.262]	(0.013)		
Father work away	0.211	0.00003	0.999	1680
	[0.408]	(0.027)		
Parents work away	0.282	0.011	0.580	1680
	[0.450]	(0.020)		
Parents check I study	2.634	0.002	0.946	1680
	[1.089]	(0.056)		
Discuss school with mother	2.161	0.059*	0.083	1680
	[0.653]	(0.032)		
Discuss friends with mother	2.061	0.087**	0.01	1680
	[0.689]	(0.030)		
Discuss worries with mother	1.963	0.087***	0.005	1680
	[0.673]	(0.027)		
Discuss school with father	2.013	0.073*	0.097	1680
	[0.673]	(0.042)		
Discuss friends with father	1.954	0.045	0.292	1680
	[0.714]	(0.042)		
Discuss worries with father	1.921	0.041	0.151	1680
	[0.657]	(0.027)		
Only child	0.133	0.018	0.415	1680
	[0.339]	(0.021)		

Notes: *** p<0.01, ** p<0.05, * p<0.1. Placebo and control groups are pooled together. Standard deviations reported in square brackets and t-test standard errors reported in parentheses. Definitions of variables are in the Appendix Table 1.

Table 2 Impact of treatment on noncognitive skills

	Mindset		Grit		Goal Orientation	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.1764***	0.1299**	-0.0115	-0.0353	-0.0578	-0.0741
	(0.0609)	(0.0557)	(0.0617)	(0.0586)	(0.0616)	(0.0604)
Placebo	-0.0075	-0.0277	0.0135	0.0129	0.0254	0.0326
	(0.0611)	(0.0558)	(0.0621)	(0.0588)	(0.0618)	(0.0606)
Male		-0.0171		-0.0224		-0.0781
		(0.0507)		(0.0536)		(0.0551)
Chinese grade 4		0.3335***		0.2307***		0.1388***
		(0.0379)		(0.0400)		(0.0414)
Maths grade 4		0.1863***		0.1595***		0.0339
		(0.0355)		(0.0375)		(0.0386)
English grade 4		-0.0325		0.0083		0.0678*
		(0.0350)		(0.0368)		(0.0381)
Low income		-0.0180		-0.0588		-0.0197
		(0.0527)		(0.0553)		(0.0572)
Boarder		-0.1210*		0.0056		-0.0019
		(0.0633)		(0.0665)		(0.0687)
Left behind		-0.0012		0.0077		0.1414**
		(0.0613)		(0.0646)		(0.0668)
Observations	1,503	1,503	1,503	1,503	1,498	1,498
R-squared	0.0871	0.2451	0.0593	0.1637	0.0690	0.1139

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All three outcome variables are based on normalized additive scores of questions batteries for each of the orientations. Details of the methodology are in the Online Appendix. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 3 Heterogeneous impact on growth mindset

		s.e.	Obs
Girls	0.1399*	(0.0778)	711
Boys	0.1231	(0.0810)	792
Low y4 result	0.0634	(0.0856)	734
High y4 result	0.1782**	(0.0729)	769
Low base GM	0.0596	(0.0898)	614
High base GM	0.1703**	(0.0708)	889
Low parent GM	0.0356	(0.0993)	493
High parent GM	0.1519**	(0.0738)	800
No boarding	0.2065***	(0.0755)	806
Boarding	0.0344	(0.0829)	697
Not low income	0.1006	(0.0664)	1,034
Low income	0.1884*	(0.1045)	469
Not left behind	0.1619***	(0.0614)	1,181
Left behind	-0.0444	(0.1334)	322

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All models estimated against treatment dummy, placebo dummy, grade 4 test scores and school fixed effects. Only estimates of treatment effects are reported. Samples are split by the median value of the sample, apart for Boarding, Low income and Left behind dummies. Year 4 result is a sum of Chinese and Mathematics scores. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 4 Impact on the real effort task – number of 'hard' choices

	Midline (games 2-3)		Endline (games 2-4)	
	(1)	(2)	(3)	(4)
Treatment	0.1008**	0.0863**	0.1060*	0.1006*
	(0.0412)	(0.0400)	(0.0642)	(0.0600)
Placebo	-0.0058	-0.0112	-0.0805	-0.0687
	(0.0414)	(0.0402)	(0.0644)	(0.0602)
Male		0.1691***		0.2265***
		(0.0363)		(0.0549)
Chinese grade 4		0.0278		-0.0074
		(0.0265)		(0.0404)
Maths grade 4		0.1553***		0.3601***
		(0.0254)		(0.0382)
English grade 4		-0.0132		0.0660*
		(0.0253)		(0.0375)
Low income		-0.0449		-0.0459
		(0.0379)		(0.0568)
Boarder		-0.0352		-0.0763
		(0.0447)		(0.0683)
Left behind		-0.0047		-0.0408
		(0.0439)		(0.0661)
Observations	1,615	1,615	1,528	1,528
R-squared	0.0740	0.1331	0.1018	0.2221
Control mean	0.554	0.554	1.088	1.088
Control S.D.	[0.680]	[0.680]	[1.084]	[1.084]

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Midline game includes two choices of difficulty, and the endline three choices. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 5 Success on the real effort task

	Midline (games 2-3)		Endline (games 2-4)	
	(1)	(2)	(3)	(4)
Treatment	-0.0618	-0.0740*	-0.0694	-0.0806
	(0.0447)	(0.0441)	(0.0613)	(0.0579)
Placebo	0.0372	0.0369	0.0067	0.0138
	(0.0450)	(0.0443)	(0.0616)	(0.0580)
Male		-0.0276		0.1039**
		(0.0400)		(0.0529)
Chinese grade 4		0.0234		0.0897**
		(0.0291)		(0.0389)
Maths grade 4		0.0906***		0.3134***
		(0.0279)		(0.0369)
English grade 4		0.0523*		-0.0054
		(0.0279)		(0.0362)
Low income		-0.0656		-0.0195
		(0.0417)		(0.0547)
Boarder		0.0217		-0.0710
		(0.0493)		(0.0657)
Left behind		-0.0355		0.0227
		(0.0483)		(0.0637)
Observations	1,615	1,615	1,529	1,529
R-squared	0.0699	0.1049	0.1052	0.2117
Control mean	0.987	0.987	1.236	1.236
Control S.D.	[0.757]	[0.757]	[1.028]	[1.028]

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Midline game includes two choices of difficulty, and the endline three choices. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 6 Impact on hard choices, with and without incentives

	Endline task 2 Not incentivised		Endline task 3 Not incentivised		Endline task 4 Incentivised	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0186	0.0170	0.0419	0.0377	0.0509*	0.0482*
	(0.0296)	(0.0281)	(0.0287)	(0.0280)	(0.0293)	(0.0286)
Placebo	-0.0488	-0.0430	-0.0058	-0.0038	-0.0228	-0.0202
	(0.0297)	(0.0282)	(0.0289)	(0.0281)	(0.0294)	(0.0287)
Male		0.0833***		0.0454*		0.0956***
		(0.0257)		(0.0256)		(0.0261)
Chinese grade 4		-0.0243		0.0018		0.0156
		(0.0189)		(0.0189)		(0.0192)
Maths grade 4		0.1661***		0.1027***		0.0907***
		(0.0179)		(0.0179)		(0.0182)
English grade 4		0.0208		0.0195		0.0241
		(0.0176)		(0.0176)		(0.0179)
Low income		-0.0001		-0.0369		-0.0089
		(0.0266)		(0.0265)		(0.0270)
Boarder		0.0038		-0.0464		-0.0242
		(0.0320)		(0.0319)		(0.0325)
Left behind		-0.0315		-0.0256		0.0177
		(0.0309)		(0.0308)		(0.0315)
Observations	1,543	1,543	1,539	1,539	1,534	1,534
R-squared	0.0454	0.1470	0.0578	0.1132	0.1028	0.1541

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Endline includes four games. In the first game the pupils have no choice, as the difficulty is randomised. The fourth game in the endline is incentivised. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 7 Heterogeneous impact on hard tasks

	Midline # Hard tasks (0-2)			Endline # Hard tasks (0-3)		
		s.e	Obs		s.e	Obs
Girls	0.0635	(0.0574)	760	0.1107	(0.0884)	717
Boys	0.0964*	(0.0568)	855	0.1073	(0.0837)	811
Low y4 result	0.1338**	(0.0531)	807	0.0582	(0.0809)	752
High y4 result	0.0555	(0.0613)	808	0.1301	(0.0876)	776
Low base GM	-0.0018	(0.0627)	657	-0.0122	(0.0933)	624
High base GM	0.1437***	(0.0529)	958	0.1935**	(0.0798)	904
Low parent GM	0.0356	(0.0697)	539	-0.1439	(0.1028)	501
High parent GM	0.1007*	(0.0566)	854	0.2848***	(0.0848)	811
No boarding	0.1092*	(0.0577)	856	0.0664	(0.0847)	816
Boarding	0.0651	(0.0570)	759	0.1033	(0.0866)	712
Not low income	0.0799	(0.0498)	1,116	0.0972	(0.0741)	1,047
Low income	0.0953	(0.0691)	499	0.1067	(0.1060)	481
Not left behind	0.1121**	(0.0462)	1,273	0.0736	(0.0691)	1,206
Left behind	0.0160	(0.0853)	342	0.1712	(0.1259)	322

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All models estimated against treatment dummy, placebo dummy, grade 4 test scores and school fixed effects. Only estimates of treatment effects are reported. Samples are split by the median value of the sample, apart for Boarding, Low income and Left behind dummies. Year 4 result is a sum of Chinese and Mathematics scores.

Table 8 Impact on test scores: Chinese

	Midline		Endline	
	(1)	(2)	(3)	(4)
Treatment	0.0489	-0.0369	0.0647	-0.0080
	(0.0564)	(0.0280)	(0.0574)	(0.0339)
Placebo	-0.0181	-0.0416	0.0012	-0.0144
	(0.0569)	(0.0282)	(0.0579)	(0.0341)
Male		-0.1534***		-0.0721**
		(0.0255)		(0.0309)
Chinese grade 4		0.7405***		0.6812***
		(0.0185)		(0.0224)
Maths grade 4		0.1138***		0.1407***
		(0.0178)		(0.0216)
English grade 4		0.0594***		0.0579***
		(0.0177)		(0.0213)
Low income		0.0107		0.0889***
		(0.0265)		(0.0320)
Boarder		0.0253		0.0502
		(0.0314)		(0.0382)
Left behind		0.0695**		0.0447
		(0.0309)		(0.0373)
Observations	1,647	1,647	1,611	1,611
R-squared	0.1353	0.7893	0.1265	0.6987

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Table 9 Impact on test scores: Mathematics

	Midline		Endline	
	(1)	(2)	(3)	(4)
Treatment	0.0325	-0.0142	0.0553	0.0118
	(0.0518)	(0.0325)	(0.0524)	(0.0330)
Placebo	-0.0144	-0.0123	0.0266	0.0316
	(0.0523)	(0.0328)	(0.0529)	(0.0332)
Male		0.1281***		0.1003***
		(0.0296)		(0.0300)
Chinese grade 4		0.1109***		0.1543***
		(0.0215)		(0.0219)
Maths grade 4		0.5698***		0.5569***
		(0.0207)		(0.0210)
English grade 4		0.1338***		0.1087***
		(0.0206)		(0.0208)
Low income		-0.0012		-0.0000
		(0.0308)		(0.0312)
Boarder		-0.0006		0.0598
		(0.0365)		(0.0372)
Left behind		-0.0582		-0.0252
		(0.0359)		(0.0363)
Observations	1,647	1,647	1,611	1,611
R-squared	0.2703	0.7156	0.2698	0.7144

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Missing dummies included for 21 missing data points for ‘Left behind’ and 3 data points for ‘Low income’ and ‘Boarder’. All models include school fixed effects.

Table 10 Heterogeneous impact on test scores

Panel A	Midline Chinese			Endline Chinese		
		s.e.	Obs		s.e.	Obs
Girls	-0.0263	(0.0366)	773	0.0004	(0.0418)	760
Boys	-0.0320	(0.0417)	874	0.0006	(0.0519)	851
Low y4 result	-0.0724	(0.0478)	823	-0.0029	(0.0585)	805
High y4 result	-0.0126	(0.0311)	824	-0.0156	(0.0328)	806
Low base GM	0.0241	(0.0503)	672	0.0375	(0.0577)	661
High base GM	-0.0813**	(0.0333)	975	-0.0433	(0.0415)	950
Low parent GM	-0.1035**	(0.0497)	546	0.0116	(0.0586)	533
High parent GM	-0.0039	(0.0382)	870	-0.0094	(0.0453)	850
No boarding	-0.0465	(0.0353)	868	-0.0473	(0.0424)	849
Boarding	-0.0184	(0.0456)	779	0.0365	(0.0547)	762
Not low income	-0.0155	(0.0342)	1,137	-0.0341	(0.0415)	1,105
Low income	-0.0910*	(0.0513)	510	0.0432	(0.0606)	506
Not left behind	-0.0184	(0.0316)	1,305	0.0062	(0.0372)	1,274
Left behind	-0.0900	(0.0650)	342	-0.0614	(0.0843)	337

Panel B	Mathematics Treatment			Mathematics Treatment		
			Obs			Obs
Girls	-0.0514	(0.0472)	773	-0.0612	(0.0482)	760
Boys	0.0087	(0.0454)	874	0.0746	(0.0457)	851
Low y4 result	0.0013	(0.0501)	823	0.0148	(0.0524)	805
High y4 result	-0.0115	(0.0405)	824	0.0193	(0.0404)	806
Low base GM	-0.0037	(0.0541)	672	0.0142	(0.0564)	661
High base GM	-0.0282	(0.0410)	975	0.0010	(0.0403)	950
Low parent GM	-0.0521	(0.0564)	546	-0.0586	(0.0574)	533
High parent GM	0.0299	(0.0455)	870	0.0551	(0.0444)	850
No boarding	-0.0273	(0.0424)	868	-0.0034	(0.0415)	849
Boarding	0.0263	(0.0508)	779	0.0386	(0.0530)	762
Not low income	-0.0031	(0.0394)	1,137	0.0169	(0.0399)	1,105
Low income	-0.0158	(0.0598)	510	0.0254	(0.0604)	506
Not left behind	-0.0076	(0.0364)	1,305	0.0259	(0.0365)	1,274
Left behind	-0.0161	(0.0752)	342	-0.0274	(0.0785)	337

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All models estimated against treatment dummy, placebo dummy, grade 4 test scores and school fixed effects. Only estimates of treatment effects are reported. Samples are split by the median value of the sample, apart for Boarding, Low income and Left behind dummies. Year 4 result is a sum of Chinese and Mathematics scores.

Table 11 Dormitory spillover effects in growth mindset, effort task, Chinese and Mathematics

	Stated GM	Hard choices	Chinese		Maths		
	Endline	Midline	Endline	Midline	Endline	Midline	Endlin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.0804	0.075	0.0575	-0.0233	0.0238	0.0358	0.023
	(0.0921)	(0.0693)	(0.1045)	(0.0485)	(0.0613)	(0.0595)	(0.062
Placebo	0.007	-0.0183	-0.0481	-0.0278	-0.0504	0.0841	0.1035
	(0.0888)	(0.0715)	(0.0988)	(0.0513)	(0.0635)	(0.064)	(0.060
Proportion treated in dormitory	-0.0678	0.096	-0.0959	0.1026	0.0369	0.0678	-0.066
	(0.1175)	(0.0771)	(0.1313)	(0.0653)	(0.0747)	(0.0787)	(0.075
Male	-0.1501*	0.1163**	0.2286***	-0.2049***	-0.1738***	0.1228**	0.1226
	(0.0842)	(0.0588)	(0.0866)	(0.0517)	(0.0649)	(0.0573)	(0.057
Chinese grade 4	0.3753***	-0.0022	-0.0038	0.7245***	0.6514***	0.0996**	0.1056*
	(0.0643)	(0.0409)	(0.0624)	(0.035)	(0.0399)	(0.0461)	(0.039
Maths grade 4	0.1697***	0.1602***	0.3383***	0.0760**	0.1204***	0.5761***	0.5433*
	(0.0611)	(0.0373)	(0.0555)	(0.0312)	(0.0364)	(0.0372)	(0.036
English grade 4	-0.0289	-0.0577	0.1133*	0.0780**	0.0543	0.1144**	0.1526*
	(0.0557)	(0.0414)	(0.0671)	(0.0345)	(0.0397)	(0.0459)	(0.041
Low income	0.0504	0.0328	0.0103	0.0208	0.0967*	0.0198	-0.025
	(0.0716)	(0.0561)	(0.0857)	(0.0447)	(0.0525)	(0.0494)	(0.046
Left behind	0.0031	-0.0701	-0.0963	0.0609	0.0419	-0.1508**	-0.06
	(0.0939)	(0.0593)	(0.0808)	(0.0447)	(0.0587)	(0.058)	(0.058
Constant	-0.4478	0.2202	0.6944***	-0.0456	-0.4330*	0.4391***	-0.2363
	(0.2742)	(0.1377)	(0.2433)	(0.1035)	(0.2258)	(0.0906)	(0.140
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	543	584	554	595	595	595	595
R-squared	0.2811	0.1801	0.2264	0.7638	0.6451	0.651	0.677

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample restricted to pupils whose dormitory status is known in Spring term. Midline effort task includes two choices of difficulty, and the endline three choices. All models include school fixed effects.

Table 12 Student characteristics for trained and non-trained teachers

	（1）	（2）	(3)	(4)
	Non-trained	Trained teachers	Treatment-control	P-value
Chinese last semester	-0.037	0.112	0.149	0.000***
	[1.006]	[0.972]	(0.031)	
Maths last semester	-0.013	0.040	0.052	0.093*
	[1.010]	[0.969]	(0.031)	
Low income	0.287	0.306	-0.019	0.181
	[0.453]	[0.461]	(0.014)	
Boarder	0.401	0.467	-0.066	0.000***
	[0.490]	[0.499]	(0.015)	
Left behind	0.202	0.292	-0.091	0.000***
	[0.401]	[0.455]	(0.013)	

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample size: 5,537 for each row. Standard deviations reported in square brackets and standard errors reported in parentheses

Table 13 Impact of trained teachers, Value Added Models

	(1) Chinese	(2) Maths	(3) ΔChinese	(4) ΔMaths
Trained	0.1038	-0.0178	0.0209	-0.0612
	(0.0958)	(0.0539)	(0.1175)	(0.0616)
Male	-0.2433***	0.0093	-0.0732***	0.0119
	(0.0233)	(0.0181)	(0.0179)	(0.0175)
Chinese last semester	0.4397***	0.0081		
	(0.0221)	(0.0316)		
Maths last semester	0.2623***	0.6934***		
	(0.0289)	(0.0348)		
Low income	-0.0566**	-0.0608**	-0.0128	0.0042
	(0.0209)	(0.0232)	(0.0263)	(0.0213)
Boarder	0.0092	0.0036	0.0584	0.0577
	(0.0360)	(0.0339)	(0.0370)	(0.0363)
Left behind	0.0015	-0.0156	0.0077	0.0061
	(0.0330)	(0.0394)	(0.0376)	(0.0377)
Observations	5,537	5,537	5,537	5,537
R-squared	0.5743	0.6107	0.0347	0.0247

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. ΔChinese = Chinese - Chinese last semester, ΔMaths = Maths - Maths last semester. 30 Chinese teachers were trained to deliver the intervention. 18 of selected teachers taught Grade 2, 6 taught Grade 3 and the rest 6 taught Grade 4. We get hold of the data linking them to their grade 2-4 students and the students' test scores from the education bureau. We study the effects of teachers being trained on academic test scores over one semester (the Spring term in the academic year 2019-2020 versus the Autumn term in the academic year 2020-2021). We show results for Chinese and Mathematics, including school fixed effects and clustering standard errors at the school-level.

Appendix Table1 Definition of variables

Variable	Definition
Male	Male = 1, Female = 0
Low-income	Officially certified low-income family: Yes=1; No=0
Boarding	Boarding at school: Yes=1; No=0
Leftover	At least either parent migrant worker: yes=1;no=0
Chinese grade 4	Chinese test scores in the second term of Grade 4
Maths grade 4	Maths test scores in the second term of Grade 4
Chinese & Maths grade 4	Chinese and Maths test scores in the second term of Grade 4
Only child	Only child in the family: Yes=1; No=0
Extra class for school courses	Participate in any extra class for Chinese, English and math: Yes=1; No=0
Extra class for arts	Participate in any extra class for music, dancing and painting: Yes=1; No=0
Cadre	Role as a Monitor, Vice monitor, Course representative or a Group leader of a class: Yes=1; No=0
Time on homework	Avg. time spent on homework: 1= < 0.5h ; 2 = 0.5-1h ; 3 = 1-1.5h ; 4 = >1.5h
Aim for college	Preference for attending college in the future: 1=No idea; 2=Not at all; 3=Not really; 4=Much; 5=Very much
Mother as main carer	Live with mother at home: Yes=1; No=0
Father as main carer	Live with father at home: Yes=1; No=0
Grandparents as main carer	Live with grandparents at home: yes=1;no=0
Mother Work away	Only mother works away from home: Yes=1; No=0
Father Work away	Only father work away from home: Yes=1; No=0
Parents Work away	Both mother and father work away from home: Yes=1; No=0
Nobody tutor homework	Nobody helps with homework: Yes=1; No=0
Mother tutor homework	Mother usually helps with homework: Yes=1; No=0
Father tutor homework	Father usually helps with homework: Yes=1; No=0
Study check	How often in a week do parents check homework: Did not check=1; 1-2 days = 2; 3-4 days = 3; almost every day = 4
Discuss school with mother	Discuss things happened at school with mother: Never=1;sometimes=2;often=3
Discuss friends with mother	Discuss relationship with friends with mother: Never=1;sometimes=2;often=3
Discuss worries with mother	Whether discuss worries and troubles with mother: Never=1;sometimes=2;often=3
Discuss school with father	Discuss things happened at school with father: Never=1;sometimes=2;often=3
Discuss friends with father	Discuss relationship with friends with father: Never=1;sometimes=2;often=3
Discuss worries with father	Discuss worries and troubles with father: Never=1;sometimes=2;often=3
Growth mindset	Normalised self-rated Growth mindset in baseline
Grit	Normalised self-rated Grit in baseline
Goal orientation	Normalised self-rated Goal orientation in baseline

Appendix Table 2 Correlates of noncognitive skills

	[1]	[2]	[3]	[4]	[5]	[6]
	Mindset		Grit		Goal Orientation	
Male	-0.1440***	-0.0743	-0.1392***	-0.024	-0.1075**	-0.0499
	(0.0484)	(0.0521)	(0.0486)	(0.0524)	(0.0492)	(0.0578)
Low income	-0.2159***	-0.1404***	-0.0901*	0.0371	-0.0097	0.0652
	(0.0538)	(0.0543)	(0.054)	(0.0546)	(0.0546)	(0.0602)
Boarder	-0.2191***	-0.0736	-0.1764***	-0.0086	-0.1716***	-0.0913
	(0.0505)	(0.065)	(0.0508)	(0.0654)	(0.0513)	(0.072)
Left behind	-0.2031***	-0.0938	-0.2864***	-0.1153*	-0.0651	-0.0315
	(0.0605)	(0.0657)	(0.0608)	(0.0662)	(0.0614)	(0.0729)
Nobody helps		-0.0545		-0.2762***		-0.0133
with homework		(0.0676)		(0.0681)		(0.075)
Mother or father helps		-0.0452		0.0086		-0.0559
with homework		(0.0595)		(0.0599)		(0.066)
Chinese, grade 4		0.1231***		0.1823***		0.0908**
		(0.0392)		(0.0394)		(0.0434)
Maths, grade 4		0.1177***		0.1412***		0.1028**
		(0.0363)		(0.0365)		(0.0403)
English, grade 4		0.0195		0.0334		0.0207
		(0.0355)		(0.0358)		(0.0394)
Parental growth mindset		0.2279***		0.1389***		0.1420***
		(0.0237)		(0.0238)		(0.0263)
Constant	0.2890***	-0.6825***	0.2447***	-0.7667***	0.1547***	0.2223
	(0.0447)	(0.2228)	(0.0449)	(0.2242)	(0.0454)	(0.2471)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,647	1,416	1,647	1,416	1,647	1,416
R-squared	0.0421	0.2322	0.0332	0.212	0.0117	0.1038

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. All three outcome variables are based on normalized additive scores of questions batteries for each of the orientations. Details of the methodology are in the Online Appendix. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Appendix Table 3 Longer-term impact on test scores (Grade 6)

	Chinese		Maths	
	(1)	(2)	(3)	(4)
Treatment	0.0201	-0.0579	0.0851	0.0304
	(0.0525)	(0.0376)	(0.0546)	(0.0411)
Placebo	0.0408	0.0279	0.0152	0.0218
	(0.0486)	(0.037)	(0.0445)	(0.0381)
Male		-0.0905**		0.0323
		(0.0407)		(0.041)
Chinese grade 4		0.6501***		0.1753***
		(0.0269)		(0.0275)
Maths grade 4		0.1145***		0.4954***
		(0.0311)		(0.0286)
English grade 4		0.0693***		0.1135***
		(0.0218)		(0.0233)
Low income		-0.0041		-0.0643
		(0.033)		(0.0435)
Boarder		0.0854*		0.1325***
		(0.0431)		(0.04)
Left behind		0.0394		-0.0517
		(0.0352)		(0.0651)
Observations	1,554	1,554	1,555	1,555
R-squared	0.1287	0.6755	0.2188	0.6283

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Missing dummies included for 21 missing data points for 'Left behind' and 3 data points for 'Low income' and 'Boarder'. All models include school fixed effects.

Online Appendix Tables

Online Appendix Table 4 Correlations between the three noncognitive outcomes

	Mindset	Grit
Grit	0.541***	
Goal Orientation	0.330***	0.317***

Notes: ***: $p < 0.01$

Online Appendix Table 5 Further Balance Tests

	(1)	(2)	(3)	(4)	(5)
	Control mean	Treatment - Control		Placebo - Control	
		Mean	P-value	Mean	P-value
Panel A School registry data					
Male	0.522	0.015	0.642	0.017	0.489
	[0.500]	(0.031)		(0.025)	
Low-income	0.317	-0.011	0.607	-0.013	0.465
	[0.466]	(0.022)		(0.018)	
Boarding	0.479	-0.001	0.984	-0.019	0.535
	[0.500]	(0.026)		(0.030)	
Leftover	0.209	0.005	0.776	-0.003	0.852
	[0.407]	(0.019)		(0.018)	
Chinese grade 4	61.760	2.053*	0.059	0.545	0.556
	[20.289]	(1.027)		(0.911)	
Maths grade 4	64.311	0.786	0.564	-0.663	0.418
	[21.026]	(1.341)		(0.801)	
Chinese & Maths grade 4	126.071	2.839	0.186	-0.118	0.939
	[37.587]	(2.074)		(1.534)	
Panel B Student survey data					
Only child	0.145	0.005	0.850	-0.026	0.206
	[0.353]	(0.025)		(0.020)	
Extra class for school courses	0.131	-0.008	0.500	-0.008	0.559
	[0.338]	(0.012)		(0.014)	
Extra class for arts	0.145	0.009	0.680	-0.012	0.640
	[0.353]	(0.021)		(0.026)	
Cadre	0.260	-0.008	0.666	-0.024	0.221
	[0.439]	(0.019)		(0.019)	
Time on homework	2.533	0.022	0.702	-0.027	0.550
	[0.884]	(0.056)		(0.044)	
Aim for collage	3.904	0.087	0.362	0.078	0.123
	[1.279]	(0.093)		(0.048)	
Mother as main carer	0.633	0.018	0.593	-0.027	0.317
	[0.483]	(0.033)		(0.027)	
Father as main carer	0.602	0.005	0.868	-0.031	0.274
	[0.490]	(0.031)		(0.028)	
Grandparents as main carer	0.485	0.007	0.766	0.025	0.395
	[0.500]	(0.023)		(0.028)	
Mother work away	0.080	-0.006	0.679	-0.012	0.495
	[0.272]	(0.015)		(0.017)	
Father work away	0.233	-0.021	0.467	-0.043**	0.050
	[0.423]	(0.029)		(0.020)	
Parents work away	0.255	0.039	0.156	0.054*	0.067
	[0.436]	(0.026)		(0.028)	

Nobody tutors homework	0.264	-0.030	0.320	-0.005	0.769
	[0.441]	(0.029)		(0.017)	
Mother tutors homework	0.278	0.061*	0.051	0.022	0.386
	[0.449]	(0.029)		(0.025)	
Father tutors homework	0.216	-0.043*	0.064	-0.042**	0.024
	[0.412]	(0.022)		(0.017)	
Study check	2.667	-0.035	0.520	-0.063*	0.077
	[1.088]	(0.053)		(0.034)	
Discuss school with mother	2.162	0.059*	0.093	-0.0004	0.991
	[0.651]	(0.033)		(0.033)	
Discuss friends with mother	2.058	0.091**	0.016	0.008	0.786
	[0.688]	(0.034)		(0.031)	
Discuss worries with mother	1.927	0.123***	0.000	0.072*	0.052
	[0.676]	(0.028)		(0.035)	
Discuss school with father	1.995	0.092**	0.043	0.037	0.321
	[0.689]	(0.042)		(0.036)	
Discuss friends with father	1.942	0.058	0.186	0.025	0.564
	[0.714]	(0.042)		(0.042)	
Discuss worries with father	1.898	0.064***	0.006	0.047	0.211
	[0.673]	(0.021)		(0.037)	
Growth mindset	0.010	0.001	0.984	-0.027	0.512
	[1.020]	(0.070)		(0.041)	
Grit	-0.024	0.048	0.236	0.028	0.593
	[0.991]	(0.039)		(0.051)	
Goal orientation	-0.036	0.056	0.181	0.048	0.413
	[1.027]	(0.041)		(0.057)	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard deviations reported in square brackets and standard errors reported in parentheses.

Online Appendix Table 6 Impact on noncognitive skills (parental treatment)

	Mindset		Grit		Goal Orientation	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.1539**	0.1041	0.0140	-0.0092	-0.0548	-0.0709
	(0.0741)	(0.0677)	(0.0751)	(0.0712)	(0.0750)	(0.0735)
Parent Treatment	0.0456	0.0521	-0.0516	-0.0531	-0.0060	-0.0064
	(0.0854)	(0.0779)	(0.0866)	(0.0820)	(0.0865)	(0.0847)
Placebo	-0.0075	-0.0277	0.0135	0.0129	0.0254	0.0326
	(0.0612)	(0.0558)	(0.0621)	(0.0588)	(0.0618)	(0.0606)
Male		-0.0165		-0.0230		-0.0781
		(0.0507)		(0.0536)		(0.0551)
Chinese grade 4		0.3341***		0.2304***		0.1387***
		(0.0380)		(0.0400)		(0.0415)
Maths grade 4		0.1861***		0.1596***		0.0339
		(0.0355)		(0.0375)		(0.0386)
English grade 4		-0.0327		0.0085		0.0678*
		(0.0350)		(0.0368)		(0.0381)
Low income		-0.0179		-0.0588		-0.0197
		(0.0527)		(0.0553)		(0.0572)
Boarder		-0.1205*		0.0050		-0.0020
		(0.0633)		(0.0665)		(0.0688)
Left behind		-0.0014		0.0082		0.1415**
		(0.0613)		(0.0647)		(0.0668)
Observations	1,503	1,503	1,503	1,503	1,498	1,498
R-squared	0.0873	0.2453	0.0595	0.1639	0.0690	0.1139

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Notes: All three outcome variables are based on normalized additive scores of questions batteries for each of the orientations. Details of the methodology are in the Online Appendix. The models now include a dummy variable for parental treatment, which is never statistically significant.

Online Appendix Table 7 Impact on real effort task, number of hard choices, including parental treatmant

	Midline (games 2-3)		Endline (games 2-4)	
	(1)	(2)	(3)	(4)
Treatment	0.0720	0.0556	0.0713	0.0673
	(0.0504)	(0.0490)	(0.0783)	(0.0733)
Parent Treatment	0.0569	0.0606	0.0694	0.0668
	(0.0574)	(0.0557)	(0.0899)	(0.0840)
Placebo	-0.0058	-0.0112	-0.0804	-0.0687
	(0.0414)	(0.0402)	(0.0644)	(0.0602)
Male		0.1697***		0.2272***
		(0.0363)		(0.0549)
Chinese grade 4		0.0287		-0.0066
		(0.0265)		(0.0404)
Maths grade 4		0.1550***		0.3599***
		(0.0254)		(0.0382)
English grade 4		-0.0137		0.0655*
		(0.0253)		(0.0375)
Low income		-0.0445		-0.0456
		(0.0379)		(0.0568)
Boarder		-0.0351		-0.0761
		(0.0447)		(0.0683)
Left behind		-0.0047		-0.0412
		(0.0439)		(0.0661)
Observations	1,615	1,615	1,528	1,528
R-squared	0.0746	0.1337	0.1022	0.2224
Control mean	0.554	0.554	1.088	1.088
Control S.D.	[0.680]	[0.680]	[1.084]	[1.084]

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Midline included three games (or tasks), and endline included four. In the first game the pupils have no choice, as the difficulty is randomised. The fourth game in the endline is incentivised. The models now include a dummy variable for parental treatment, which is never statistically significant

Online Appendix Table 8 The effect of mindset treatment on the ability to solve puzzles (endline only)

	Likelihood of failure in 1st task		
	If 1st task Easy	If 1st task Complex	Full sample
	(1)	(2)	(3)
Treatment	-0.0402	0.0088	-0.0056
	(0.0342)	(0.0284)	(0.0227)
Placebo	0.0382	0.0202	0.0360
	(0.0338)	(0.0289)	(0.0228)
1st Task Complex			0.5773***
			(0.0186)
Male	-0.0073	-0.0162	-0.0137
	(0.0309)	(0.0262)	(0.0207)
Chinese grade 4	-0.0173	-0.0152	-0.0123
	(0.0230)	(0.0192)	(0.0153)
Maths grade 4	-0.2090***	-0.0679***	-0.1356***
	(0.0219)	(0.0180)	(0.0145)
English grade 4	-0.0152	0.0046	-0.0138
	(0.0208)	(0.0186)	(0.0142)
Low income	0.0212	0.0253	0.0239
	(0.0322)	(0.0275)	(0.0215)
Boarder	-0.0132	-0.0215	-0.0158
	(0.0392)	(0.0321)	(0.0258)
Left behind	-0.0045	-0.0093	-0.0172
	(0.0383)	(0.0308)	(0.0250)
Missing	0.4236**	-0.2622	0.1074
	(0.1767)	(0.1772)	(0.1267)
Constant	0.2383	1.0150***	0.4478***
	(0.1609)	(0.0998)	(0.0784)
School FE	Yes	Yes	Yes
Observations	789	757	1,546
R-squared	0.3103	0.0952	0.4673
P-value Treat = Placebo	0.0218**	0.6397	0.0330**

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In the first game the pupils have no choice, as the difficulty is randomised.