Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*

Sule Alan, University of Essex

Teodora Boneva, University College London

Seda Ertac, Koc University

March 2016

Abstract

We show that grit, a non-cognitive skill that has been shown to be highly predictive of achievement, is malleable in the childhood period and can be fostered in the classroom environment. Our evidence comes from the evaluation of a randomized educational intervention implemented in two independent elementary school samples in Istanbul. Outcomes are measured via a novel incentivized real effort task and performance in standardized tests. We find that treated students are 1) more likely to choose to undertake a more challenging and more rewarding task against an easier but less rewarding alternative, 2) less likely to give up after failure, 3) more likely to exert effort to accumulate task-specific ability, and consequently, 4) more likely to succeed and collect higher payoffs. The intervention also has a large positive impact on students' standardized test scores. Treated students score 0.28 standard deviations higher in a standardized math test, and 0.13 standard deviations higher in a standardized Turkish test at follow-up.

JEL Categories: C91, C93, D03, I28

Keywords: non-cognitive skills, grit, perseverance, field experiments, randomized interventions

*This paper is part of a field project partially funded by the Turkish division of the ING Bank. Other funders are TUBITAK (Career Grant 111K444), Koc University, the TUBA-GEBIP Program and the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC) at the University of Essex, grant number ES/L009153/1, whom we would like to thank for generous financial support. Boneva acknowledges financial support from the ESRC and the British Academy. We are grateful to Orazio Attanasio, Marco Castillo, Thomas Crossley, Thomas Dohmen, Angela Duckworth, Armin Falk, James Heckman, John List, Gautam Rao, conference participants at the Experimental Methods in Policy Conference, the Stanford University SITE workshop, the ECBE conference, IZA workshop on Education, Interventions and Experiments and seminar participants at the University of Chicago, Chicago Fed, the World Bank, University of Maryland, George Mason University, University of Michigan, Insitute for Fiscal Studies, and the University of Essex for helpful comments. We would also like to thank Elif Kubilay, Nergis Zaim, Banu Donmez, Mert Gumren and Enes Duysak, as well as numerous other students who provided excellent research assistance. All errors are our own.

"Ever tried. Ever failed. No matter. Try again.

Fail again. Fail better."

— Samuel Beckett, *Worstward Ho*

1 Introduction

The growing literature on human capital accumulation has emphasized the importance of non-cognitive skills in explaining individual differences in achievement in various economic and social domains (Heckman, Stixrud and Urzua, 2006; Borghans et al., 2008). These skills encompass a broad range of individual character traits, often measured via standardized questionnaires by psychologists and, more recently, via incentivized experimental elicitation techniques by economists. Non-cognitive skills such as patience, self-control and conscientiousness, and preference parameters such as attitudes towards risk have been shown to be highly predictive of outcomes ranging from educational attainment, occupational and financial success to criminal activity and health outcomes; see Heckman, Stixrud and Urzua (2006); Almlund et al. (2011); Dohmen et al. (2011); Sutter et al. (2013); Heckman, Humphries and Mader (2011); Moffit et al. (2011); Castillo et al. (2011); Golsteyn, Gronqvist and Lindahl (2013). In fact, the predictive power of non-cognitive skills appears to rival that of cognitive skills (Roberts et al., 2007; Kautz et al., 2014). More importantly from a policy standpoint, there is now ample evidence suggesting that these important skills are malleable especially in the childhood period and can be fostered through educational interventions (Almlund et al., 2011; Kautz et al., 2014).[1]

Among these skills, "grit" - which has not been studied extensively by economists but is likely to influence a myriad of economic decisions and outcomes such as entrepreneurial success, career development, college dropout rates and absenteeism - is the focus of this paper.[2] Grit is generally defined as perseverance in a productive task, and is also related to conscientiousness and being able to set long-term goals. Hence, it is a non-cognitive skill that influences the motivation to set a goal, exert effort towards that goal and persevere in pursuing it in response to negative performance feedback. Given the ubiquity of challenging tasks and frequent performance feedback encountered in educational and employment settings, and the central question of how to motivate individuals to work harder, it is

[1]Well-known examples of early childhood and elementary school programs include the Perry Preschool program (Heckman et al., 2010; Heckman, Pinto and Savelyev, 2013), the Abecedarian Program (Heckman, Moon and Pinto, 2010; Conti et al., 2014), and the project STAR (Schanzenbach, 2006; Dee and West, 2011; Chetty et al., 2011). As an example of a targeted education, Alan and Ertac (2014) show that an educational intervention designed to improve forward-looking behavior in elementary school children leads to favorable outcomes not only in incentivized elicitation tasks but also in disciplinary conduct in school.

[2]Grit has been shown to be associated with college GPAs and educational attainment. It also predicts retention in different contexts: Grittier students are more likely to graduate from high school, grittier employees are more likely to keep their jobs, grittier soldiers are more likely to be retained in the army and grittier men are more likely to remain married. See Duckworth et al. (2007); Duckworth and Quinn (2009); Maddie et al. (2012); Eskreis-Winkler et al. (2014).

important to understand the nature of grit as a non-cognitive skill and to explore ways of enhancing it.

In producing "gritty" behavior, beliefs are likely to play a pivotal role since an individual will set ambitious performance goals and persevere in response to failures if her perceived return to exerting effort is sufficiently high. While beliefs about the existing stock of skill can play an important role in such decisions, beliefs about the role of effort in the production or performance process are also likely to be crucial. Considering longer-term, high-reward targets, these beliefs pertain to the role of effort in general ability development, i.e. to the malleability of ability through effort. If an individual believes that she can develop skills over time by exerting effort (e.g. by continued practice), she will be less discouraged by early failures and more likely to keep at the task, which will lead to higher achievement, especially if the performance technology is conducive to skill accumulation.

In this paper, we provide evidence that grit, an important non-cognitive skill, is malleable in the childhood period and can be fostered through targeted education in the classroom environment. Our evidence comes from the evaluation of a randomized-controlled educational intervention we develop and implement in Istanbul, Turkey. The intervention aims to positively influence children's beliefs about the malleability of ability and the productivity of effort in the skill accumulation process, and thereby induce gritty behavior. The program exposes children to a worldview in which ability, rather than being innately fixed, can be developed through sustained, goal-oriented effort. The core message is to highlight the role of effort in the skill accumulation process and thereby in achievement, and to discourage students from interpreting early setbacks and failures as evidence for a lack of innate ability. The premise is that holding such beliefs about the performance technology will increase the motivation to undertake challenging but rewarding tasks and to exert sustained effort, resulting in higher achievement.

The intervention material involves animated videos, mini case studies and classroom activities that highlight i) the plasticity of the human brain against the notion of innately fixed ability, ii) the role of effort in enhancing skills and achieving goals, iii) the importance of a constructive interpretation of failures, and iv) the importance of goal setting. These materials are shaped by a multidisciplinary team of education consultants and elementary school teachers, and are conveyed by the students' own teachers, who are trained before the program starts. In addition, teachers are encouraged to adopt a teaching philosophy that emphasizes the role of effort in everyday classroom tasks, e.g. while giving performance feedback and interpreting test results. In this sense, the program is not merely a set of materials to be covered in a specified period of time, like a common curriculum item. Instead, it aims to change the students' beliefs about the importance of effort partly by changing the mindset of the teachers and the nature of the classroom environment.[3]

We evaluate the effect of this unique training program using two independent samples of 4th grade students (ages 8-10). Within each sample, the intervention is randomized across schools in which at least one teacher was willing to participate in the program. We measure the outcomes through a multi-faceted methodology that includes a novel incentivized real effort task, test scores and pre- and post-treatment questionnaires. The incentivized real effort task is designed to elicit a number of characteristics of grit; challenge seeking, perseverance through setbacks, goal setting and the propensity to engage in effortful behavior to accumulate skill. Specifically, we elicit students' choices between a challenging high-reward and an easy low-reward task, and the dynamic response of this choice to negative performance feedback. The experiment also involves a temporal component, which allows us to observe longer-term skill accumulation in the challenging task through practice. In addition to experimental choices and outcomes, we administer standardized tests in mathematics and Turkish. We also measure students' beliefs about the malleability of ability and the role of effort in achievement, as well as self-reported attitudes and behaviors regarding perseverance, using pre- and post-treatment questionnaires.

In both samples, our results reveal a striking impact of the intervention on students' behaviors and outcomes in the experimental task. In particular, we find treated students to be significantly more likely to opt for a difficult high-reward task when offered the choice against an easier low-reward alternative. Treated students are also significantly more likely to re-attempt the difficult task after receiving negative performance feedback. The design of our experimental task additionally allows us to investigate whether treated students are more likely to set for themselves the goal of succeeding in the difficult task when they are given the opportunity to accumulate task-specific skill. When given time (exactly one week) to acquire the skill needed to succeed in the difficult task, treated students are not only significantly more likely to set for themselves the goal of succeeding in the difficult task, but they are also significantly more likely to achieve this goal. More specifically, they are about 18 to 22% more likely to succeed in the difficult task, and consequently, they collect about 16 to 26% higher rewards than students in the control group. These findings suggest that treated students are more likely to set ambitious goals, are more likely to engage in skill accumulating activities, and end up with higher success as a result. We find that the estimated treatment effects are remarkably similar across the two independent samples in which we randomize the educational intervention. The replicability of

[3]Blackwell, Trzesniewski and Dweck (2007) show that the students' mindset with regard to the malleability of intelligence has an effect on the trajectory of mathematics grades among 7th graders. More evidence on the relationship between students' mindsets and achievement is provided by Aronson, Fried and Good (2002) and Good, Aronson and Inzlicht (2003).

our results is encouraging and clears the path for a potential scale-up (see Maniadis, Tufano and List, 2014).

The positive effects we estimate in the incentivized task also extend to real outcomes. Although we do not estimate a significant treatment effect on grades given by teachers, we find that treated students score 0.28 and 0.13 standard deviations higher in standardized math and Turkish tests, respectively. Compared to other estimates in the literature, these effects are large. For example, Schanzenbach (2006) reviews the existing evidence on the project STAR and finds that being randomly assigned to a small class increases student test scores by 0.15 standard deviations.

Using a production function framework similar to the one proposed in Cunha, Heckman and Schennach (2010), we elaborate on the potential mechanisms through which the treatment effects on the real effort task might be achieved. We use our survey data on beliefs and behaviors collected at baseline and follow-up to support our proposed mechanism. Estimated treatment effects on beliefs and behaviors suggest that the treatment might have affected students' beliefs about the productivity of effort in the skill accumulation process, thereby leading to more ambitious goal-setting, a higher propensity to engage in skill accumulating activities, and consequently to higher success. This mechanism is also likely to be responsible for the economically and statistically significant treatment effects we obtain on test scores.

The paper makes several contributions to the literature. To our knowledge, this is the first large-scale randomized intervention which provides causal evidence that grit, as a specific non-cognitive skill, can be improved through targeted education. The unique measurement methodology offers important advantages in capturing a rich set of beliefs, behaviors and outcomes related to grit, and in studying potential mechanisms through which the impact has been achieved. With its dynamic real-effort framework, the paper also contributes to the recent experimental economics literature on the motivational effects of performance feedback and quitting behavior in work settings (Azmat and Iriberri, 2010; Eriksson, Poulsen and Villeval, 2009; Gill and Prowse, 2012; Barankay, 2014). More broadly, the study contributes to the existing literature which uses randomized controlled trials to test the effectiveness of interventions on children's behaviors and educational outcomes (e.g. Fryer, 2011; Bettinger et al., 2012; Alan and Ertac, 2014; Kosse et al., 2015; Levitt et al., 2016).

We show that educational materials and activities, implemented in the children's natural classroom environment by their own teachers, can produce remarkable effects on behaviors related to goal-setting and perseverance, and on eventual success and payoffs in an incentivized real effort task, as well as performance in standardized tests. Given the pivotal role of non-cognitive skills for academic achievement and labor market success (Duckworth et al., 2007; Almlund et al., 2011; Kautz et al., 2014), evidence on the positive impact of such education is of utmost policy importance. Our results provide an affirmative answer to the question of whether non-cognitive skills are malleable, and highlight a particular low-cost alternative that can be implemented to foster these skills in the natural environment of the classroom. Being able to achieve such an impact in the school environment offers hope for reducing persistent achievement gaps observed in many countries, where many educational policy actions aiming to enhance family inputs face challenges in engaging families of low socio-economic strata.

The paper is organized as follows. Section 2 presents details on the design and implementation of the educational intervention, and on the measurement of the different outcome variables of interest. Section 3 contains details on the data, while section 4 presents the results. Section 5 provides a discussion and section 6 concludes.

2 Design and Outcome Measurement

2.1 Nature of the Intervention

The educational intervention consists of a carefully designed curriculum to be covered by teachers in official extra-curricular hours. Each topic is introduced by the teacher with the help of a kit that provides specific guidelines for implementing activities. The curriculum consists of animated videos, mini case studies and classroom activities that highlight i) the plasticity of the human brain against the notion of innate ability, ii) the role of effort in enhancing skills and achieving goals, iii) the importance of the constructive interpretation of setbacks and failures, and iv) the importance of goal setting. The aim of the training is to expose students to a worldview in which any one of them can set goals in an area of their interest and can work toward these goals by exerting effort. The materials highlight the idea that in order to achieve goals, it is imperative to avoid interpreting immediate failures as a lack of innate ability or intelligence. This worldview embraces any productive area of interest, whether it be music, art, science or sports. While the target concepts of the educational materials were determined by the scientific team, specific contents (e.g. scripts) were shaped with input from an interdisciplinary team of education psychologists, a group of voluntary elementary school teachers, children's story writers and media animation artists, according to the age and cognitive capacity of the students.[4] A minimum of 10 sessions were recommended to the teachers to complete the curriculum.[5]

[4]The Appendix contains an example of such material. A summary of the curriculum can be found in the online Appendix posted at https://sites.google.com/site/salancrossley/. The full translated curriculum is available upon request.

[5]Based on the feedback we received from participating teachers, on average 12 weeks were necessary to complete the curriculum. Most teachers reported that they spent at least 2 hours/week on the project.

To give an example, in an animated video, two students who hold opposite views on the malleability of ability engage in a dialog. The student who believes that ability is innate and therefore believes that there is no scope for enhancing ability through effort, points out that the setbacks she experiences are reminders of the fact that she is not intelligent. Following this remark, the student who holds the opposite view replies that she knows that setbacks are usually inevitable on the way to success; she interprets them as opportunities to learn, and therefore, they do not discourage her. The video contains further dialogs between these two students on similar ideas such as the importance of sustained effort in achieving one's long-term goals. Training materials also include stories in the form of mini-case studies containing similar ideas in different contexts. Visual materials and stories are supplemented by classroom activities created and supervised by teachers, based on general suggestions and guidelines put forward in the teacher training. For example, in a large number of schools, students prepared colorful posters that contain famous phrases of renowned individuals pertaining to the importance of grit and perseverance. These posters were exhibited in these schools in the week during which the lives of famous scientists and explorers in history were covered as part of the life sciences curriculum.[6]

In teacher training seminars, teachers were encouraged to adopt the ideas put forward in the materials as part of a teaching philosophy, and to implement them by praising the students' effort and their positive attitude toward learning, rather than just praising good outcomes. Teachers were also encouraged not to praise a successful student in a way which would imply that the student possesses superior intelligence (in the innate sense of the word), but were rather advised to highlight the role of effort in the student's success. In this sense, the intervention is not merely a set of materials to be covered in a specified period of time, like any other curriculum item, but rather an attempt to change the mindset of children by changing the mindset of the teachers.[7]

2.2 Evaluation Design

The Turkish Ministry of Education encourages schools and teachers to participate in socially useful projects offered by the private sector, NGOs, the government and international organizations. These projects, upon careful examination and endorsement by the Ministry, are made available to interested schools. The Ministry allows up to 5 lecture hours per week for project-related classroom activities,

[6]Oversight of the ministry and the input received from independent school teachers in preparation of the materials ensured that all activities and reading materials complemented the existing curricula.

[7]In order to assess how successful this attempt was, we conducted an anonymous survey among teachers at the end of the academic year and asked about their views on the ideas put forward in the materials. More than 95% of all teachers report that they agree with the ideas conveyed by the training and more than 90% report having implemented the training moderately or intensely. Anecdotally, the field partner who maintained contact with teachers after the conclusion of the intervention, received feedback from teachers supporting our point that the ideas put forward in the offered materials were adopted as a teaching philosophy. Several teachers revealed that they were implementing these ideas for their new cohort of students (in the absence of an official program).

and participation in these projects is at the discretion of teachers. Subject matters for these projects are many, and typical examples include environment, art, foreign languages, health, and dental care. These are generally high quality projects designed and offered by the Turkish Ministry of Health, the Ministry of the Environment and international organizations such as the Regional Environmental Center. In the absence of any projects, students use the free hours as unstructured playtime, so these projects do not crowd-out any core teaching.

The program we develop and evaluate is offered as part of a corporate social responsibility project of the Turkish division of a major international bank. The main objective of the program is to improve key non-cognitive skills in elementary school children in the classroom environment. The program is implemented by the students' own teachers in a large number of state elementary schools in Istanbul with the permission and oversight of the Turkish Ministry of Education. In the last few decades middle class families in Turkey tend to prefer private schools over under-resourced state schools for their children. Therefore, the program mainly reaches students from lower socio-economic backgrounds.

We implement and evaluate the program as two independent studies, resulting in two independent samples. In both samples, we randomize the intervention across schools in which at least one teacher stated his/her willingness to participate in the program. In the first study, students who received training on grit in Fall 2013 had already received another training in Spring 2013, which aimed to improve the ability to make decisions in a forward-looking manner and encourage patience.[8] This sample (Sample A henceforth) consists of 36 schools, 15 of which are assigned to treatment (grit+patience) and 21 to control, and has a total of about 1,700 students in 64 classrooms. Among those 21 control schools, 9 of them received only the patience treatment. While we can evaluate the independent impact of the patience treatment on our outcome measures (and indeed show that the patience treatment by itself does not produce gritty behavior), we are not able to isolate or rule out potential complementarities across the grit and patience training using Sample A. The second study, which was implemented in the school year 2015-2016 and essentially provides a replication sample, resolves this issue.

In the second randomized field experiment, we randomly assign the same grit intervention across a new set of schools in Istanbul. This sample (Sample B henceforth) consists of 16 schools (8 treatment, 8 control) and has a total of about 1,200 students in 42 classrooms. While the intervention followed the same procedures (same curricular materials and the same teacher training approach), there are a couple of important differences in the way the study was conducted. These differences were made to alleviate potential issues with the design of the first study, which were due to logistical constraints. First, in the replication study neither the treatment nor the control schools were subject to any other treatment

[8]The results of the evaluation of this intervention are reported in Alan and Ertac (2014).

than grit. This allows us isolate the effect of the grit intervention. Second, we administer our own standardized tests both at baseline and at follow-up. These tests measure students' performance in math and Turkish, two core subjects that are of utmost importance in terms of the further academic endeavors of students. The test scores give us a solid measure of real outcomes on which we can evaluate the impact of the training.[9]

In both studies, the randomization was performed in the following way. First, the Istanbul Directorate of Education sent the official documentation of the program to all elementary schools in designated districts of Istanbul.[10] The teachers in these schools were then contacted in random sequence and offered to participate in the program.[11] Once a teacher stated a willingness to participate, we assigned their school into the treatment or control group.[12] In Sample A, baseline data were collected in Spring 2013, the first intervention (patience) was implemented in Spring 2013, and the intervention on grit was implemented in Fall 2013. The follow-up data related to the grit intervention was collected approximately four months after the implementation of the intervention, in Spring 2014. In Sample B, the baseline data were collected in Spring 2015, the intervention (grit only) was implemented in Fall 2015, and the follow-up data were collected in January 2016.[13] Full details of the evaluation designs for each study sample are given in Table 1.

2.3 Experimental Outcome: A Real Effort Task

We estimate the effect of the intervention on students' behaviors and outcomes in an incentivized experimental task designed to measure grit. Our design includes two different visits, a week apart from each other. In the first visit, children go through five rounds of a mathematical real effort task.

[9] Another difference between Sample A and Sample B is that the students in Sample B are about 6 months younger than the students in Sample A. This is because of an unexpected educational reform implemented in 2012 that lowered the age at which children start school.

[10] The general intervention program was titled "financial literacy, savings and economic decisions" and no further information on the particulars of the program were disclosed to the teachers prior to the teacher training seminars.

[11] Teachers were informed that upon participation they would be assigned to different training phases within the coming two academic years. All teachers who agreed to participate were promised to eventually receive all training materials, and to participate in training seminars, but they were not told when within the next two academic years they would receive the treatment until the random assignment was completed. The promise of the training offer was made to the teacher and not to current students, i.e. while children in control groups will never receive the training as they move on to middle school after year 4, their teachers will, albeit at a later time.

[12] The sample generated with this design contains schools in which at least one teacher stated their willingness to participate in the program. Therefore, the estimated impact of the program is the average treatment effect on the treated and is not readily generalizable to the population. However, in the study sample A approximately 60% of the contacted teachers accepted our offer and the most common reason for non-participation was being "busy with other projects, although happy to participate in this program at a later date" (about 20%). The rest of the non-participation was due to "impending transfer to a school in another city, with a willingness to participate if the program is implemented there" (about 5%), and "not being in a position to participate due to private circumstances" (about 10%). In study sample B, acceptance of the training offer reached 80%. Given these numbers, we conjecture that the external validity of our results is strong.

[13] Note that in order to ensure data quality, authors Alan and Ertac coordinated the field logistics, trained a select group of students and experienced interviewers to assist with data collection, and physically visited all schools to collect data. All measurements were conducted with the approval of the local IRB and the permission of the Ministry.

In particular, they are presented with a grid which contains different numbers where the goal is to find pairs of numbers that add up to 100. At the end of the five rounds, one of the rounds is selected at random and subjects get rewarded based on their performance in that round. Rewards depend on meeting a performance target. In all the tasks we present to the children, the target is to find three pairs of numbers which sum up to 100, within 1.5 minutes.[14] Before each round starts, subjects have the chance to choose between two different types of tasks for that round: (1) the "4-gift game", which yields four gifts in the case of success and zero in the case of failure, and (2) the "1-gift game", which yields one gift in the case of success and zero in the case of failure. Although in both games the goal is to find at least three pairs of numbers adding to 100, the 4-gift game is more difficult than the 1-gift game. In particular, in the 1-gift game the grid is smaller, and the matching pairs are easier to identify.[15] In fact, the mean empirical success rate in the easy task ranges from 97% to 100% over the five rounds.

Before the five periods start, all subjects are given a large grid that contains many matching numbers and they are given two minutes to find as many pairs of numbers that add to 100 as possible. This is intended to both familiarize the children with the task before they make decisions, and measure task-specific ability. The rewards are such that children get a small gift for each pair they can find.[16] In the main 5-round part of the experiment, subjects are distributed two booklets of 5 pages each, the 4-gift game booklet and the 1-gift game booklet. Each booklet contains 5 pages that correspond to the 5 rounds of the relevant type of game. In addition, subjects are distributed a "choice sheet". Before a typical round starts, subjects are instructed to circle their game of choice for the upcoming round in their choice sheet, and then get ready to open the relevant page of their booklet of choice. They are then given 1.5 minutes to find as many matching number pairs as they can. All students are instructed to fold their arms once the 1.5 minutes are over.[17] During this time, experimenters go around the class and circle either "Succeeded" or "Failed" on the students' sheets for that round, based on whether at least 3 pairs were correctly found. As mentioned above, students have the opportunity to switch back and forth between the two types of tasks as the rounds progress.

The above procedures, whereby students work on their task of choice in each round have one exception. In the first round, the students' choices are implemented with 50% chance, and with 50%

[14] Note that while sample A students are given 1.5 minutes for each round, sample B students are given 1 minute 45 seconds for each round. We chose to give sample B students more time because they were on average younger than sample A students. See section 3 for more details on the characteristics of the two samples.

[15] See Appendix for examples of the two types of task.

[16] These small gifts (e.g. a regular pencil, single hairpin etc.) are significantly lower in value than the gifts used as rewards in the actual task, and the children are aware of this. In addition, information about actual rewards they receive from this task is not revealed until the end of the 1st visit.

[17] We learned from pilot sessions (at an out-of-sample school) that folding arms and being dead-silent, which is called "becoming a flower", is a well-known state to children, and indeed children were very obedient when instructed to do this.

chance they play the difficult game irrespective of their choice. This procedure allows us to have a random subset of children where the effects of 1st round feedback on 2nd round task choices can be analyzed free of selection. From the 2nd round onwards, students are completely free in their choices, and their choices are implemented with 100% chance.

After the five rounds are completed, we inform the children that we will visit their classrooms once more in exactly a week's time. The children are told that they will play the game one more time during the second visit, and that they need to decide now whether they would like to play the 4-gift (difficult) game or the 1-gift (easy) game in a week's time. They are also told that they will have access to an "exercise booklet" which contains examples and practice questions that have a similar difficulty level to the 4-gift game. Just as in the first round, in order to get a subsample to play the difficult game free of selection, the students' choices are implemented with 50% chance, and with 50% chance they play the challenging game in the next visit. Students are aware of this procedure when they make their choices, and they are also informed about which game they are going to play in the 2nd visit at the end of the 1st visit.[18]

In the second visit, children perform the task they chose at the end of the first visit or the difficult task, depending on whether the difficult task was imposed for them or not. They again have 1.5 minutes to find pairs of numbers that add up to 100. The game is played for one round, and rewards are based on performance during that round.[19]

3 Data and Baseline Information

We estimate the treatment effects separately for each study sample. Sample A contains information on 64 classrooms in 36 schools (15 treatment, 21 control), and it includes about 1,700 students. Sample B contains information on 42 classrooms in 16 schools (8 treatment, 8 control), and it includes about 1,200 students.

For both samples, the baseline data contain a rich set of key student characteristics. In addition to collecting information on demographic variables such as gender and age, we administer a Raven's progressive matrices test to obtain a measure of cognitive ability (Raven, Raven and Court, 2004). Moreover, we measure students' risk tolerance using a version of the Gneezy and Potters (1997) risk preference elicitation task. Importantly, we also have information on students' (i) baseline beliefs about the malleability of ability, and (ii) attitudes and behaviors related to grit and perseverance.[20] From

[18] Actual rewards from the first visit are not revealed until after all the choices have been made for the second visit. In total, the first visit takes one lecture hour.

[19] The reward basket in the second visit contains the same array of items that were used as rewards in the first visit.

[20] To obtain these measures we extract the first principal component from the students' responses to questionnaire

teacher assessments we also have a measure of the students' academic success and the socio-economic status of the students' families.

Both samples contain measures of prior academic achievement. In Sample A, we have information on students' pre-intervention exam grades in two core subjects (mathematics and Turkish), which we obtain from administrative records. For the purpose of the analysis we normalize the grades to have a mean of zero and a standard deviation of one. In Sample B, we administer standardized tests at baseline in math and Turkish. We also normalize the standardized test score results so that they have a mean of zero and a standard deviation of one. We use these baseline measures to assess the samples' balance across treatment status. Table 2 provides the balance tests for Sample A and Sample B. In Sample A, we do not observe any statistically significant differences in student characteristics, test scores or beliefs. In Sample B, most characteristics, test scores and beliefs are also balanced, although there are some significant differences across treatment and control. We use a number of baseline variables as covariates in the estimation of the average treatment effects to increase the precision of our estimates and to account for potential imbalances in baseline covariates which are predictive of our outcome measures.

Next we investigate whether students with different treatment status have different task-specific ability at the beginning of the incentivized experiment. As explained in section 2, at the beginning of the first visit, there is an initial round that aims to familiarize students with the actual task. This round is designed to facilitate informed decision making for the following five rounds, and it allows us to measure the students' task-specific skill level. This task consists of finding as many matching pairs as possible in a large grid of numbers and it is incentivized. As can be seen in table 2, the number of pairs found in this task is not different across treatment status in either sample (see variable "Task Ability"). Since this measure is balanced across treatment groups and very highly predictive of the experimental outcomes, we use it as one of our covariates in estimating average treatment effects. Note, however, that the randomization ensures that the results (estimated effect sizes) are not affected by the choice of covariates, and that the highly predictive covariates such as task-specific ability, cognitive ability and math grades greatly improve the precision of our estimates.

Before turning to the estimation of the treatment effects, we first investigate whether decisions in our experimental task are predictive of real outcomes. Using observations from the control group only, we show that in both samples behavior in the real effort task predicts baseline math and Turkish scores, over and above what can be predicted by cognitive ability measured by the Raven test (table 3). In

items which relate to these beliefs and behaviors. The variables are normalized to have mean zero and standard deviation one.

Sample A, students who choose the difficult task in all five rounds of the first visit have significantly higher math and Turkish scores at baseline (columns 1 and 3). Whether or not the student chooses the difficult task for the second visit similarly predicts significantly higher math and Turkish scores (columns 2 and 4). In Sample B, we obtain qualitatively similar results although the significance levels of the estimates differ somewhat from those obtained in Sample A, possibly due to the smaller sample size. Overall, these results are consistent with previous findings which suggest that measures of non-cognitive skills can predict outcomes over and above what can be predicted by cognitive ability alone (Roberts et al., 2007).

4 Results

4.1 Estimation of Treatment Effects

In order to test the null hypothesis that the program had no impact on the experimental outcome y^E, we estimate the average treatment effect conditioning on baseline covariates:

$$y_{ij}^E = \alpha_0 + \alpha_1 T_j + X_{ij}^{'}\gamma + \varepsilon_{ij}$$

where T_j is a dummy variable which equals 1 if school j is in the treatment group and zero otherwise, and X_{ij} is a vector of observables for student i in school j that are potentially predictive of the outcome measures we use. These include performance in the task ability elicitation round, cognitive ability which we measure using the Raven test (Raven, Raven and Court, 2004), the student's math and Turkish scores prior to the intervention, risk tolerance measured via an incentivized task (Gneezy and Potters, 1997), gender, and two factors extracted from the baseline survey which capture (i) students' beliefs about the malleability of ability and (ii) students' attitudes and behaviors regarding grit and perseverance.

The estimated $\hat{\alpha}_1$ is the average treatment effect on the treated. Estimates are obtained via a logit regression when the outcome considered is binary. This is the case for students' choices between the difficult and the easy task, and for their success/failure in meeting the target. The binary outcome variable "success" is defined as finding three correct pairs or more. In the case of payoffs, the above equation is estimated via ordinary least squares. The outcome variable "payoff" takes the value 0 if the target of finding three pairs is not met, 1 if the easy game is played and the target is met, and 4 if the difficult game is played and the target is met.

In order to test the null hypothesis that the program had no impact on the real outcome y^R, we estimate the average treatment effect using the same specification but a different set of covariates. In particular, when estimating the average treatment effects on test scores we control for gender, the Raven score, risk tolerance, the students' baseline test scores in math and Turkish, and class size.[21]

4.2 Treatment Effect on the Real Effort Task

In the following, we examine the effect of treatment on students' choices and outcomes in the incentivized real effort task. While section 4.2.1 presents the results for the first visit, section 4.2.2 presents the results of the second visit.[22]

4.2.1 Treatment Effect on Choices and Outcomes in Real Effort Task: First Visit

First, we estimate the effect of treatment on students' choice of task difficulty during the first visit. Table 4 presents marginal effects from logit regressions in which we regress the choice of task difficulty on a treatment dummy and a set of covariates. The first finding to note in this table is that in both samples, the proportion of students in the control group who attempt the difficult task declines visibly through the rounds. While about 73% of the control students attempt the difficult task in the first round in Sample A (71% in Sample B), only 41% attempt the difficult task in round five (27% in Sample B). This switching trend from difficult task to the easy one is true for the treated students as well and is indicative of a rational response to the incentives the task entails. Recall that the easy task, in which the success rate is very high, brings one gift in the case of success. However, as can be seen in the table, treated students are significantly more likely to attempt the difficult task in each of the five rounds. In fact, treated students are 9 and 13 percentage points more likely to choose the difficult task in all five rounds in Sample A and Sample B, respectively. Figure 1 shows the estimated treatment effects for the two samples with 95% confidence intervals.

Why does the treatment cause such persistence in attempting the more challenging task? One explanation may be that the treated students believe that they can improve their performance on the difficult task with repeated attempts. However, this belief tends to fade as it becomes apparent that acquiring ability in this task requires more time and effort than possible in the short period of time

[21] In all empirical analyses, standard errors are clustered at the school level which is the unit of randomization. Our results are also confirmed if we assume the most conservative intra-cluster correlation, which is 100%. That is, our results carry through even when we collapse our data into school means and estimate treatment effects using only 36 (Sample A) or 16 (Sample B) observations. All related analyses are available upon request and are also available in the online Appendix posted at https://sites.google.com/site/salancrossley/.

[22] For the sake of brevity, all tables in this section present the estimated treatment effects without presenting the coefficient estimates of the other covariates. Full estimation results where we also present the coefficient estimates of the covariates can be found in Appendix A (Sample A) and Appendix B (Sample B).

provided for each round in the first visit, which makes the incentive to switch to the easy task stronger. We provide a discussion of the potential mechanisms of behavioral change in the first and the second visit in section 5.2.

Next, we investigate whether treatment affects the students' experimental outcomes, namely, success and payoffs. Table 5, column 1 presents the estimated treatment effects on success in round 1 of the first visit for the sample which was forced to play the difficult game. This particular round is designed in such a way that allows us to estimate the treatment effect on success in the difficult game free of selection. In both samples, we find that treatment had no significant effect on success rates. This is also true for payoffs in all rounds: the estimated treatment effects on payoffs in all five rounds are not statistically different from zero, with the exception of the first round in Sample B, which is positively significant at the 5% level.

Figure 2 depicts the estimated treatment effects on payoffs for the two samples, along with 95% confidence intervals. These figures clearly indicate that despite the fact that a higher proportion of students in the treatment group chooses to do the difficult task, and despite the fact that the probability of success is much lower for the difficult task, treated students do not end up with lower payoffs on average.

The Effect of Negative Feedback

A constructive response to negative feedback is perhaps one of the most important skills associated with achievement in many domains. An important aspect of our intervention is to teach students not to immediately attribute failure to a lack of innate ability. Perseverance in the face of a setback is consistently praised and encouraged during training hours.

In an observational setting, the effect of negative feedback on an individual's subsequent actions cannot be assessed in a framework that is free from selection; feedback, by definition, comes in response to an action. Since engaging in an action in a voluntary manner is likely to be endogenous to individual characteristics, so is the feedback such an action generates and so is the subsequent response to such feedback. This is evident in the case of our experimental setting, as students who self-select into the difficult task are more likely to receive negative feedback, and they differ sharply from those who do not select into the difficult task. In order to assess the response to feedback, one must ensure an exogenous variation in feedback, which is typically not possible in observational settings.

To identify the effect of the training on students' response to negative feedback, a random subset of all students is forced to play the difficult game in the first round irrespective of their choice. This experimental design renders playing the difficult game free from selection so we can assess the effect of failure on subsequent choices, i.e. the choice for the second round.

Table 6 column 1 presents the estimated treatment effect on the probability of choosing to play the difficult game in the second round after failure in the first round. This analysis is done for the students who were forced to play the difficult game regardless of their choices in the first round and failed to meet the target of finding three correct pairs. The estimated difference between the treatment and control groups is striking. Among the students who failed in the imposed difficult task in Sample A, treated students are 18 percentage points more likely to re-attempt the difficult task in the subsequent round (p-value=0.002). The corresponding estimate in Sample B is 14 percentage points (p-value=0.064).

This provides strong evidence that treated students persevere more after negative feedback, and we note this as one of the most encouraging impacts of the intervention. The fact that treated students are more likely to persevere after failure is also consistent with the idea that treated students believe that they can get better at the difficult game with repeated attempts.

Goal Setting and Commitment to Ability Development

As explained in Section 2.3, at the end of the first visit we let the students know that we will come back exactly one week later and that they will play the same game for one additional round. We also inform them that if they like, they can take a study booklet covering numerous examples of the difficult game and study/practice over the week using this booklet. We emphasize that this is entirely voluntary. We then collect their decisions on which type of task they would like to do in the following week. After we collect these choices, students are informed whether they will have to play the difficult game in the following week, or the game of their choice. The purpose of this exercise is to test whether the treatment generates "goal-setting" behavior in the form of commitment to improve task related ability in the six days before the second visit. We predict that students who believe that ability in this task is malleable through sustained effort and perseverance are more likely to set the goal of succeeding in the difficult game and therefore more likely to commit to playing the difficult game. This is exactly what we see in Table 6, column 2: treated students are estimated to be 14 percentage points more likely to plan to play the difficult game in the following week in Sample A, and 19 percentage points more likely to plan to play the difficult game in Sample B.

Treated students did in fact set themselves the goal of succeeding on the difficult task in the second week - but did they actually achieve this goal? This is the question we explore in the next subsection.

4.2.2 Treatment Effect on Choices and Outcomes in Real Effort Task: Second Visit

The temporal component of our experimental task, which involves another visit exactly one week later, serves a very important purpose for our study. While it is unlikely that students can improve their ability on a task within five rounds of only 1.5 minutes, it may well be that students can accumulate task-specific ability when given sufficient time. Undoubtedly, ability accumulation takes time and effort, and the amount of time and effort required to master a task varies according to the characteristics of the task and the characteristics of the person who is trying to master it. In this specific real effort task, we chose to give students one week, with the conjecture that it would be sufficient for motivated students to work through the exercises provided in the study booklet and that such effort would lead to a higher probability of success in the second visit.

As in the first round of the first visit, a random subset of students is forced to do the difficult task during the second visit, irrespective of their choice. This allows us to investigate whether the treatment affects the probability of success in the difficult task in the second visit. Table 7 presents the estimated treatment effects on outcomes of the second visit. The first column presents the treatment effects on success obtained from the sample on which the difficult task is imposed, while columns 2-4 present the treatment effects on payoffs. For the latter, we estimate treatment effects on the entire sample as well as conditional on whether the difficult task was imposed in the class or not. Looking at the first column for both samples, we see that treated students are about 9 (11) percentage points more likely to succeed in the difficult game in Sample A (Sample B). These effects are statistically significant. The increased success rate is also reflected in payoffs: we estimate a statistically significant 16% and 26% treatment effect on payoffs in Sample A and Sample B, respectively (0.30 and 0.46 more gifts for treated students in Sample A and Sample B, respectively). Note also that the estimated effects are similar for the imposed and unimposed samples. Considering the combined payoffs of both visits (the last column), we estimate 11% higher payoffs in Sample A, and 20% higher payoffs in Sample B.

A natural question is whether there is a type of student for whom the treatment was particularly successful. Presumably, treatment may have a differential impact on students with different task-related ability levels. For example, it is conceivable that a treatment of this sort might be effective in pushing a potentially able but reluctant student into planning to do the difficult task and in encouraging her to study. On the other hand, it may encourage a student with low ability to study as well. Since the performance technology is conducive to ability accumulation, we might also observe increased success rates in the second week for these students. Our analyses do not reveal any systematic heterogeneity in treatment effects with respect to gender, task ability and cognitive function.[23]

4.3 Treatment Effect on Test Scores

The implication of a change in beliefs regarding the malleability of skills can be far reaching. For one thing, a student who used to think that there is not much one can do to excel in an area, whether that be related to art or science, may now be convinced that all it takes is goal-setting and hard work. If this is the case, even in the short-run, we may be able to see improvements in other domains where sustained effort results in better outcomes. The obvious outcome to look at in this regard is actual school grades. For this purpose, for Sample A, we collect official grades (marked by the teacher) that reflect the students' math and Turkish performance in Spring 2014, several months after the implementation of the training. For Sample B, instead of relying on teacher-marked exam results, we administer our own tests to measure students' ability in math and Turkish. These tests were prepared based on the official curriculum for the 4th grade.

While we find no significant impact of treatment on average teacher-given grades in Sample A (table 8), we find remarkably large and significant treatment effects on standardized test scores in Sample B (table 9).[24] In particular, we find that treatment increases student performance in the standardized math test by 0.28 standard deviations (p-value=0.008), and student performance in the standardized Turkish test by 0.13 standard deviations (p-value=0.027). Compared to other estimates in the literature, these effects are large. For example, Schanzenbach (2006) provides a review of the existing evidence on the project STAR and concludes that being randomly assigned to a small class raises student test scores by 0.15 standard deviations. Our intervention can be delivered by the students' own teachers in the classroom environment, which is why it comes at a relatively low cost and is therefore likely to have a high rate of return.

These results are encouraging. First, if such an intervention is able to increase student performance via changing their mindset about the role of effort in achievement, a similar but longer, more intense intervention has a lot of scope. Second, the aspirations of students who experienced high achievement through effort may change significantly and these changes may become permanent by creating a productive cycle of further effort and further success. The latter is particularly important from the policy perspective for regions in which parents are generally not engaged in the skill formation of their children for reasons such as illiteracy and poverty, and as a result achievement rates are dismal.

[23] Full results on heterogeneity are available upon request.

[24] A potential reason why teacher-given grades are unaffected by the intervention is that teachers tend to apply a relative grading scheme. For this reason, and the possibility that the teacher's assessment may have been affected by the treatment in an unknown way, we chose to administer standardized tests in Sample B that do not rely on teacher assessments.

5 Discussion

5.1 Are Choices Payoff-Maximizing?

An important question as regards an intervention of this sort is whether being gritty is good for everyone, i.e. whether it is optimal for children to always set challenging goals, persevere in the case of setbacks and engage in costly skill accumulation activities. Certain endeavors might not be worth the time and effort if they are unachievable or if the costs of perseverance required for success are so high that they outweigh the potential gains. In general, perseverance is more likely to pay off when the performance technology is conducive to skill accumulation and the costs of effort or investment are not too high.

In order to get some insight into this question, we investigate to what extent individual choices of task difficulty are payoff-maximizing in expectation. More specifically, we first obtain an individual measure of each student's probability of success in each task given the student's baseline characteristics, using the empirical distribution of success.[25] We then calculate the student's expected payoff from choosing the difficult task and compare that with the expected payoff from choosing the easy task. Once we have obtained an estimate of which task choice would be payoff-maximizing for each student, we compare this payoff-maximizing choice to the student's actual task choice.

In Sample A, treated students are no more likely to choose the payoff-maximizing task in the first round of the first visit (table 10, column 1) but they are 9 percentage points more likely to choose the payoff-maximizing task for the second visit (table 10, column 2). In Sample B, students are more likely to choose the payoff-maximizing task in both visits. In particular, treated students are 14 percentage points more likely to make the payoff-maximizing choice in visit 1, and 10 percentage points more likely to make the payoff-maximizing choice in visit 2. Overall, we conclude that treated students were more likely to make decisions that were payoff-maximizing in expectation.[26]

5.2 A Potential Mechanism

Taken together, the findings indicate that the intervention was remarkably successful in changing both students' behaviors and outcomes in the incentivized experimental task, and their performance on the

[25]To obtain a measure of the student's probability of success in the first (second) visit we run a logit regression of a dummy variable which equals 1 if the student was successful in the imposed difficult task in the first (second) visit on a set of covariates such as baseline task ability, gender, Raven score, pre-treatment beliefs, math and Turkish test scores and risk tolerance, and we obtain the predicted success probability for each student based on their individual characteristics. We apply this procedure separately for treatment and control group students, and we follow a similar strategy for obtaining the probability of success in the easy task.

[26]Note that it is difficult to make statements about utility (rather than payoffs), since effort costs are unobservable. However, the incidence of treated students choosing the difficult task for the 2nd visit and succeeding suggests, through revealed preference, that these choices might also be utility-maximizing for that sample.

standardized math and Turkish tests. Guided by our empirical results from the experimental task, we now propose a potential mechanism through which these changes might have been achieved. For this purpose, we use (i) a production function that maps task ability and on-the-task effort into the number of pairs found (output), and (ii) an underlying ability accumulation technology that helps us characterize behavior in-between the two visits. With this model, we can offer an interpretation for our findings in the experimental task and, more importantly, have a theoretical framework we can use to think about the impact of the intervention on performance outcomes outside of the experimental task, such as performance in standardized tests.

Suppose that student i has a "true" production function which takes the standard constant elasticity of substitution (CES) form. For the sake of brevity assume that output q_k, which is the number of pairs found in the difficult game in a given round k, is a function of two main inputs, task ability and effort:[27]

$$q^1_{i,k} = A[\alpha a^{\rho}_{1,i} + \beta_i E^{\rho}_{i,k}]^{\frac{1}{\rho}} \varepsilon_{i,k} \tag{1}$$

where $q^1_{i,k}$ is the number of pairs student i finds in round k in visit 1, A is the productivity parameter, $a_{1,i}$ is individual task ability in visit 1, $E_{i,k}$ is the effort the individual chooses to exert on the task, and $\varepsilon_{i,k}$ captures all omitted inputs and individual- and round-specific shocks. This production function is sufficiently flexible, as the substitutability of inputs is governed by the parameter ρ, via which perfect complementarity (as $\rho \to -\infty$), perfect substitutability (as $\rho \to 1$), and Cobb-Douglas (as $\rho \to 0$) cases can be obtained. Using this framework, we postulate that the intervention may have changed the *perceived* production function by changing the students' beliefs about the parameters. In principle, all the parameters (A, α, β and ρ) in this representation could be influenced by the intervention. However, given the nature of the intervention and our empirical results, we argue that the intervention has likely increased the perceived marginal product of effort E such that, using the potential outcomes framework, the perceived productivity of effort is higher for student i in round k if she is treated:

$$\beta_{i,d=1} \geq \beta_{i,d=0} \tag{2}$$

where the subscript d denotes treatment status, which equals 1 if student i is treated. If students choose effort to maximize payoffs taking their perceived production function and the perceived level of their task ability as given, then this condition will lead to treated students being more likely to choose the difficult task in any of the five rounds of the first visit, as they will be more likely to be believe

[27] We define success as finding $\bar{q}$ pairs or more in the difficult game but the same threshold is imposed for the easy game. For the latter, the observed success rate is almost 100%.

that they can meet the performance target by exerting on-the-task effort.[28]

While in the data we observe that the treatment effect persists throughout all five rounds, the proportion of students choosing the difficult task declines in both the treatment and the control group as the rounds progress. This result suggests that students in both groups revise their beliefs about the probability of success in the difficult task downwards. This decline could be due to students updating their beliefs about their task ability or the marginal productivity of effort downwards as they receive negative performance feedback.

After the first visit, students decide whether to engage in skill accumulating investment activities between the two visits. We assume a CES skill accumulation technology similar in spirit to Cunha, Heckman and Schennach (2010)

$$a_{2,i} = \Psi[\gamma a_{1,i}^{\sigma} + \lambda_i I^{\sigma}]^{\frac{1}{\sigma}}, \tag{3}$$

where $a_{1,i}$ denotes student i's task ability in the first visit, $a_{2,i}$ denotes student i's task ability in the second visit, and I denotes the investment made by the student to develop task ability in-between the two visits. In our context, the intervention most likely influenced the perceived marginal product of this investment:

$$\lambda_{i,d=1} > \lambda_{i,d=0}. \tag{4}$$

A change in the perceived value of λ would not only result in students setting more ambitious goals for the following visit but also result in students making higher investments in-between the two visits and actually being more successful in the second visit. More specifically, re-writing the true production function of the task in the second week as

$$q_i^2 = A[\alpha a_{2,i}^{\rho} + \beta_i E_i^{\rho}]^{\frac{1}{\rho}} \epsilon_i, \tag{5}$$

student i is more likely to succeed in the second visit if treated since, everything else equal,

$$a_{2,d=1} > a_{2,d=0}. \tag{6}$$

Overall, changes in the perceived role of on-the-task effort and the perceived role of investments emerge as a likely mechanism that can explain both the first-visit and the second-visit results.[29] While

[28] A student who chooses effort to maximize payoffs will choose to play the difficult game if and only if her perceived payoffs from choosing the difficult task are higher than her perceived payoffs from choosing the easy task, i.e. if $Pr_i(Success_{Diff})\pi^{Diff} > Pr_i(Success_{Easy})\pi^{Easy}$, where $Pr_i(Success) = Pr_i(q \geq \bar{q})$, π denotes payoff and $\bar{q}$ denotes the required threshold.

[29] Other potential changes in the perceived production function as a result of treatment might have been through

a shift in students' beliefs about the importance of on-the-task effort can explain why treated students persistently choose the difficult task more often during the first visit, a shift in students' beliefs about the importance of effortful investments can not only explain why treated students are more likely to choose the difficult task for the following week but also why treated students are more likely to actually succeed in the second visit. It is likely that the shift in beliefs has resulted in treated students engaging more in skill accumulating activities in-between the two visits, which is why they were more likely to succeed and obtain higher payoffs during the second visit.

More support for this particular mechanism comes from our survey data. Table 11 presents the estimated treatment effects on the factors measuring students' beliefs about the malleability of ability (column 1) and students' self-reported levels of perseverance (column 2). In both samples, treated students report about 0.3 standard deviations higher belief in malleability. Consistently with these results, treated students report about 0.3 standard deviations higher perseverance. Figures 3 and 4 show the shift in the distributions in visual clarity.[30] These results support our proposed mechanism. The intervention is likely to have changed the students' beliefs about the skill accumulation technology by emphasizing the role of effort in eventual success. A student who once believed that ability is a fixed individual trait, may now think that it is malleable through effort. Such a student would be expected to now invest more time and effort into a given task, accumulate more skills and consequently be more likely to succeed in meeting the performance target. Such an increase in the students' beliefs about the malleability of ability is also likely to lead to an increase in perseverant actions as reported by the students, since these actions will be perceived to be more rewarding in a world where repeated attempts/trials with respect to a specific task lead to skill accumulation.[31] This mechanism may also explain the significant treatment effects we estimate on standardized test scores. .

lowering the perceived importance of ability (α) or increasing the perceived substitutability parameter ρ in the production function. Treatment making students believe that ability is not that important (lower α) would, however, be inconsistent with the 2nd-visit results. Students who believe that ability is not that important would be *less* willing to invest in costly skill accumulation in-between the two visits, which would result in lower success rates. Similarly, treated students believing more in the substitutability of task ability with on-the-task effort would invest *less* in ability development for the second visit, and would hence also be less likely to actually succeed.

[30]To construct these graphs, we obtain the residuals from the regression analysis performed in table 11, and we plot the kernel density of the residuals by treatment status for each sample.

[31]Note that an intervention of this sort may have acted as a confidence booster, since it emphasizes the importance of not attributing setbacks to a lack of ability. For this reason, a student who once interpreted failure as strong evidence for a lack of ability, might now end up with a less diminished self-confidence level and be less reluctant to re-attempt a challenging task. In the above framework this may also affect the parameters of the perceived production function and skill accumulation technology. However, we do not estimate any significant treatment effect on confidence scores (column 3), which suggests that a boost in self-confidence is unlikely to be the mechanism through which the estimated results were obtained.

6 Conclusion

Using two independent study samples, we evaluate a large-scale randomized educational intervention that aims to enhance grit in the classroom environment. We estimate the effect of treatment on (i) students' behaviors and outcomes in an incentivized experimental task and (ii) students' grades and performance in standardized tests after the implementation of the intervention. We find significant treatment effects of the intervention on students' behaviors and outcomes in the real effort task, which are remarkably similar across the two independent samples. In both samples, treated students are significantly more likely to select challenging tasks, more likely to engage in skill accumulating activities, and more likely to accumulate more skills and obtain higher payoffs as a result. Moreover, treated students perform significantly better on standardized tests at follow-up. In particular, the treatment raises students' performance on a standardized math test by 0.28 standard deviations, and on a standardized Turkish test by 0.13 standard deviations. Compared to the impacts of other interventions (e.g. Schanzenbach, 2006), these effects are of considerable size.

From the policy perspective, the paper contributes to the ongoing debate about the malleability of non-cognitive skills and the role of educational programs in enhancing individual achievement through interventions that specifically target those skills (Almlund et al., 2011; Kautz et al., 2014). Our results provide an affirmative answer to the question of malleability within the context of an important non-cognitive skill, and highlight a particular low-cost alternative that can be implemented to foster it in the natural environment of the classroom. Being able to achieve such an impact in the school environment offers hope for reducing persistent achievement gaps observed in many countries, where many educational policy actions aiming to improve family inputs face challenges to engage families of low socio-economic strata.

A potential caveat relates to external validity, since program participation was voluntary. However, the majority of teachers who did not participate in our program expressed constraints rather than a lack of willingness to participate (see section 2). In both of the studies, less than 10% of the contacted teachers rejected our program offer by either stating that they are "not in a position to participate due to private circumstances" or simply by not returning our call. Given these numbers, we believe that the external validity of our results is strong and scaling-up a low-cost program like this is likely to be effective.

On a final note, our results are short-term and we do not yet know whether the impacts achieved will change over time as students become adolescents. However, the results may not be short-lived if the intervention altered beliefs and aspirations and encouraged some students to exert extra effort and consequently to succeed, since success realizations attributed to high effort might create a productive cycle of further effort and further success. In this regard, future avenues of research this study motivates abound, ranging from settings where alternative intensities (durations) of interventions are tested to measuring longer term effects on both students in our samples and new cohorts of students who will be taught by the teachers in our samples.

References

Alan, S., and S. Ertac. 2014. "Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference." Available at SSRN 2566405.

Almlund, M., A.L. Duckworth, J.J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." *Handbook of the Economics of Education*, 1–181.

Aronson, J., C. Fried, and C. Good. 2002. "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." *Journal of Experimental Social Psychology*, 38: 113–125.

Azmat, G., and N. Iriberri. 2010. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students." *Journal of Public Economics*, 94: 435–452.

Barankay, I. 2014. "Incentives and Social Tournaments." unpublished manuscript.

Bettinger, E., B. Long, P. Oreopoulos, and L. Sanbonmatsu. 2012. "The Role of Simplification and Information: Evidence from the FAFSA Experiment." *Quarterly Journal of Economics*, 123(3): 1205–1242.

Blackwell, L., K.H. Trzesniewski, and C.S. Dweck. 2007. "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention." *Child Development*, 78: 246–263.

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

Castillo, M., P.J. Ferraro, J.L. Jordan, and R. Petrie. 2011. "The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children." *Journal of Public Economics*, 95(11): 1377–1385.

Chetty, R., J.N. Friedman, N. Hilger, E. Saez, S.W. Diane, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics*, 125: 1593–1660.

Conti, G., J.J. Heckman, S. Moon, and R. Pinto. 2014. "The Long-term Health Effects of Early Childhood Interventions." unpublished manuscript.

Cunha, F., J.J. Heckman, and S.M. Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78(3): 883–931.

Dee, T., and M. West. 2011. "The Non-Cognitive Returns to Class Size." *Educational Evaluation and Policy Analysis*, 33: 23–46.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences." *Journal of the European Economic Association*, 9(3): 522–550.

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

Duckworth, A.L., C. Peterson, M.D. Matthews, and D.R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology*, 92: 1087–1101.

Eriksson, T., A. Poulsen, and M.C. Villeval. 2009. "Feedback and Incentives: Experimental Evidence." *Labour Economics*, 16: 679–688.

Eskreis-Winkler, L., E.P. Shulman, S. Beal, and A.L. Duckworth. 2014. "Survivor Mission: Why Those Who Survive Have a Drive to Thrive at Work." *Journal of Positive Psychology*, 9: 209–218.

Fryer, R. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics*, 126(4): 1755–1798.

Gill, D., and V. Prowse. 2012. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review*, 102(1): 469–503.

Gneezy, U., and J. Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics*, 112: 631–645.

Golsteyn, B.H.H., H. Gronqvist, and L. Lindahl. 2013. "Adolescent Time Preferences Predict Lifetime Outcomes." *Economic Journal*, 124: 739–761.

Good, C., J. Aronson, and M. Inzlicht. 2003. "Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat." *Applied Developmental Psychology*, 24: 645–662.

Heckman, J.J., J. Humphries, and N. Mader. 2011. "The GED." *Handbook of the Economics of Education*, 423–484.

Heckman, J.J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour." *Journal of Labor Economics*, 24: 411–482.

Heckman, J.J., R. Pinto, and P. Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *The American Economic Review*, 103(6): 1–35.

Heckman, J.J., S.H. Moon, R. Pinto, P.A. Savelyev, and A.Q. Yavitz. 2010. "The Rate of Return to the High Scope Perry Preschool Program." *Journal of Public Economics*, 94: 114–128.

Heckman, J.J., S. Moon, and R. Pinto. 2010. "The Effects of Early Intervention on Abilities and Social Outcomes: Evidence from the Carolina Abecedarian Study." *Unpublished manuscript, University of Chicago*, 67.

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

Kosse, F., T. Deckers, H. Schildberg-Hoerisch, T. Deckers, and A. Falk. 2015. "Formation of Prosociality: Causal Evidence on the Role of Sociocultural Environment." unpublished manuscript.

Levitt, S., J. List, S. Neckermann, and S. Sadoff. 2016. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." *American Economic Journal: Economic Policy.*

Maddie, S.R., M.D. Matthews, D.R. Kelly, B. Villarreal, and M. White. 2012. "The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets." *Military Psychology*, 24: 19–28.

Maniadis, Z., F. Tufano, and J. List. 2014. "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects." *American Economic Review*, 104: 277–290.

Moffit, T. E., L. Arseneault, D. Belsky, N. Dickson, R.J. Hancox, H. Harrington, R. Houts, R. Poulton, B.W. Roberts, S. Ross, N. R. Sears, W. M. Thomsom, and A. Caspi. 2011. "A Gradient of Childhood Self-control Predicts Health, Wealth, and Public Safety." *Proceedings of the National Academy of Sciences*, 108: 269–398.

Raven, J., J.C. Raven, and J.H. Court. 2004. "Manual for Raven's Progressive Matrices and Vocabulary Scales." *San Antonio, TX: Harcourt Assessment.*

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

Schanzenbach, D. 2006. "What Have Researchers Learned from Project STAR?" *Brookings Papers on Education Policy*, 9: 205–228.

Sutter, M., M. Kocher, D. Rutzler, and S. Trautman. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review*, 103: 510–531.

Tables

Table 1: Design

	A: Sample A			B: Sample B	
	Patience+Grit (15 schools)	Patience (9 schools)	Control (12 schools)	Grit (8 schools)	Control (8 schools)
Baseline Data Collection	Mar ’13	Mar ’13	Mar ’13	May ’15	May ’15
Patience Training	Spring ’13	Fall ’13	-	-	-
Grit Training	Fall ’13	-	-	Fall ’15	-
Follow-up Data Collection	May ’14	May ’14	May ’14	Jan ’16	Jan ’16

Table 2: Mean Comparisons of Pre-Treatment Variables

	A: Sample A		**B: Sample B**	
	Mean [SD]	Difference (p-value)	Mean [SD]	Difference (p-value)
Malleability	0.00	-0.02	0.00	0.00
	[1.00]	(0.90)	[1.00]	(1.00)
Perseverance	0.00	0.00	0.00	-0.15
	[1.00]	(0.97)	[1.00]	(0.22)
Gender (Male=1)	0.52	-0.01	0.43	-0.02
	[0.50]	(0.49)	[0.50]	(0.49)
Age	10.02	0.00	9.45	0.01
	[0.42]	(0.97)	[0.50]	(0.73)
Raven	0.00	-0.03	0.00	-0.26*
	[1.00]	(0.81)	[1.00]	(0.06)
Task Ability	5.07	-0.14	3.88	0.20
	[2.22]	(0.45)	[2.15]	(0.31)
Risk Tolerance	2.52	-0.07	2.18	0.04
	[1.49]	(0.60)	[1.60]	(0.86)
Wealth	2.84	-0.08	2.66	-0.04
	[0.97]	(0.58)	[1.04]	(0.88)
Success in School	3.41	-0.13	3.40	-0.13
	[1.06]	(0.13)	[1.06]	(0.37)
Class Size	37.14	3.94	37.51	5.12
	[8.34]	(0.17)	[7.44]	(0.11)
Math Test Score	0.00	-0.06	0.00	-0.09
	[1.00]	(0.68)	[1.00]	(0.51)
Turkish Test Score	0.00	-0.09	0.00	-0.23**
	[1.00]	(0.63)	[1.00]	(0.04)

* Columns 1 and 3 contain the means of the pre-treatment variables with standard deviations in brackets. Columns 2 and 4 show the estimated difference in means which is obtained from regressing the variable of interest on the treatment dummy. Standard errors are clustered at the school level (unit of randomization) and p-values are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

** The variables malleability and perseverance are extracted factors from questionnaire items in the pre-treatment student survey. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). Task ability refers to the student's performance in the ability measuring round of the experiment. Risk tolerance is elicited using the incentivized Gneezy and Potters (1997) task. The student's wealth and success in school is based on reports by teachers (scale: 1-5). Students' math and Turkish baseline test scores are normalized (mean 0, standard deviation 1). For sample A, these test scores refer to the grades given to the students by their teachers, while for sample B they refer to the students' performance on the standardized tests we administer.

Table 3: Associations in Control Group

	A: Sample A				B: Sample B			
	Math Score	Math Score	Turkish Score	Turkish Score	Math Score	Math Score	Turkish Score	Turkish Score
Raven	0.437***	0.444***	0.436***	0.438***	0.455***	0.457***	0.368***	0.356***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
All difficult visit 1	0.346***		0.152**		0.191		0.199	
	(0.07)		(0.07)		(0.11)		(0.16)	
Plan difficult visit 2		0.246***		0.167**		0.262**		0.193***
		(0.06)		(0.06)		(0.08)		(0.05)
R-squared	0.26	0.25	0.23	0.24	0.19	0.21	0.15	0.15
N	753	746	753	745	610	619	609	596

* All estimates are obtained using least squares and using observations from the control group only. Standard errors are clustered at the school level (unit of randomization) and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
** The outcome variable is the student's math or Turkish test score measured at baseline. The dummy variable "All Difficult Visit 1" equals 1 if the student chooses difficult in all 5 rounds of the first visit and zero otherwise. The dummy variable "Plan Difficult Visit 2" equals 1 if the student chooses to do the difficult task for the second visit.

Table 4: Treatment Effect on Choice of Difficult Task

	Difficult (Round 1)	Difficult (Round 2)	Difficult (Round 3)	Difficult (Round 4)	Difficult (Round 5)	Difficult (All Rounds)
A: Sample A						
Treatment	0.104***	0.093**	0.133***	0.118***	0.083***	0.085***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Control Mean	0.73	0.55	0.43	0.42	0.41	0.25
N	1716	1715	1715	1714	1717	1696
B: Sample B						
Treatment	0.101**	0.151***	0.157***	0.153***	0.142***	0.129***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Control Mean	0.71	0.53	0.37	0.32	0.27	0.17
N	1199	1195	1196	1193	1199	1180

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in columns 1-5 is a dummy variable which equals one if the student chooses to do the difficult task in the respective round. The outcome variable in column 6 is a dummy variable which equals 1 if the student chooses to do the difficult task in all five rounds of the first visit. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Table 5: Treatment Effect on Success and Payoffs

	Success (Round 1)	Payoff (Round 1)	Payoff (Round 2)	Payoff (Round 3)	Payoff (Round 4)	Payoff (Round 5)
A: Sample A						
Treatment	0.039	0.085	0.021	0.032	0.109	0.055
	(0.04)	(0.11)	(0.06)	(0.08)	(0.10)	(0.09)
Control Mean	0.29	1.54	1.06	1.45	1.28	1.34
N	817	1714	1704	1710	1709	1711
B: Sample B						
Treatment	0.048	0.230**	0.050	0.095	0.046	0.094
	(0.03)	(0.10)	(0.08)	(0.08)	(0.05)	(0.08)
Control Mean	0.21	0.79	0.66	1.04	0.93	1.01
N	674	1195	1195	1195	1192	1197

* Reported estimates in column 1 are marginal effects from a logit regression. Reported estimates in columns 2-6 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in column 1 is a dummy variable which equals 1 if the student was successful in meeting the target. Estimates in column 1 are obtained for students for whom the difficult task was imposed. The outcome variable in columns 2-6 is the student's payoff in the respective round. Estimates are obtained for all students. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Table 6: Treatment Effect on Choice of Difficult Task

	Difficult (After Failing in Difficult)	Difficult (Choice for Next Week)
A: Sample A		
Treatment	0.176***	0.143***
	(0.06)	(0.04)
Control Mean	0.36	0.45
N	558	1689
B: Sample B		
Treatment	0.139*	0.191***
	(0.08)	(0.04)
Control Mean	0.51	0.41
N	522	1197

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in column 1 is a dummy which equals 1 if the student chooses to do the difficult task in the 2nd round of the first visit. Estimates are obtained for students who were forced to do the difficult game in round 1 and who failed to meet the target. The outcome variable in column 2 is a dummy which equals 1 if the student chooses to do the difficult task for the following week. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Table 7: Success and Payoffs in the Second Visit

	Success	Payoff			Total Payoff
	Imposed	All	Imposed	Not Imposed	All, 1st+2nd visit
A: Sample A					
Treatment	0.087***	0.301***	0.338**	0.228*	0.362**
	(0.03)	(0.09)	(0.13)	(0.12)	(0.14)
Control Mean	0.50	1.91	2.00	1.76	3.32
N	1004	1736	1004	732	1563
B: Sample B					
Treatment	0.107**	0.462***	0.427**	0.536***	0.554***
	(0.05)	(0.13)	(0.18)	(0.14)	(0.15)
Control Mean	0.46	1.77	1.85	1.60	2.70
N	849	1205	849	356	1110

* Reported estimates in column 1 are marginal effects from logit regressions. Estimates in columns 2-4 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in column 1 is a dummy which equals 1 if the student was successful in meeting the target. The outcome in columns 2-4 is the student's payoff in visit 2. The sample used in the analysis either contains all observations ("All"), the observations for whom the difficult game was imposed ("Imposed") or for whom it was not imposed ("Not Imposed"). The outcome variable in column 5 is the sum of the average payoff in visit 1 and the payoff in visit 2. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Table 8: Treatment Effect on Grades Given by Teacher (Sample A)

	Math Test Score	Turkish Test Score
Treatment	-0.069	0.042
	(0.09)	(0.07)
Gender (Male=1)	-0.013	-0.218***
	(0.03)	(0.03)
Raven Score	0.156***	0.146***
	(0.03)	(0.04)
Risk Tolerance	-0.012	-0.004
	(0.02)	(0.01)
Turkish score (pre)	0.173***	0.356***
	(0.03)	(0.06)
Math score (pre)	0.512***	0.389***
	(0.04)	(0.05)
Class size	0.103	0.013
	(0.11)	(0.13)
Control Mean	0.06	0.01
N	2149	2148

* Estimates are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The dependent variables are the students' math and Turkish test scores at follow-up, which were given by the teacher. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include gender, the Raven score, risk tolerance and baseline scores in math and Turkish. Note that since information on grades is obtained from administrative records, we have observations on all students in the classes, even if they were not present during the experimental task.

Table 9: Treatment Effect on Standardized Test Scores (Sample B)

	Math Test Score	Turkish Test Score
Treatment	0.276***	0.134**
	(0.09)	(0.05)
Gender (Male=1)	0.037	-0.124**
	(0.05)	(0.04)
Raven Score	0.368***	0.230***
	(0.04)	(0.04)
Risk Tolerance	-0.039*	-0.014
	(0.02)	(0.01)
Turkish score (pre)	0.207***	0.249***
	(0.04)	(0.02)
Math score (pre)	0.193***	0.263***
	(0.04)	(0.03)
Class size	-0.009*	-0.006
	(0.00)	(0.00)
Control Mean	-0.04	0.01
N	1203	1206

* Estimates are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The dependent variables are the students' math and Turkish standardized test scores at follow-up. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include gender, the Raven score, risk tolerance and baseline standardized scores in math and Turkish.

Table 10: Treatment Effect on Payoff-Maximizing Choices

	Payoff-Maximizing Choice First Visit	Payoff-Maximizing Choice Second Visit
A: Sample A		
Treatment	0.002	0.091**
	(0.03)	(0.04)
Control Mean	0.65	0.52
N	1704	1578
B: Sample B		
Treatment	0.141***	0.096***
	(0.02)	(0.03)
Control Mean	0.46	0.50
N	1190	1129

* Estimates are marginal effects after logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in both columns is a dummy variable which equals 1 if the student makes a payoff-maximizing choice. To define the dummy variable which indicates whether the student makes a payoff-maximizing choice, we compare the actual task choice to the choice which is payoff-maximizing in expectation. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Table 11: Post-Treatment Survey

	Malleability	Perseverance	Confidence
A: Sample A			
Treatment	0.375***	0.300***	-0.010
	(0.07)	(0.06)	(0.07)
Control Mean	-0.21	-0.17	0.01
N	1690	1612	1675
B: Sample B			
Treatment	0.360***	0.368***	-0.104
	(0.07)	(0.08)	(0.08)
Control Mean	-0.13	-0.14	0.08
N	1155	1036	1181

* Estimates are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variables are malleability, perseverance and confidence which are factors extracted from the post-treatment student survey. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Controls include task ability, gender, the Raven score, baseline beliefs and test scores, and risk tolerance. See Appendix for more details.

Figures

Figure 1: Effect of Treatment on Choosing Difficult Task

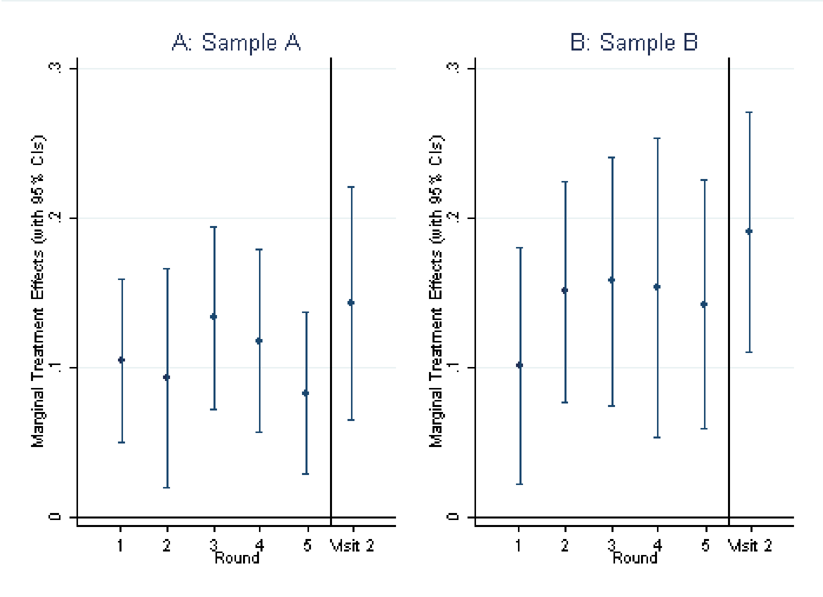


Figure 2: Effect of Treatment on Payoffs

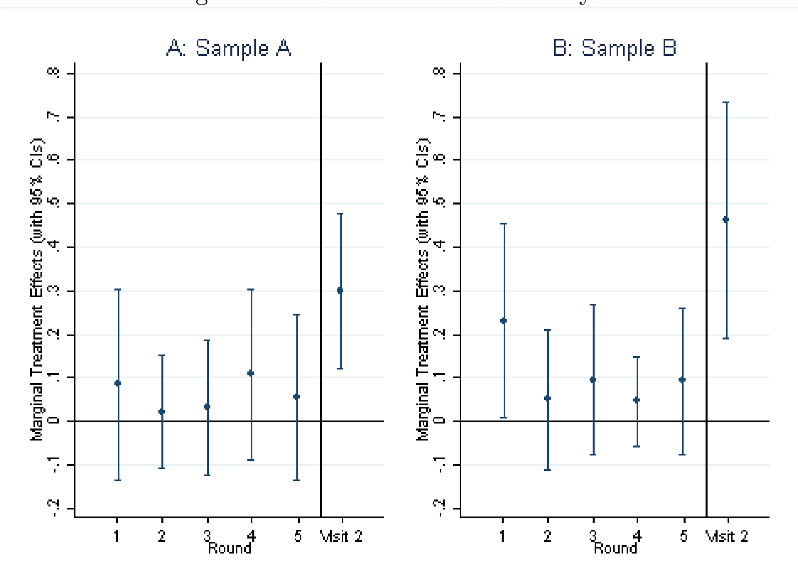


Figure 3: Effect of Treatment on Self-Reported Malleability Beliefs

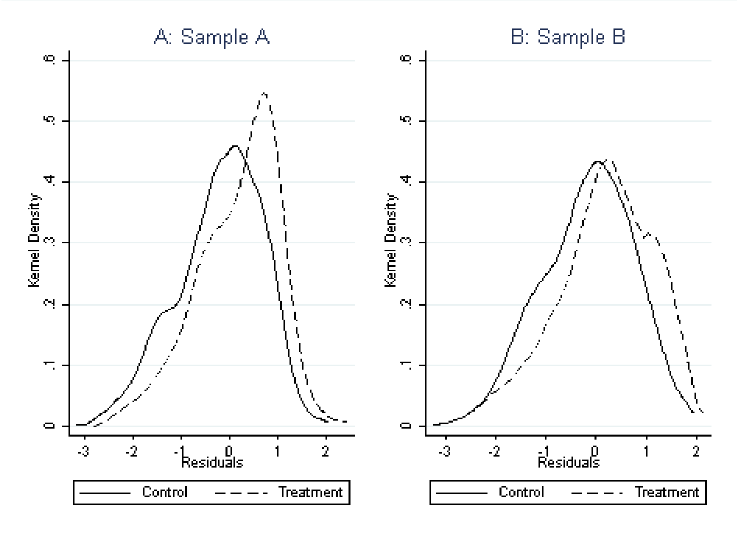
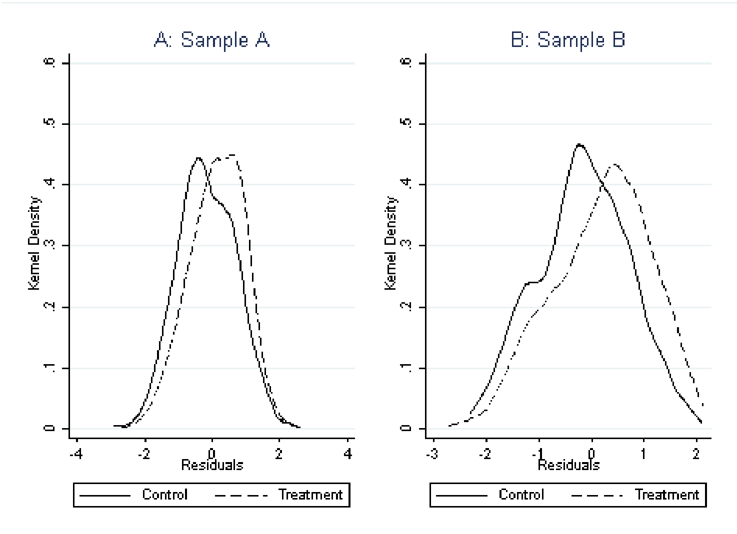


Figure 4: Effect of Treatment on Self-Reported Perseverance



Appendix A: Results Sample A

Table A.1: Treatment Effect on Choice of Difficult Task (Sample A)

	(1) Difficult (Round 1)	(2) Difficult (Round 2)	(3) Difficult (Round 3)	(4) Difficult (Round 4)	(5) Difficult (Round 5)	(6) Difficult (All Rounds)
Treatment	0.104***	0.093**	0.133***	0.118***	0.083***	0.085***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Task Ability	0.047***	0.062***	0.065***	0.069***	0.066***	0.068***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Gender (Male=1)	-0.014	0.088***	0.061***	0.045**	0.080***	0.036**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Raven Score	0.014	0.034**	0.032*	0.043***	0.036**	0.033*
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Malleability (pre)	0.009	-0.006	-0.002	-0.006	0.011	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Perseverance (pre)	-0.012	0.020	-0.000	0.029*	0.012	0.008
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Math score (pre)	0.030*	0.047*	0.059***	0.042**	0.039**	0.043*
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Turkish score (pre)	-0.005	-0.037	-0.029*	-0.009	0.005	-0.021
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance	0.018**	0.012	0.012**	0.012*	0.008	0.017***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Control Mean	0.73	0.55	0.43	0.42	0.41	0.25
N	1716	1715	1715	1714	1717	1696

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variable in columns 1-5 is a dummy variable which equals 1 if the student chooses to do the difficult task in the respective round. The outcome variable in column 6 is a dummy variable which equals 1 if the student chooses to do the difficult task in all five rounds of the first visit. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.
*** Control mean refers to the unconditional mean of the outcome in the control group.

Table A.2: Treatment Effect on Success and Payoffs (Sample A)

	(1) Success (Round 1)	(2) Payoff (Round 1)	(3) Payoff (Round 2)	(4) Payoff (Round 3)	(5) Payoff (Round 4)	(6) Payoff (Round 5)
Treatment	0.039	0.085	0.021	0.032	0.109	0.055
	(0.04)	(0.11)	(0.06)	(0.08)	(0.10)	(0.09)
Task Ability	0.083***	0.301***	0.162***	0.247***	0.191***	0.190***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Gender (Male=1)	0.031	0.202**	0.131*	0.180***	0.099	0.143*
	(0.03)	(0.08)	(0.07)	(0.06)	(0.07)	(0.08)
Raven Score	0.041*	0.132**	0.131***	0.083**	0.079	0.081
	(0.02)	(0.05)	(0.03)	(0.04)	(0.05)	(0.05)
Malleability (pre)	-0.006	-0.024	0.047	0.015	-0.025	0.010
	(0.01)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Perseverance (pre)	-0.006	0.008	-0.036	0.015	0.006	0.063
	(0.02)	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)
Math score (pre)	0.067***	0.265***	0.193***	0.165***	0.147***	0.092
	(0.02)	(0.06)	(0.04)	(0.04)	(0.05)	(0.06)
Turkish score (pre)	-0.000	-0.016	0.010	-0.039	-0.014	0.023
	(0.01)	(0.06)	(0.04)	(0.04)	(0.03)	(0.06)
Risk Tolerance	-0.002	0.024	0.036	0.057***	0.011	0.044*
	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Control Mean	0.29	1.54	1.06	1.45	1.28	1.34
N	817	1714	1704	1710	1709	1711

* Reported estimates in column 1 are marginal effects from a logit regression. Reported estimates in columns 2-6 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in column 1 is a dummy variable which equals 1 if the student was successful in meeting the target. Estimates in column 1 are obtained for students for whom the difficult task was imposed. The outcome variable in columns 2-6 is the student's payoff in the respective round. Estimates are obtained for all students. The regression in column 2 additionally controls for whether the student was forced to do the difficult task in that round. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table A.3: Treatment Effect on Choice of Difficult Task (Sample A)

	(1) Difficult (After Failing in Difficult)	(2) Difficult (Choice for Next Week)
Treatment	0.176***	0.143***
	(0.06)	(0.04)
Task Ability	0.032***	0.055***
	(0.01)	(0.01)
Gender (Male=1)	0.120***	0.012
	(0.04)	(0.02)
Raven Score	0.046***	0.023
	(0.02)	(0.02)
Malleability (pre)	-0.010	0.009
	(0.03)	(0.02)
Perseverance (pre)	0.004	0.044***
	(0.04)	(0.02)
Math score (pre)	0.008	0.041**
	(0.04)	(0.02)
Turkish score (pre)	-0.032	0.005
	(0.04)	(0.02)
Risk Tolerance	0.011	0.013
	(0.02)	(0.01)
Control Mean	0.36	0.45
N	558	1689

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

** The outcome variable in column 1 is a dummy which equals 1 if the student chooses to do the difficult task in the 2nd round of the first visit. Estimates are obtained for students who were forced to do the difficult game in round 1 and who failed to meet the target. The outcome variable in column 2 is a dummy which equals 1 if the student chooses to do the difficult task for the following week. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table A.4: Success and Payoffs in the Second Visit (Sample A)

	Success	Payoff			Total Payoff
	Imposed	All	Imposed	Not Imposed	All, 1st+2nd visit
Treatment	0.087***	0.301***	0.338**	0.228*	0.362**
	(0.03)	(0.09)	(0.13)	(0.12)	(0.14)
Task Ability	0.079***	0.283***	0.297***	0.262***	0.512***
	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Gender (Male=1)	0.026	0.075	0.102	0.031	0.267**
	(0.03)	(0.08)	(0.13)	(0.08)	(0.10)
Raven Score	0.036**	0.183***	0.146**	0.231***	0.274***
	(0.02)	(0.05)	(0.06)	(0.07)	(0.08)
Malleability (pre)	0.019	0.095**	0.076	0.118*	0.089
	(0.02)	(0.04)	(0.07)	(0.06)	(0.06)
Perseverance (pre)	0.006	0.010	0.020	-0.000	0.012
	(0.02)	(0.05)	(0.07)	(0.07)	(0.07)
Math score (pre)	0.072***	0.308***	0.327***	0.273**	0.418***
	(0.02)	(0.05)	(0.07)	(0.10)	(0.08)
Turkish score (pre)	0.011	0.009	0.039	-0.035	0.001
	(0.02)	(0.05)	(0.08)	(0.06)	(0.07)
Risk Tolerance	0.008	0.039*	0.037	0.043	0.068**
	(0.01)	(0.02)	(0.03)	(0.04)	(0.03)
Control Mean	0.50	1.91	2.00	1.76	3.32
N	1004	1736	1004	732	1563

* Reported estimates in column 1 are marginal effects from logit regressions. Estimates in columns 2-4 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in column 1 is a dummy which equals 1 if the student was successful in meeting the target. The outcome in columns 2-4 is the student's payoff in visit 2. The sample used in the analysis either contains all observations ("All"), the observations for whom the difficult game was imposed ("Imposed") or for whom it was not imposed ("Not Imposed"). The outcome variable in column 5 is the sum of the average payoff in visit 1 and the payoff in visit 2. Column 2 additionally controls for whether the student was forced to play the difficult task in the second visit. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table A.5: Treatment Effect on Payoff-Maximizing Choices (Sample A)

	(1) Payoff-Maximizing Choice First Visit	(2) Payoff-Maximizing Choice Second Visit
Treatment	0.002	0.091**
	(0.03)	(0.04)
Task Ability	0.060***	0.034***
	(0.01)	(0.01)
Gender (Male=1)	-0.005	0.002
	(0.02)	(0.02)
Raven Score	0.051***	0.035**
	(0.01)	(0.02)
Malleability (pre)	0.011	0.017
	(0.01)	(0.02)
Perseverance (pre)	-0.019	0.049***
	(0.01)	(0.02)
Math score (pre)	0.114***	0.023
	(0.02)	(0.02)
Turkish score (pre)	-0.048***	-0.031
	(0.02)	(0.02)
Risk Tolerance	0.004	0.014
	(0.01)	(0.01)
Control Mean	0.65	0.52
N	1704	1578

* Estimates are marginal effects after logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$
** The outcome variable in both columns is a dummy variable which equals 1 if the student makes a payoff-maximizing choice. To define the dummy variable which indicates whether the student makes a payoff-maximizing choice, we compare the actual task choice to the choice which is payoff-maximizing in expectation. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The control variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.
*** Control mean refers to the unconditional mean of the outcome in the control group.

Table A.6: Post-Treatment Survey (Sample A)

	(1) Malleability	(2) Perseverance	(3) Confidence
Treatment	0.375***	0.300***	-0.010
	(0.07)	(0.06)	(0.07)
Gender (Male=1)	0.073*	-0.145***	0.005
	(0.04)	(0.03)	(0.05)
Raven Score	0.085**	0.050*	0.015
	(0.04)	(0.02)	(0.03)
Malleability (pre)	0.195***	-0.002	-0.027
	(0.03)	(0.03)	(0.03)
Perseverance (pre)	0.173***	0.354***	0.094**
	(0.03)	(0.03)	(0.04)
Math score (pre)	0.108***	0.138***	0.403***
	(0.04)	(0.04)	(0.05)
Turkish score (pre)	0.136***	0.165***	-0.076*
	(0.04)	(0.04)	(0.04)
Risk Tolerance	-0.007	-0.010	0.016
	(0.01)	(0.01)	(0.02)
Control Mean	-0.21	-0.17	0.01
N	1690	1612	1675

* Estimates are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

** The outcome variables are malleability, perseverance and confidence which are factors extracted from the post-treatment student survey. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The control variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Appendix B: Results Sample B

Table B.1: Treatment Effect on Choice of Difficult Task (Sample B)

	(1) Difficult (Round 1)	(2) Difficult (Round 2)	(3) Difficult (Round 3)	(4) Difficult (Round 4)	(5) Difficult (Round 5)	(6) Difficult (All Rounds)
Treatment	0.101**	0.151***	0.157***	0.153***	0.142***	0.129***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Task Ability	0.047***	0.046***	0.045***	0.047***	0.039***	0.044***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Gender (Male=1)	-0.022	0.050	0.070**	0.108***	0.096***	0.063***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
Raven Score	0.014	-0.001	-0.012	0.001	0.016	0.014
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Malleability (pre)	0.000	0.002	0.003	0.008	-0.007	-0.001
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Perseverance (pre)	0.009	0.007	-0.020	-0.002	0.019	-0.002
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Math score (pre)	0.025**	0.006	0.027	0.023	0.012	-0.001
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Turkish score (pre)	-0.025**	-0.008	0.013	0.011	0.001	0.008
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Risk Tolerance	0.005	0.017***	0.008	0.012	0.029***	0.022**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Control Mean	0.71	0.53	0.37	0.32	0.27	0.17
N	1199	1195	1196	1193	1199	1180

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in columns 1-5 is a dummy variable which equals 1 if the student chooses to do the difficult task in the respective round. The outcome variable in column 6 is a dummy variable which equals 1 if the student chooses to do the difficult task in all five rounds of the first visit. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table B.2: Treatment Effect on Success and Payoffs (Sample B)

	(1) Success (Round 1)	(2) Payoff (Round 1)	(3) Payoff (Round 2)	(4) Payoff (Round 3)	(5) Payoff (Round 4)	(6) Payoff (Round 5)
Treatment	0.048	0.230**	0.050	0.095	0.046	0.094
	(0.03)	(0.10)	(0.08)	(0.08)	(0.05)	(0.08)
Task Ability	0.062***	0.241***	0.124***	0.186***	0.129***	0.168***
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Gender (Male=1)	0.121***	0.402***	0.269***	0.260***	0.110***	0.023
	(0.02)	(0.07)	(0.06)	(0.06)	(0.04)	(0.05)
Raven Score	0.001	0.009	0.086**	0.054	0.103***	0.052
	(0.02)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)
Malleability (pre)	-0.014	-0.016	0.008	-0.004	-0.028	-0.037
	(0.02)	(0.06)	(0.02)	(0.04)	(0.03)	(0.03)
Perseverance (pre)	0.045**	0.139***	0.042	0.065	0.002	-0.009
	(0.02)	(0.04)	(0.04)	(0.05)	(0.03)	(0.05)
Math score (pre)	0.011	0.102**	0.040	0.098**	0.030	0.089**
	(0.02)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Turkish score (pre)	0.043**	0.119*	0.041	0.012	0.066*	0.045
	(0.02)	(0.07)	(0.04)	(0.03)	(0.04)	(0.04)
Risk Tolerance	-0.010	-0.013	-0.020	-0.004	-0.027	0.005
	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Control Mean	0.21	0.79	0.66	1.04	0.93	1.01
N	674	1195	1195	1195	1192	1197

* Reported estimates in column 1 are marginal effects from a logit regression. Reported estimates in columns 2-6 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in column 1 is a dummy variable which equals 1 if the student was successful in meeting the target. Estimates in column 1 are obtained for students for whom the difficult task was imposed. The outcome variable in columns 2-6 is the student's payoff in the respective round. Estimates are obtained for all students. The regression in column 2 additionally controls for whether the student was forced to do the difficult task in that round. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table B.3: Treatment Effect on Choice of Difficult Task (Sample B)

	(1) Difficult (After Failing in Difficult)	(2) Difficult (Choice for Next Week)
Treatment	0.139*	0.191***
	(0.08)	(0.04)
Task Ability	0.030***	0.038***
	(0.01)	(0.01)
Gender (Male=1)	0.017	-0.019
	(0.06)	(0.03)
Raven Score	-0.006	-0.006
	(0.02)	(0.02)
Malleability (pre)	0.038	-0.022
	(0.03)	(0.02)
Perseverance (pre)	-0.001	0.046***
	(0.03)	(0.02)
Math score (pre)	-0.010	0.027
	(0.03)	(0.02)
Turkish score (pre)	-0.015	-0.006
	(0.03)	(0.01)
Risk Tolerance	0.010	0.010
	(0.01)	(0.01)
Control Mean	0.51	0.41
N	522	1197

* Reported estimates are marginal effects from logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in column 1 is a dummy which equals 1 if the student chooses to do the difficult task in the 2nd round of the first visit. Estimates are obtained for students who were forced to do the difficult game in round 1 and who failed to meet the target. The outcome variable in column 2 is a dummy which equals 1 if the student chooses to do the difficult task for the following week. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table B.4: Success and Payoffs in the Second Visit - Difficult Imposed (Sample B)

	Success	Payoff			Total Payoff
	Imposed	All	Imposed	Not Imposed	All, 1st+2nd visit
Treatment	0.107**	0.462***	0.427**	0.536***	0.554***
	(0.05)	(0.13)	(0.18)	(0.14)	(0.15)
Task Ability	0.075***	0.275***	0.296***	0.234***	0.446***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Gender (Male=1)	0.090***	0.317***	0.356***	0.166	0.534***
	(0.03)	(0.07)	(0.10)	(0.13)	(0.08)
Raven Score	0.057*	0.190*	0.215*	0.126	0.231*
	(0.03)	(0.09)	(0.11)	(0.09)	(0.11)
Malleability (pre)	-0.039**	-0.118*	-0.150**	-0.027	-0.146**
	(0.02)	(0.06)	(0.06)	(0.08)	(0.05)
Perseverance (pre)	0.052**	0.227***	0.217*	0.294***	0.319***
	(0.03)	(0.07)	(0.10)	(0.08)	(0.09)
Math score (pre)	0.037**	0.099*	0.150**	-0.034	0.119
	(0.02)	(0.05)	(0.06)	(0.13)	(0.09)
Turkish score (pre)	0.043***	0.115**	0.171**	-0.010	0.147**
	(0.02)	(0.05)	(0.06)	(0.11)	(0.06)
Risk Tolerance	-0.014	-0.033	-0.053	0.065**	-0.036
	(0.01)	(0.04)	(0.05)	(0.02)	(0.06)
Control Mean	0.46	1.77	1.85	1.60	2.70
N	849	1205	849	356	1110

* Reported estimates in column 1 are marginal effects from logit regressions. Estimates in columns 2-4 are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in column 1 is a dummy which equals 1 if the student was successful in meeting the target. The outcome in columns 2-4 is the student's payoff in visit 2. The sample used in the analysis either contains all observations ("All"), the observations for whom the difficult game was imposed ("Imposed") or for whom it was not imposed ("Not Imposed"). The outcome variable in column 5 is the sum of the average payoff in visit 1 and the payoff in visit 2. Column 2 additionally controls for whether the student was forced to play the difficult task in the second visit. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table B.5: Treatment Effect on Payoff-Maximizing Choices (Sample B)

	(1) Payoff-Maximizing Choice First Visit	(2) Payoff-Maximizing Choice Second Visit
Treatment	0.141***	0.096***
	(0.02)	(0.03)
Task Ability	0.034***	0.031***
	(0.01)	(0.01)
Gender (Male=1)	0.156***	-0.046
	(0.03)	(0.04)
Raven Score	-0.024	0.011
	(0.02)	(0.02)
Malleability (pre)	-0.023	0.001
	(0.02)	(0.01)
Perseverance (pre)	0.076***	-0.011
	(0.02)	(0.02)
Math score (pre)	-0.026**	0.007
	(0.01)	(0.02)
Turkish score (pre)	0.063*	-0.016
	(0.03)	(0.02)
Risk Tolerance	0.031***	0.021*
	(0.01)	(0.01)
Control Mean	0.46	0.50
N	1190	1129

* Estimates are marginal effects after logit regressions. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

** The outcome variable in both columns is a dummy variable which equals 1 if the student makes a payoff-maximizing choice. To define the dummy variable which indicates whether the student makes a payoff-maximizing choice, we compare the actual task choice to the choice which is payoff-maximizing in expectation. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The control variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pre-treatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.

*** Control mean refers to the unconditional mean of the outcome in the control group.

Table B.6: Post-Treatment Survey (Sample B)

	(1) Malleability	(2) Perseverance	(3) Confidence
Treatment	0.360***	0.368***	-0.104
	(0.07)	(0.08)	(0.08)
Gender (Male=1)	0.060	-0.114**	-0.019
	(0.06)	(0.05)	(0.05)
Raven Score	0.170***	0.140***	0.032
	(0.03)	(0.04)	(0.03)
Malleability (pre)	0.123***	0.133***	-0.120***
	(0.04)	(0.04)	(0.03)
Perseverance (pre)	0.140***	0.167***	0.298***
	(0.03)	(0.05)	(0.05)
Math score (pre)	0.001	0.089***	0.138***
	(0.03)	(0.03)	(0.04)
Turkish score (pre)	0.134***	0.155***	0.045
	(0.04)	(0.05)	(0.03)
Risk Tolerance	-0.043**	0.017	-0.022
	(0.02)	(0.01)	(0.02)
Control Mean	-0.13	-0.14	0.08
N	1155	1036	1181

* Estimates are obtained using least squares. Standard errors are clustered at the school level (unit of randomization) and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01
** The outcome variables are malleability, perseverance and confidence which are factors extracted from the post-treatment student survey. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. Task ability refers to the student's performance in the ability measuring round of the experiment. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). The control variables malleability and perseverance are extracted factors from the pre-treatment student survey. Math and Turkish scores are measured pretreatment. Risk tolerance was elicited using the incentivized Gneezy and Potters (1997) task.
*** Control mean refers to the unconditional mean of the outcome in the control group.

Appendix C: Educational Material and Tasks

Figure 1: Screenshot of Educational Video Material

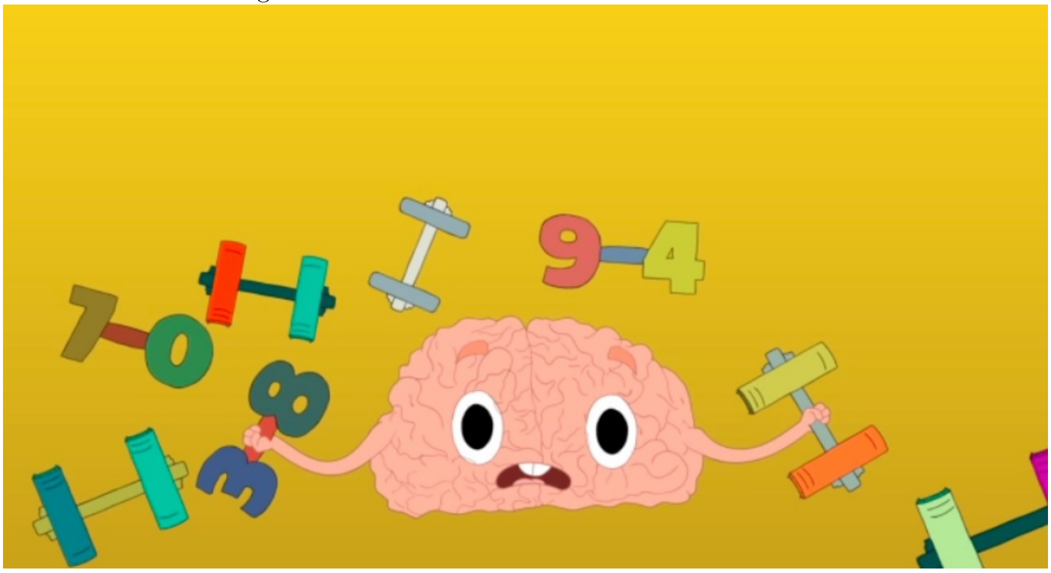


Figure 2: Example of Easy Task

80	7
70	95
5	20
10	30
93	90

Figure 3: Example of Difficult Task

17	86	23	12
71	42	27	38
51	62	83	30
77	59	46	67
81	58	29	54