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

Sule Alan, University of Essex

Teodora Boneva, University of Cambridge

Seda Ertac, Koc University

April 2015

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 an evaluation of a randomized educational intervention implemented in elementary schools in Istanbul. Outcomes are measured via a novel incentivized real effort task and actual school grades on core subjects. 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 significant impact on school grades: We find that treated students are about 3 percentage points more likely to receive top grades in core academic subjects.

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 ESRC Research Centre on Micro-social Change (MISOC), whom we would like to thank for generous financial support. We are grateful to Orazio Attanasio, Thomas Crossley, James Heckman, conference participants at the Experimental Methods in Policy Conference and seminar participants at the IFS and the University of Essex for helpful comments. We would also like to thank Elif Kubilay, Nergis Zaim, Banu Donmez 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 et al., 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 Almlund et al. (2011), Sutter et al. (2013), Moffit et al. (2011), Castillo et al. (2011), Golsteyn et al. (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 to being able to set and work towards long-term goals. Hence, it is a non-cognitive skill that influences the motivation to set a goal, exert effort towards a goal and persevere in pursuing that goal 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

[1]Well-known examples of early childhood and elementary school programs include the Perry Preschool program (Heckman et al., 2010, Heckman et al. 2013), the Abecedarian Program (Heckman et al., 2014, Conti et al., 2013), and the project STAR (Dee and West, 2008, Chetty et al., 2010). 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 favourable 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. Grit also significantly 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) and Eskreis-Winkler et al. (2014).

individuals to work harder, it is important to understand the nature of grit as a non-cognitive skill and to explore ways of enhancing it.

In determining "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 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 large-scale randomized educational intervention in elementary schools in Istanbul. 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 is based on exposing 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 exert effort and reduce fears of failure in challenging tasks, 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 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 material 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 a large sample of 4th grade students (ages 9-10), in schools that have at least one teacher willing to participate in the program and are randomly assigned into treatment and control groups.[4] We measure the outcomes through a multi-faceted methodology that includes a novel incentivized experiment, pre- and post-treatment questionnaires, and actual school grades. The experiment is designed to elicit goal setting, perseverance and skill accumulation in an incentivized way, using a mathematical real effort task. 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 have access to actual school outcomes measured by official grades on core academic subjects. We also measure students' beliefs about the malleability of ability and the role of effort in achievement, and self-reported attitudes and behaviors such as perseverance and goal orientation, using pre- and post-treatment questionnaires.

Our results reveal a striking impact of the intervention in terms of behavior and outcomes in the experimental task. In particular, we find that treated students are about 10 percentage points more likely to opt for the difficult high-reward task when offered a choice against the easier low-reward alternative. Furthermore, when they receive negative feedback in the difficult task, they are 18 percentage points more likely than students in the control group to persevere and attempt the difficult task again. When they are given the chance to do the task again in the following week, treated students are 14 percentage points more likely to set the goal of succeeding in the difficult task. We find that in the following week, treated students are indeed 9 percentage points more likely to succeed and eventually collect 16% higher payoffs than students in the control group. These findings suggest that treated students are more likely to engage in ability accumulation exercises in the week between visits. These positive effects also extend to outcomes outside of the experiment: while we do not observe a significant improvement in average school grades, we estimate that treated students are about 3 percentage points more likely to receive top grades in overall core academic subjects - mathematics, social and life sciences, and Turkish. Results from the questionnaire hint that students' more positive

[3] Blackwell et al. (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 mindset and achievement is provided by Aronson et al. (2002) and Good et al. (2003).

[4] Details of the phase-in evaluation design are given in Section 2.

beliefs about the malleability of ability and the productivity of effort are likely to have been a major factor leading to more ambitious goal-setting, more perseverance and more skill accumulation in the treatment group.

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 might have been achieved. With its dynamic real-effort framework, the paper contributes to the recent experimental economics literature on the motivational effects of performance feedback and dquitting behavior in work settings (Azmat and Iriberri, 2010, Eriksson et al., 2009, Barankay, 2014, Gill et al., 2013). The results also relate to the literature on the link between beliefs and investments for human capital accumulation (e.g. Cunha et al., 2013).

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 a mathematical task, as well as actual school outcomes. Given the pivotal role of these 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 Measurement

2.1 Nature of the Intervention

Our educational intervention involves providing 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 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 these 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 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.[5]

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 there is no scope for enhancing it 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]

[5]See the appendix for an example.

[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.

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. 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. Figure 1 shows the responses of teachers in the treatment group to the question "how much do you agree with the ideas conveyed in the training?", while Figure 2 shows the responses to the question "how intensely did you implement the training material and suggested activities?". 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.[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, 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 World Bank and the Regional Development 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

[7] 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).

non-cognitive skills in elementary school children in the classroom environment, under the supervision of their own teachers. The program is implemented 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.

In its full-fledged form, the program involves three major treatment arms, implemented using a version of a phase-in randomized-controlled design, whereby groups of treatment schools receive training in different semesters.[8] The first arm is designed as an intervention on time preference, for which students receive materials that highlight the importance of forward-looking behavior. The treatment group that is the focus of this paper (IT henceforth, for initial treatment) received this training in Spring 2013 for eight weeks.[9] The same students moved on to 4th grade and, in Fall 2013, they received training on grit for the entire semester, which constitutes the second intervention arm. Meanwhile, another group of students (CT henceforth, for control-then-treatment), received no training in Spring 2013 but received the training on forward-looking behavior in Fall 2013. This group never received the training on grit, which helps us isolate the potential confounding effects coming from the forward-looking behavior training in our main treatment arm. We will revisit this issue in Section 3 and show that this training does not have an independent impact on any of our outcome measures. Finally, a group of schools received neither of the training materials during our measurement phases. This group serves as the "pure control" group (PC henceforth).[10]

After official documents were sent to all elementary schools in designated districts of Istanbul by the Istanbul Directorate of Education, 3rd grade teachers in these schools were contacted in random sequence and were offered to participate in the program. Teachers were informed that upon participation they would be assigned to different training phases within the coming two academic years.[11] Once a teacher stated a willingness to participate, we assigned their school into one of the three arms of the intervention. A willing teacher had a 40 percent chance of being in the initial treatment group

[8]Figure 3 shows the complete evaluation design including the teacher training, training implementation, and measurement phases.

[9]See Alan and Ertac (2014) for the evaluation of this arm.

[10]All involved teachers 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 the pure control group will never receive the training as they move on to middle school after year 4, their teachers will, albeit at a later time.

[11]The general intervention program was titled "financial literacy, saving and economic decisions" and no further information on the particulars of the program was disclosed.

IT, a 30 percent chance of being in the control-then-treatment group CT, and a 30 percent chance of being in the pure control group PC. It should be noted that random assignment was done at the school level and not at the classroom level, since the physical proximity of classrooms and teachers would have been likely to generate significant spillover effects. With this procedure, we finished up with 15 schools in IT, 10 in CT and 12 in PC, totaling 37 clusters (schools).[12],[13]

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, 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%). Given these numbers, we conjecture that the external validity of our results is strong. We will revisit this issue when we discuss our results.

After completing the randomization and before launching training seminars for teachers in the initial treatment group (IT), data on a large set of variables were collected from the entire sample through student and teacher surveys to obtain baseline data (Phase 1).[14] After the training seminars for teachers were completed, the first arm of the program (training on forward-looking behavior) was implemented in Spring 2013 by the teachers in the IT group, and in Fall 2013 by the teachers in the CT group. The training on grit was given in the Fall 2013 semester, and we collected our experimental outcome measures related to the grit intervention in May 2014, approximately 4 months after the implementation of the grit training, and before the third arm of the intervention was implemented.[15]

[12]We lost one school in the CT group after the Spring 2013 intervention due to the participating teacher being transferred to another school outside of Istanbul, giving us 36 schools in total.

[13]In practice, we filled the IT, CT and PC bins using a random number generator. We stopped calls after reaching 37 schools due to feasibility concerns in terms of physically visiting classrooms to collect experimental data and the constraints imposed by the academic calendar. 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.

[14]All measurements are conducted with the approval of the local IRB and the permission of the Ministry.

[15]This third arm was a 2-week intervention that involved showing children a video of 6 successful Turkish female role models in male-dominated occupations, and discussing aspirations in the classroom with an emphasis on girls. The women in the video talk about their childhood aspirations and how they overcame the difficulties they faced on the way to success. This arm was implemented by stratified sampling from each of three treatment groups (IT, CT, and PC). Although our experimental outcome data were collected prior to this intervention, official grades were given at the end of June 2014, preventing us from using final grades as outcome measures. However, we can still use the marks given for the spring examinations implemented prior to this intervention, which is what we do.

2.3 Experimental Task

Our design consists of two different visits, a week apart from each other. In the first visit, children go through five rounds of a mathematical real effort task. 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. 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.[16] 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.[17] 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.[18] 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

[16] See appendix for examples of the two types of task.

[17] 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.

[18] 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.

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. Before they make their choices for the first round, students are told that the choice they are about to make for the first round has a 50% chance of "counting". Specifically, there are two folded pieces of paper in a small bag, with one that says "free choice" and another that says "no choice". If "free choice" is drawn, students work on the type of game of their choice in the first round.[19] If the "no-choice" paper is drawn, all students in the classroom play the difficult task. 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.[20] 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.[21]

In the second visit, children perform the task that they chose in the previous week or the difficult task, if that was imposed in their classroom. 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.[22]

[19] One child comes forward to make the selection in front of the class.

[20] All children are given this booklet.

[21] 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.

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

3 Data and Baseline Information

Our sample contains information on 64 classrooms in 36 schools (clusters), and includes about 2,150 students, more than 1,700 of whom were present during our experimental task. Among the schools in our sample, 15 are in the treatment group and received the training on grit. In a preliminary analysis we compare students in schools that received the forward-looking behavior training to those in pure control schools, and we show that forward-looking behavior training has no independent effect on any of our outcome measures.[23] We therefore carry out our empirical analyses by pooling the 21 schools whose students did not receive grit training, and consider them as the control group.

Our baseline data contain a rich set of variables on key student characteristics reported by teachers such as social behavior, overall teacher-assessed academic success and the socio-economic status of the student's family. We also have access to the students' pre-intervention grades on three core subjects: mathematics, life and social sciences, and Turkish. In addition to these measures, we have self-reported questionnaire data on the students' (i) baseline beliefs about the malleability of intelligence, (ii) attitudes and behaviors related to grit and perseverance, and (iii) self-confidence. Finally, we elicit measures of students' baseline cognitive function and risk attitudes. The former is collected using a Raven's progressive matrices test (Raven et al., 2004), and the latter using a version of the Gneezy and Potters (1997) risk preference elicitation task. We use these baseline measures to assess the sample's balance across treatment status. We also use a number of baseline variables that are highly predictive of our outcome variables as covariates in estimating the average treatment effects.

Table 1 presents the balance tests. Column 1 reports the overall mean in the sample, while columns 2 and 3 report the mean characteristics of the treatment and the control group, respectively. Column 4 shows the estimated difference between the treatment and the control group, which is obtained by running a univariate regression of the variable of interest on treatment status. Standard errors are clustered at the school level and p-values are reported in parentheses. We do not observe any statistically significant differences in key student characteristics such as socio-economic status, age, gender, the Raven test score and risk tolerance. More importantly, we also do not find any significant differences in students' baseline beliefs about the malleability of ability, their attitudes and behaviors related to grit and perseverance, or their self-confidence. Finally, we do not find any statistically significant differences in average core grades or the probability of receiving a top grade (defined as

[23] All related analyses are available upon request and are also available in the online appendix posted at https://sites.google.com/site/salancrossley/ Note also that this design does not allow us to test potential complementarities between the two types of training.

scoring 90 or above) in any of the subjects. We note that we use this particular variable, measured post-treatment, as our outcome variable. Together, these results suggest that the sample is balanced in terms of pre-treatment characteristics and that the randomization was successful.

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. As presented in Table 1, the number of pairs found in this task is not different across treatment status (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 function and math grades greatly improve the precision of our estimates.

We first investigate whether decisions in our experimental task and self-reported attitudes and behaviors elicited in the survey are predictive of real outcomes. Using observations from the control group only, Table 2 shows that behavior in the experimental task significantly predicts test scores in school, over and above what can be predicted by cognitive ability.[24] In particular, students who choose the difficult task in all five rounds of the first visit have significantly higher math scores (column 1). Whether or not the student chooses the difficult task for the second visit similarly predicts significantly higher math scores (column 2). Looking at the predictive power of the self-reported attitudes and behaviors, which were elicited before the intervention, we find that students who are more inclined to believe that ability is malleable and can be enhanced through effort are more likely to obtain higher math scores (column 3). The same is true for students who have rated themselves as being more perseverant (column 4). The estimated associations are considerable in size: The magnitudes of the point estimates in column 4 show that a one standard deviation increase in the Raven score predicts an increase in the math score by 6.8 points, while a one standard deviation increase in the perseverance score predicts an increase in the math test score by 3.8 points (about 55% of the effect size of the Raven test score). These results suggest that our measures, which capture at least some aspects of non-cognitive skills, have predictive power for important outcomes of interest, over and above cognitive skills.

[24]Note that grades are measured on a 0 to 100 scale.

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 by 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 function, mathematics grades prior to the intervention, risk tolerance measured via an incentivized task based on Gneezy and Potters (1997), gender, and two factors extracted from the baseline survey which capture (i) students' beliefs about the malleability of intelligence 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. In the case of payoffs, the above equation is estimated via ordinary least squares.

In order to test the null hypothesis that the program made no impact on the real outcome y^R, we estimate the average treatment effect using the same specification but using a different set of covariates. In particular, when estimating the average treatment effects on grades we control for gender, the Raven test score, class size and all core subject grades prior to the intervention.[25]

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

We now assess the effect of treatment on the students' choices and outcomes in the fully incentivized experimental task. We begin by noting that there were a total of 123 students who were reported to have difficulties in understanding the task procedures and for whom we do not have valid entries. Our

[25] 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, all our results carry through even when we collapse our data into school means and estimate treatment effects using only 36 observations. All related analyses are available upon request and are also available in the online appendix posted at https://sites.google.com/site/salancrossley/

analysis shows that there is no statistical difference in the proportion of such students across treatment status (p-value=0.72). Therefore, we drop these observations from our analysis.

In every round, students decide whether to do the difficult more rewarding task or the easy less rewarding task. The binary outcome variable "success" is defined as finding three correct pairs or more. The categorical "payoff" variable takes the value 0 if the target of finding three pairs is not met, 1 if the easy game is chosen and the target is met, and 4 if the difficult game is chosen and the target is met.

Table 3 presents the marginal effects from a logit regression where the dependent variable is a dummy which takes the value 1 if the student chooses to play the difficult game and zero otherwise. As we see in the first row of the table, treated students are more likely to choose to play the difficult game in all five rounds. For example, in the first round, treated students are about 10 percentage points more likely to choose the difficult game (proportion of students in the control group who choose the difficult game in the first round is 73%). Looking at the last column, we see that treated students are about 9 percentage points more likely to consistently attempt the difficult game in all five rounds (baseline value for attempting the difficult game in all five rounds is 25%). What is striking about these results is the persistence of the treatment effect through all five rounds; while the number of students who attempt to play the difficult game falls toward the final round, the positive and significant treatment effect persists. Panel A of Figure 4 shows in visual clarity the estimated marginal treatment effects for the choice of the difficult game with 95% confidence intervals for all five rounds.

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 provided in the first visit, which makes the incentive to switch to the easy game stronger. We further discuss this later in Section 5.2 where we propose a potential mechanism of behavioral change.

The next question is whether there is a treatment effect on outcomes in the first visit. Table 4 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 (i) estimate the treatment effect on success in the difficult game, and (ii) estimate the impact of feedback received in this round on subsequent rounds' choices without selection. We will come back to the latter result in the following section. Table 4 shows that the probability of success in the difficult game is not significantly different across treatment groups. 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 (see also Panel A of Figure 5).

These results show that despite the fact that a higher proportion of students in the treatment group chose to do the difficult task and despite the fact that the probability of success is much lower for this task, treated students did not end up with lower payoffs on average. Looking at the distribution of overall payoffs in the first visit we see that the distributions are not statistically different from each other.[26]

The Effect of Negative Feedback

A constructive response to negative feedback is perhaps one of the most important skills that is 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 ability, and 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 voluntarily 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 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 the way in which students respond to negative feedback, after collecting choices regarding the first round, we impose the difficult game on all students in randomly selected classes. 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. Recall that we find that the estimated treatment effect on success is not statistically different from zero when the difficult game was imposed in round 1 (see Table 4 column 1).

Table 5 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

[26] To test the equality of distributions we use a two-sample Kolmogorov-Smirnov test and compare the residual payoff distributions of treated and control group students; p-value=0.19.

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. Treated students are about 18 percentage points more likely to choose to play the difficult game in the second round after failing in the first round's difficult game. 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. Note also that male students and, not surprisingly, students with higher initial ability are more likely to respond to the failure feedback by re-attempting the difficult game.

Goal Setting and Commitment to Ability Development

As explained in Section 2, 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 very same game for one additional round. We also inform them that we have prepared a study booklet covering numerous examples of the game and tell them that if they want they can study/practice over the week using this booklet. We emphasize that this is entirely voluntary but most students decide to take the study booklet. We then collect their decision on which game they would like to play in the following week. After we collect these choices, all students in each class are informed whether they will all 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 allowed 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 and therefore to commit to playing the difficult game, and are more likely to study and increase their chances to succeed. This is exactly what we see in Table 5 column 2: treated students are estimated to be 14 percentage points more likely to plan to play the difficult game in the following week (baseline proportion is 45%).

Treated students did in fact take up the challenge of committing themselves to the difficult game - but did they actually study over the week? This is the question we explore in the next subsection.

4.3 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. It could be too optimistic to think that over rounds of only 1.5 minutes a student can improve her ability in a task via repeated trials. 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 those 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 this effort would lead to a higher probability of success in the second visit.

Table 6 presents the estimated treatment effects on outcomes of the second visit. The first column presents the treatment effects on success obtained from the sample that is forced to play the difficult game, 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 game was imposed in the class or not. Looking at the first column, we see that treated students are about 9 percentage points more likely to succeed in the difficult game (baseline proportion is 50%). This effect is statistically significant at the 1% level. The increased success rate is reflected in payoffs: we estimate a statistically significant 16% treatment effect on payoffs for the overall sample (baseline average payoff is 1.91). The estimated effects are similar for the imposed and unimposed samples (17% and 13% respectively). Here, the Kolmogorov-Smirnov test reveals that the distribution of the mean deviated payoffs is not different across treatment groups, suggesting only a mean shift (p-value=0.37, Figure 7). Considering the combined payoffs of both visits, we estimate 11% higher payoffs for the treated students (column 5, Table 6).

Table 7 shows suggestive evidence that the strong impact we estimate in the second week for the imposed sample is mainly driven by students who chose to play the easy game. For this group, the estimated impact on the success rate is about 11 percentage points, while the effect on payoffs is about 28%. This result suggests that the treatment was especially successful in motivating students who would have otherwise chosen not to study. Not surprisingly, we do not detect this effect for students who already planned to play the difficult game, i.e. those who are already motivated enough to study (see columns 3 and 4).

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, and since the performance technology is open to ability accumulation, we might observe success in the second week for these students. Our analyses do not reveal any significant heterogeneity in treatment effects with respect to gender, task ability and cognitive function.[27]

4.4 Treatment Effect on Real Outcomes: School Grades

The implication of the change in beliefs regarding the malleability of intelligence is 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 target 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, we collect official grades on core subjects - mathematics, life and social sciences, and Turkish - that reflect students' performance in Spring 2014, several months after the implementation of the training.

Table 8 reveals that the treatment did not significantly affect the level of grades in each core subject and the level of overall core grades (defined as the average of the three subjects). Only when we focus on the top grades, defined as a score of 90 and above, we see truly striking results: the treated students are estimated to be 3.2 percentage points more likely to receive top grades in overall core areas, and this effect is statistically significant at the 5% level (Table 9, Figure 6). This is an effect of considerable size, given that the proportion of control group students with top grades is only 12%. We also observe that this effect is consistent through all three core subjects, i.e. the estimated coefficients are positive although they do not reach significance.

These results are encouraging. For one thing, if such an intervention is able to push several students to the top level (that is, from 4 to 5 in Turkish grading scale of 1 to 5) via changing their mindset about the role of effort in achievement, a similar but longer, more intense intervention has a lot of scope. Secondly, the aspirations of students who experienced high achievement through effort may change significantly and these changes may be permanent by creating a productive cycle of further

[27] Full results on heterogeneity are available upon request and also available in the online appendix posted at https://sites.google.com/site/salancrossley/

effort and further success. The latter is particularly important from the policy perspective for areas in the world where parents are generally not engaged in the skill formation of their children for reasons such as illiteracy and poverty, and where as a result achievement rates are dismal.

5 Discussion

5.1 Optimality

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.[28] 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 to the student's actual task choice. This allows us to categorize individuals into three different groups: (i) students whose choices were optimal in expectation (from the perspective of the econometrician), (ii) students who chose the difficult task although in expectation the easy task would have yielded higher payoffs, and (iii) students who chose the easy task although in expectation the difficult task would have yielded higher payoffs. We can do this for the first round in the first visit and estimate the treatment effect on over/under choosing the difficult task, since we can estimate the success rate in the imposed sample. Similarly, we can do the same analysis for the second visit using the plans made at the end of the first visit and success rates in the imposed sample and estimate the

[28] 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, the Raven score, pre-treatment beliefs, the math test core 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.

treatment effect on over/under investment in the difficult task in between visits.

Table 10 presents the results. For the first round of the first visit, the results suggest that treated students are no more likely to choose the payoff maximizing task compared to students in the control group (baseline mean is 65%). They are also no more likely to over- or under-choose the difficult task in this round. When we turn to the choices for the second visit, we see a remarkable difference. In particular, treated students are 9.2 percentage points more likely to choose the payoff-maximizing task (baseline mean is 52%). While there is no effect of treatment on over-investment (baseline mean is 3%), we can see that treated students are 9.8 percentage points less likely to under-invest (baseline mean is 45%).

Why do we observe striking differences across the first and the second visit? A potential mechanism which could be at play is that the opportunities for skill accumulation differ between the two visits. Having the belief that ability can be increased through goal-oriented effort will pay off when there is sufficient time for students to engage in skill accumulation activities. While choosing the challenging task might not have been better for treated students in terms of payoffs in the short run, it becomes the payoff maximizing choice when sufficient time is given to practice and accumulate skill.[29] We discuss a theoretical model that is based on this insight below.

5.2 A Potential Mechanism

Taken together, the findings indicate that the intervention was remarkably successful in changing children's behaviors and outcomes in the experimental task. Guided by our empirical results, we now contemplate a potential mechanism through which these changes might have been achieved. For this, we use a production function that maps ability and on-the-task effort into the number of pairs found (output), and an underlying ability accumulation technology that can help us characterize behavior in-between the two visits. With this model, we can better interpret our findings and, more importantly, we will have a theoretical framework we can use to think about the impact of the intervention on performance outcomes outside the experimental task, such as school grades.

Suppose that student i has a "true" production function which takes the standard CES form. For the sake of brevity assume that the output q_k, which is the number of pairs found in the difficult game

[29] A potential caveat to note here is 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.

in a given round k, is a function of two main inputs, ability and effort:[30]

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

where $q^1_{i,k}$ is the number of pairs student i finds in round k in visit 1, A is the productivity parameter, $a_{1,i,k}$ is task ability in visit 1 and round k, $E_{i,k}$ is the effort exerted during 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 first that the intervention has likely increased the perceived marginal product of effort E such that, using the potential outcomes framework, the perceived role of effort is higher for student i in round k if she is treated:

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

where the subscript d denotes the treatment status which equals 1 if student i is treated and zero otherwise. Everything else equal, this condition alone will lead to treated students being more likely to choose the difficult game in the first round if the student makes payoff-maximizing choices.[31]

How the propensity to choose the difficult task changes after performance feedback will depend on how the students update their perceived production function. Empirical findings of note here are that (i) the estimated treatment effects on the choice of difficult task are significant in all rounds but the point estimates decline visibly as rounds progress (see Figure 4), and (ii) a decline in the propensity to choose the difficult task after the first round is observed in both the treatment and control group. These findings suggest that the proportion of students attempting the difficult task approaches the empirical rate of success over time, which is consistent with a rational updating mechanism.

Such updating is based on revising down the belief about the marginal productivity of effort, which in turn can involve updating the role-of-effort parameter β and/or actual ability, a. When a

[30] 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%.

[31] A student will choose to play the difficult game if and only if $Pr(Success_{Diff})\pi^{Diff} > Pr(Success_{Easy})\pi^{Easy}$, where $Pr(Success) = Pr(q \geq \bar{q})$, π denotes payoff and $\bar{q}$ denotes the required threshold.

student is faced with a new task, it is likely that she is not fully informed about her ability level in this task. Recall that the estimated treatment effect on the success rate in the first round (in the imposed sample) of the first visit is not different from zero. This, and the lack of a difference on the ability-measuring task suggest that there was no task-specific ability difference across treatment status ex-ante. The decline in the proportion of students choosing the difficult task is consistent with students of both treatment status revising their perceived ability downward (toward its true level) as they receive negative performance feedback. This could be complemented, especially in treated children, with updates of the role of effort in the production process, since treated kids are likely to initially overestimate the role of on-the-task effort in production in the first visit where they have very little time in each round.[32] Treated children revising this parameter down after negative feedback leads to a decline in the propensity of choosing the difficult task, and a reduction in the treatment effect therein, as rounds progress.

After the first visit, with an opportunity to accumulate skills in-between visits, students decide whether to engage in ability investment. For this process, we assume a CES ability accumulation technology similar in spirit to Cunha, Heckman and Schennach (2010)

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

where $a_{1,i}$ denotes student i's ability in the first visit, $a_{2,i}$ denotes student i's ability in the second visit, and I denotes the investment made by the student to develop ability that is specific to this task. Here, the intervention most likely influenced the *perceived* marginal product of this investment. Given the results we presented above on optimality, which suggest that there is significant under-investment by control group students and no significant over-investment by treated students, the following is likely to be true (potential outcomes for student i)

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

This would result in treated students setting more ambitious goals for the following visit by planning to play the difficult game, and making higher investments to enhance ability. Consequently, re-writing

[32]Note that the perceived β may be higher than its true value for an untreated student as well. But as long as the condition $\beta_{i,k,d=1} \geq \beta_{i,k,d=0}$ is true, we will observe a positive treatment effect on the choice of the difficult task.

the true production function of the task in the second week as

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

student i is more likely to succeed if treated,

$$Pr(q_i^2 \geq \bar{q})_{d=1} > Pr(q_i^2 \geq \bar{q})_{d=0},$$

since, everything else equal,

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

The above mechanism may again work via changes in beliefs with respect to the role of effort in success and ability accumulation. The belief that on-the-task effort (β) is a very important input for success would generate persistence in the choice of the difficult game in the first visit, and the belief that ability can be developed through effortful investment (λ) would lead to children studying and eventually achieving a higher likelihood of success, since the experimental task is conducive to ability development. Overall, changes in the perceived role of effort in production emerge as a likely mechanism that can explain both the first-visit and the second-visit results.[33]

More support for this particular mechanism comes from our survey data. Table 11 presents the estimated treatment effects on the factors measuring the students' beliefs about the malleability of ability and the students' self-reported levels of perseverance. Treated students have significantly higher belief and perseverance scores controlling for associated baseline scores and some key student characteristics such as gender, cognitive function and school grades.[34] The estimated increases in the scores of malleability of ability and self-reported perseverance 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 skill and consequently be more likely to succeed in meeting the target. Such an increase in the students' beliefs about the malleability of ability is also

[33]Other potential changes in the perceived production function as a result of treatment might have been through lowering the importance of ability (α) or increasing the 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. Children who believe that ability is not that important would be less willing to invest in costly skill accumulation. Similarly, treated students believing more in substitutability of ability with on-the-task effort would not invest in ability development for the second visit, and would not be more likely to actually succeed.

[34]See Figures 8 and 9 for the pre-treatment and post-treatment distributions of the residual factor scores.

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.[35]

6 Conclusion

We evaluate a large-scale randomized educational intervention that aims to enhance grit in the classroom environment using outcome measures we obtained from a real effort task and actual school grades. We find that treated students are significantly more likely to engage in skill accumulation exercises and consequently, more likely to succeed and collect higher payoffs in the experimental setting. The intervention also has a positive effect on school grades in core academic subjects.

From the policy perspective, the paper contributes to the on-going debate about the malleability of non-cognitive skills and the role of educational programs in enhancing individual achievement through interventions which 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 input face challenges to engage families of low socio-economic strata.

A potential caveat relates to external validity, since program participation was voluntary. However, as mentioned in Section 2, the majority of teachers who did not participate in our program expressed constraints rather than a lack of willingness to participate. Only about 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

[35]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 intelligence. 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, which suggests that a boost in self-confidence is unlikely to be the mechanism through which the estimated results were obtained. See Figure 10.

will change over time as students age and 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 sample and new cohorts of students who will be taught by the teachers in our sample.

References

[1] Alan, S. and Ertac, S. (2014), Good Things Come to Those Who (Are Taught How to) Wait: Results from a Randomized Educational Intervention on Time Preference, Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2566405

[2] Almlund, M., Duckworth, A.L., Heckman, J.J. and Kautz, T.D. (2011), "Personality Psychology and Economics", *Handbook of the Economics of Education*, pp. 1-181

[3] Aronson, J, Fried, C.B. and Good, C. (2002), "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence", *Journal of Experimental Social Psychology*, 38, pp. 113-125

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

[5] Barankay, I. (2014), Incentives and Social Tournaments, unpublished manuscript

[6] Blackwell, L., Trzesniewski, K.H. and Dweck, C.S. (2007), "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention", *Child Development*, 78(1), pp. 246-263

[7] Borghans, L., Duckworth, A.L., Heckman, J.J. and ter Weel, B. (2008), "The Economics and Psychology of Personality Traits", *Journal of Human Resources*, 43(4), pp. 972-1059

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

[9] Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Diane, S.W., Yagan, D. (2010), How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR, NBER Working Paper 16381

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

[11] Cunha, F., Heckman, J.J., Lochner, L., Masterov, D.V. (2006), "Interpreting the Evidence on Life Cycle Skill Formation", *Handbook of the Economics of Education*, 1, pp. 679-812

[12] Cunha, F., Heckman, J.J. and Shennach, S.M. (2010), "Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Econometrica*, 78(3), pp. 883-931

[13] Cunha F., Elo I., Culhane J. (2013), "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation", NBER Working Paper 19144

[14] Dee, T. and West, M. (2008), "The Non-Cognitive Returns to Class Size", NBER Working Paper 13994

[15] Duckworth, A.L., Peterson, C., Matthews, M.D., and Kelly, D.R. (2007), "Grit: Perseverance and Passion for Long-Term Goals", *Journal of Personality and Social Psychology*, 92(6), pp. 1087-1101

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

[17] Dweck, C. Mindset: The New Psychology of Success. Random House Digital, Inc., 2006

[18] Eriksson, T., Poulsen, A. and Villeval, M.C. (2009), "Feedback and Incentives: Experimental Evidence", *Labour Economics*, 16(6), pp. 679-688

[19] Eskreis-Winkler, L., Shulman, E.P., Beal, S. and Duckworth, A.L. (2014), "Survivor mission: Why those who survive have a drive to thrive at work", *Journal of Positive Psychology*, 9(3), pp. 209-218

[20] Gill, D., Prowse, V. and Vlassopoulos, M. (2013), "Cheating in the Workplace: An Experimental Study of the Impact of Bonuses and Productivity", *Journal of Economic Behavior and Organization*, 96, pp. 120-134

[21] Gneezy, U. and Potters, J. (1997), "An experiment on risk taking and evaluation periods", *Quarterly Journal of Economics*, 112(2), pp. 631-645

[22] Golsteyn, B.H.H., Gronqvist, H. and Lindahl, L. (2013), "Adolescent Time Preferences Predict Lifetime Outcomes", *Economic Journal*, 124, pp. 739-761

[23] Good, C., Aronson, J. and Inzlicht, M. (2003), "Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat", *Applied Developmental Psychology*, 24(6), pp. 645-662

[24] Heckman, J.J., Stixrud, J. and Urzua, S. (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour", *Journal of Labor economics*, 24, pp. 411-482

[25] Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A.Q. (2010), "The Rate of Return to the High Scope Perry Preschool Program", *Journal of Public Economics*, 94(1-2), pp. 114-128

[26] Heckman, J.J., Pinto, R. and Savelyev, P.A. (2013), "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes", *American Economic Review*, 103(6), pp. 1-35

[27] Heckman, J.J., Moon, S.H. and Pinto, R. (2014), The Effects of Early Intervention on Abilities and Social Outcomes: Evidence from the Carolina Abecedarian Study, unpublished manuscript

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

[29] Maddie, S.R., Matthews, M.D., Kelly, D.R., Villarreal, B. and White, M. (2012), "The Role of Hardiness and Grit in Predicting Performance and Retention of USMA Cadets", *Military Psychology*, 24, pp. 19-28

[30] Moffit, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R.,Roberts, B.W., Ross, S., Sears, N. R., Thomsom, W. M. and Caspi, A. (2011), "A Gradient of Childhood Self-control Predicts Health, Wealth, and Public Safety", *Proceedings of the National Academy of Sciences*, 108 (7), 269-398

[31] Raven, J., Raven, J.C., & Court, J.H. (2004), "Manual for Raven's Progressive Matrices and Vocabulary Scales", San Antonio, TX: Harcourt Assessment

[32] Roberts, B.W., Kuncel, N.R., Shiner, R.L., Caspi, A. and Goldberg, L.R. (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(4), pp. 313-345

[33] Sutter, M., Kocher, M.G., Ruetzler, D. and Trautmann, S.T. (2013), "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior", *American Economic Review*, 103(1), pp. 510-531

Tables

Table 1: Mean Comparisons of Pre-Treatment Variables

	Mean [SD]	Treatment	Control	Difference (p-value)
Malleability (pre)	0.00	-0.01	0.01	-0.02
	[1.00]			(0.90)
Perseverance (pre)	0.00	0.00	0.00	0.00
	[1.00]			(0.97)
Confidence (pre)	0.00	0.00	0.01	-0.01
	[1.00]			(0.87)
Gender (Male=1)	0.52	0.51	0.53	-0.01
	[0.50]			(0.49)
Age	10.02	10.02	10.02	0.00
	[0.42]			(0.97)
Raven	0.00	0.02	-0.02	-0.03
	[1.00]			(0.81)
Task Ability	5.07	5.00	5.14	-0.14
	[2.22]			(0.45)
Risk Tolerance	2.52	2.49	2.56	-0.07
	[1.49]			(0.60)
Wealth	2.84	2.80	2.89	-0.08
	[0.97]			(0.58)
Behaviour Score	4.35	4.37	4.32	0.06
	[0.87]			(0.66)
Success Score	3.41	3.35	3.49	-0.13
	[1.06]			(0.13)
Math	67.14	66.44	67.94	-1.50
	[23.84]			(0.68)
Turkish	70.42	69.59	71.37	-1.78
	[20.00]			(0.63)
Social Sc./Science	77.23	76.26	78.35	-2.09
	[17.40]			(0.50)
Top Math	0.21	0.20	0.21	-0.01
	[0.41]			(0.78)
Top Turkish	0.19	0.20	0.17	0.03
	[0.49]			(0.56)
Top Social Sc./Science	0.28	0.29	0.26	0.03
	[0.45]			(0.54)
Class Size	37.14	38.91	34.96	3.94
	[8.34]			(0.17)

* Column 1 contains the means of the pre-treatment variables with standard deviations in brackets. Columns 2 and 3 present the means of the treated and control group respectively. Column 4 shows 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 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 was elicited using the incentivized Gneezy and Potters (1997) task. The student's wealth, behavior and success is based on reports by teachers. Top grades in each subject are defined as scoring 90 or above.

Table 2: Associations in Control Group

	(1) Math Score	(2) Math Score	(3) Math Score	(4) Math Score
Raven	7.257***	7.342***	7.339***	6.816***
	(0.54)	(0.50)	(0.46)	(0.35)
All Difficult Visit 1	5.695***			
	(1.41)			
Plan Difficult Visit 2		4.835***		
		(1.38)		
Malleability (pre)			2.872***	
			(0.59)	
Perseverance (pre)				3.797***
				(0.49)
R-squared	0.28	0.27	0.30	0.33
N	715	708	677	594

* 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 in all four regressions is the student's math test score. 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. The variables malleability and perseverance are extracted factors from the pre-treatment student survey.

Table 3: Treatment Effect on Choice of Difficult Task

	(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***	0.095**	0.135***	0.118***	0.082***	0.087***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Task Ability	0.047***	0.061***	0.064***	0.068***	0.066***	0.068***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Gender (Male=1)	-0.013	0.100***	0.070***	0.048**	0.078***	0.042**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Raven Score	0.014	0.030**	0.029*	0.042***	0.037**	0.030*
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Malleability (pre)	0.009	-0.008	-0.004	-0.007	0.012	0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Perseverance (pre)	-0.013	0.014	-0.005	0.027*	0.012	0.005
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Math Score	0.001*	0.001	0.002***	0.002**	0.002**	0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Risk Tolerance	0.018**	0.012	0.013*	0.012*	0.007	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 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. 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 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 4: Treatment Effect on Success and Payoffs

	(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	0.020	0.033	0.109	0.054
	(0.04)	(0.11)	(0.06)	(0.08)	(0.10)	(0.09)
Task Ability	0.083***	0.301***	0.162***	0.246***	0.190***	0.190***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Gender (Male=1)	0.031	0.207**	0.128*	0.192***	0.103	0.135*
	(0.02)	(0.09)	(0.07)	(0.06)	(0.06)	(0.08)
Raven Score	0.041*	0.130**	0.132***	0.079**	0.077	0.084
	(0.02)	(0.06)	(0.03)	(0.04)	(0.05)	(0.05)
Malleability (pre)	-0.006	-0.024	0.048	0.013	-0.026	0.011
	(0.01)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Perseverance (pre)	-0.006	0.006	-0.034	0.010	0.004	0.066
	(0.02)	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)
Math Score	0.003***	0.011***	0.008***	0.006***	0.006***	0.004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
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. 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 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 5: Treatment Effect on Choice of Difficult Task

	(1) Difficult (After Failure)	(2) Difficult (Next Week)
Treatment	0.178***	0.142***
	(0.06)	(0.04)
Task Ability	0.032***	0.055***
	(0.01)	(0.01)
Gender (Male=1)	0.131***	0.010
	(0.04)	(0.02)
Raven Score	0.044**	0.023
	(0.02)	(0.01)
Malleability (pre)	-0.013	0.009
	(0.03)	(0.02)
Perseverance (pre)	-0.001	0.045***
	(0.04)	(0.02)
Math Score	-0.000	0.002**
	(0.00)	(0.00)
Risk Tolerance	0.012	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 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 6: Success and Payoffs in the Second Visit

	Success	Payoff			
	(1)	(2)	(3)	(4)	(5)
	Imposed	All	Imposed	Not Imposed	Total
Treatment	0.086***	0.300***	0.337**	0.229*	0.362**
	(0.03)	(0.09)	(0.13)	(0.12)	(0.14)
Task Ability	0.079***	0.283***	0.298***	0.263***	0.512***
	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Gender (Male=1)	0.022	0.073	0.090	0.042	0.267**
	(0.03)	(0.08)	(0.13)	(0.08)	(0.10)
Raven Score	0.038**	0.184***	0.152**	0.227***	0.274***
	(0.01)	(0.05)	(0.06)	(0.07)	(0.08)
Malleability (pre)	0.019	0.096**	0.077	0.116*	0.090
	(0.02)	(0.04)	(0.06)	(0.06)	(0.06)
Perseverance (pre)	0.008	0.011	0.026	-0.005	0.012
	(0.02)	(0.05)	(0.06)	(0.07)	(0.06)
Math Score	0.003***	0.013***	0.015***	0.011**	0.018***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Risk Tolerance	0.008	0.039*	0.037	0.042	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. 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 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 7: Success and Payoffs in the Second Visit - Difficult Imposed

	Choice Easy		Choice Difficult	
	(1) Success	(2) Payoff	(3) Success	(4) Payoff
Treatment	0.106**	0.428**	0.034	0.088
	(0.05)	(0.20)	(0.04)	(0.16)
Task Ability	0.078***	0.304***	0.067***	0.266***
	(0.01)	(0.05)	(0.01)	(0.04)
Gender (Male=1)	0.003	0.015	0.053	0.186
	(0.04)	(0.18)	(0.04)	(0.17)
Raven Score	0.028	0.106	0.034*	0.138*
	(0.03)	(0.12)	(0.02)	(0.08)
Malleability (pre)	0.024	0.094	-0.001	0.000
	(0.02)	(0.10)	(0.02)	(0.08)
Perseverance (pre)	0.022	0.082	-0.010	-0.051
	(0.03)	(0.12)	(0.02)	(0.09)
Math Score	0.001	0.005	0.004***	0.018***
	(0.00)	(0.00)	(0.00)	(0.00)
Risk Tolerance	-0.010	-0.041	0.022*	0.101*
	(0.02)	(0.06)	(0.01)	(0.05)
Control Mean	0.39	1.54	0.72	2.87
N	425	425	468	468

* Reported estimates in columns 1 and 3 are marginal effects from logit regressions. Estimates in columns 2 and 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 columns 1 and 3 is a dummy which equals 1 if the student was successful in meeting the target. The outcome in columns 2 and 4 is the student's payoff. Estimates in columns 1 and 2 use observations for whom the difficult task was imposed but who would have preferred to do the easy game. Estimates in columns 3 and 4 are obtained using observations for whom the difficult task was imposed and who also preferred the difficult game. 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 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 8: Spring Grades - Total Score

	(1) Math	(2) Turkish	(3) Life/Social Sc.	(4) All
Treatment	-1.187	1.227	1.921	0.620
	(2.00)	(1.36)	(1.40)	(1.09)
Gender (Male=1)	0.049	-3.471***	0.338	-1.035***
	(0.58)	(0.50)	(0.44)	(0.35)
Raven	2.950***	1.965***	2.326***	2.416***
	(0.55)	(0.59)	(0.46)	(0.41)
Math (pre)	0.404***	0.203***	0.107***	0.237***
	(0.04)	(0.03)	(0.03)	(0.03)
Life/Social Sc. (pre)	0.231***	0.321***	0.404***	0.319***
	(0.04)	(0.04)	(0.04)	(0.02)
Turkish (pre)	0.100**	0.202***	0.177***	0.158***
	(0.04)	(0.07)	(0.04)	(0.04)
Class size (ln)	1.409	-0.942	-3.166	-0.995
	(2.57)	(2.71)	(2.38)	(1.73)
Control Mean	68.40	71.94	74.67	71.80
N	2149	2148	2149	2132

* 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
** Outcomes are the math, social science and Turkish scores, and the overall mean of the three scores. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). Additional controls are the pre-treatment grades in all three core subjects and class size (ln).
*** Control mean refers to the unconditional mean of the outcome in the control group.

Table 9: Spring Grades - Percent Students with Top Grades

	(1) Math	(2) Turkish	(3) Life/Social Sc.	(4) All
Treatment	0.020	0.007	0.047	0.032**
	(0.02)	(0.03)	(0.03)	(0.02)
Gender (Male=1)	0.022	-0.046***	0.023*	0.012
	(0.01)	(0.01)	(0.01)	(0.01)
Raven	0.062***	0.047***	0.037***	0.057***
	(0.01)	(0.01)	(0.01)	(0.01)
Math (pre)	0.006***	0.003***	0.003***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)
Life/Social Sc. (pre)	0.006***	0.006***	0.008***	0.006***
	(0.00)	(0.00)	(0.00)	(0.00)
Turkish (pre)	0.001*	0.005***	0.005***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Class size (ln)	0.038	0.071	-0.037	0.031
	(0.03)	(0.05)	(0.06)	(0.03)
Control Mean	0.18	0.16	0.18	0.12
N	2149	2148	2149	2132

* 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 is a dummy variable which equals 1 if the student has top grades (defined as scoring 90 or above) in math, social science, Turkish and the overall mean of the three scores. Treatment is a dummy variable which equals 1 if the student attends a school which has been treated with the grit intervention. The Raven score is measured using a progressive Raven's matrices test (Raven et al., 2004). Additional controls are the pre-treatment grades in all three core subjects and class size (ln).
*** Control mean refers to the unconditional mean of the outcome in the control group.

Table 10: Optimality Analysis

	First Visit			Second Visit		
	(1) Optimal	(2) Over-Choose	(3) Under-Choose	(4) Optimal	(5) Over-Invest	(6) Under-Invest
Treatment	-0.011	0.014	-0.004	0.092**	0.008	-0.098**
	(0.03)	(0.02)	(0.02)	(0.04)	(0.01)	(0.04)
Task Ability	0.060***	-0.069***	0.008**	0.034***	-0.023***	-0.010
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Gender (Male=1)	0.031*	-0.036*	0.003	0.018	0.006	-0.028
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Raven Score	0.042***	-0.032**	-0.011	0.031*	-0.013**	-0.015
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Malleability (pre)	0.010	0.007	-0.017***	0.020	-0.010**	-0.011
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.02)
Perseverance (pre)	-0.027*	0.020	0.006	0.043**	-0.005	-0.036**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
Math Score	0.004***	-0.004***	0.001***	0.000	-0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Risk Tolerance	-0.001	0.010	-0.011**	0.014	-0.004	-0.012
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Control Mean	0.65	0.28	0.07	0.52	0.03	0.45
N	1704	1704	1704	1578	1578	1578

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

** To define the dummy variables which indicate whether the student chooses optimally, over-invests or under-invests we compare the actual task choices to the choices which are 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, perseverance and confidence are extracted factors from the pre-treatment student survey. Math 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 11: Post-Treatment Survey

	(1) Malleability	(2) Perseverance	(3) Confidence
Treatment	0.373***	0.294***	-0.018
	(0.07)	(0.06)	(0.05)
Gender (Male=1)	0.031	-0.190***	0.054
	(0.04)	(0.03)	(0.05)
Raven Score	0.103**	0.071***	0.004
	(0.04)	(0.02)	(0.03)
Malleability (pre)	0.202***	0.007	-0.033
	(0.03)	(0.03)	(0.03)
Perseverance (pre)	0.199***	0.360***	-0.014
	(0.04)	(0.03)	(0.03)
Confidence (pre)	-0.023	0.093***	0.440***
	(0.02)	(0.02)	(0.03)
Math Score	0.008***	0.008***	0.010***
	(0.00)	(0.00)	(0.00)
Risk Tolerance	-0.007	-0.012	0.008
	(0.01)	(0.01)	(0.01)
Control Mean	-0.21	-0.17	0.01
N	1690	1612	1675
R-Squared	0.25	0.30	0.27

* 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, perseverance and confidence are extracted factors from the pre-treatment student survey. Math 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.

Figures

Figure 1: Teacher Survey: How much do you agree with ideas?

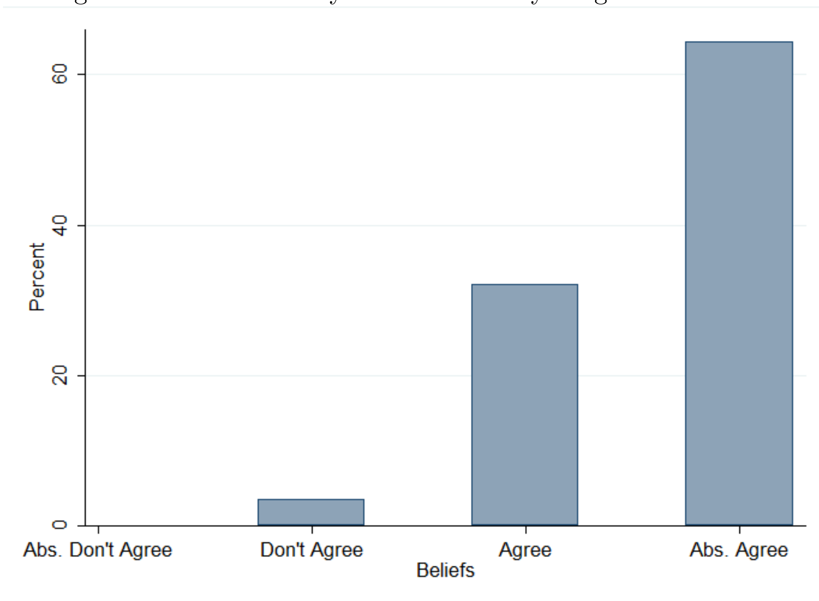


Figure 2: Teacher Survey: How intensely did you implement the ideas?

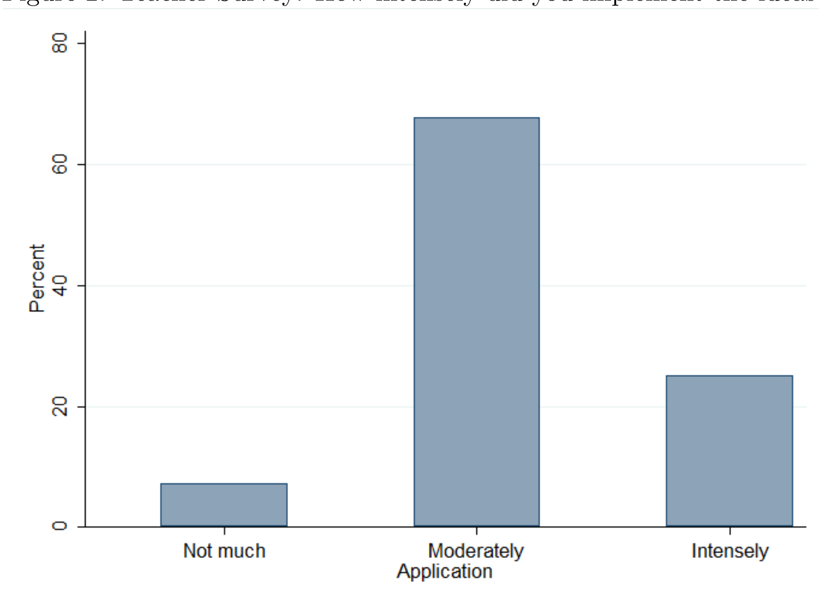


Figure 3: Design

		Grit & Fwd-Looking Behavior (IT)	Fwd-Looking Behavior (CT)	Pure Control (PC)
Phase 1	Training	-	-	-
	Measurement	March 2013: Baseline Data Collection		
Phase 2	Training	Spring 2013 (Fwd-L)	-	-
Phase 3	Training	Fall 2013 (Grit)	Fall 2013 (Fwd-L)	-
Phase 4	Training	-	-	-
	Measurement	May 2014: Follow-up Data Collection (Grit)		

Figure 4: Effect of Treatment on Choosing Difficult Task

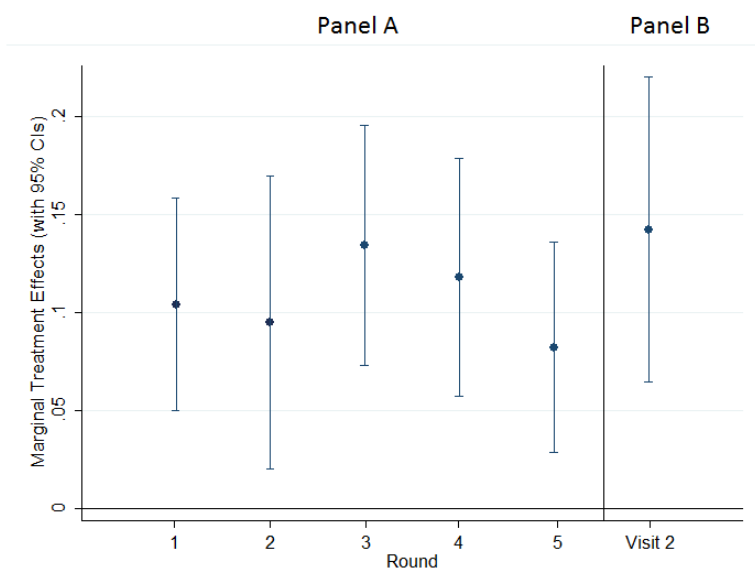


Figure 5: Effect of Treatment on Payoffs

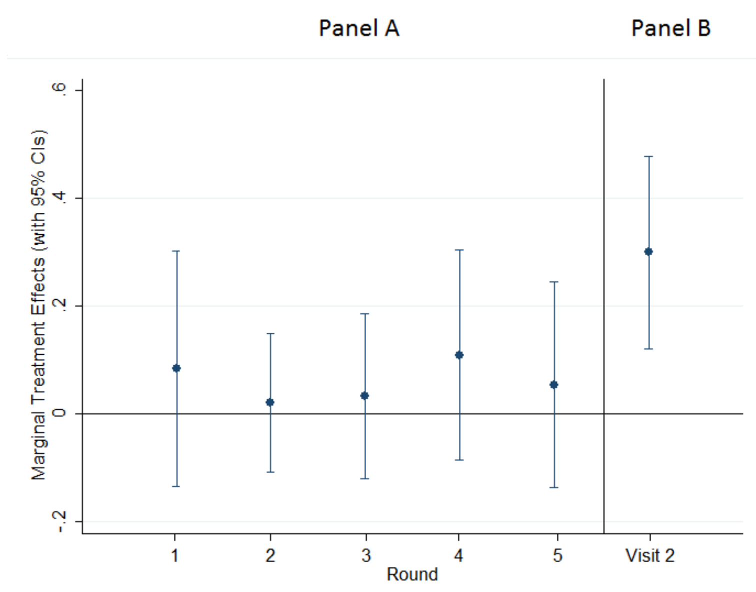


Figure 6: Effect of Treatment on Grades

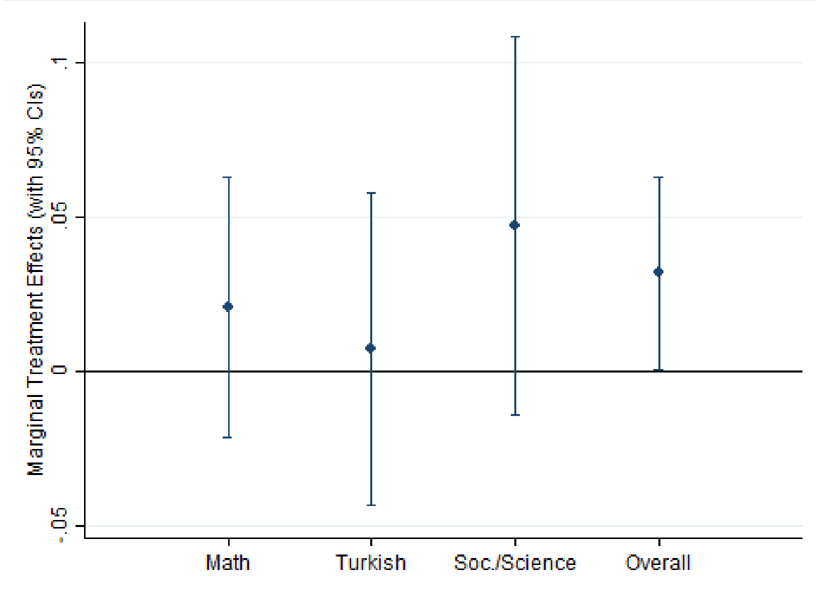


Figure 7: Kernel Density of Residual Payoff Distribution Week 2

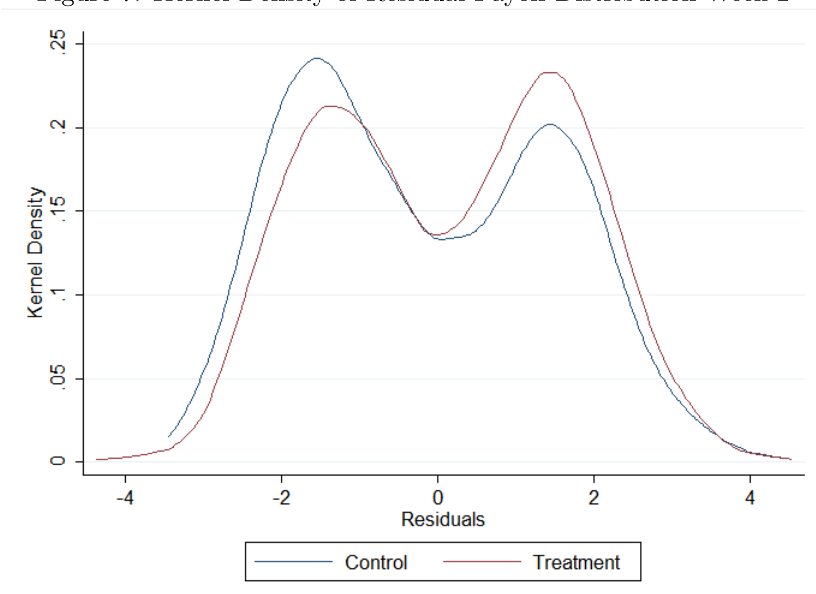


Figure 8: Effect of Treatment on Self-Reported Malleability Beliefs

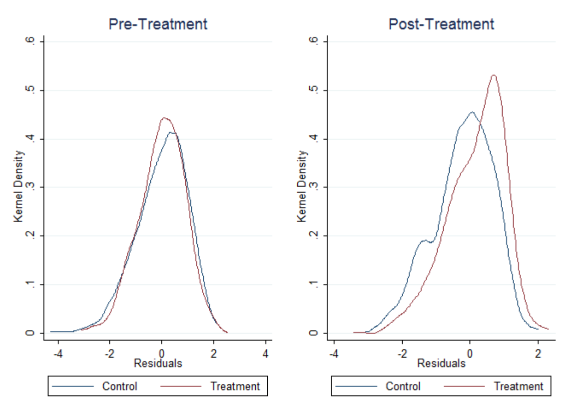


Figure 9: Effect of Treatment on Self-Reported Perseverance

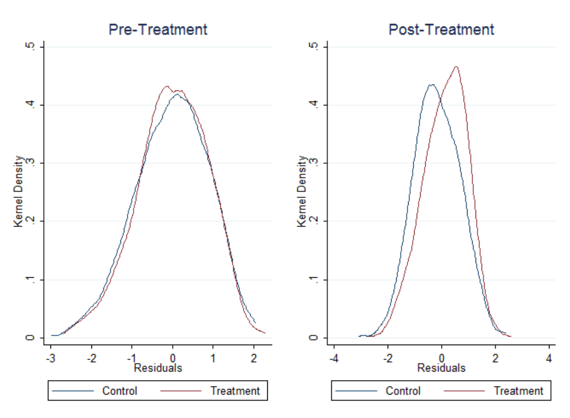
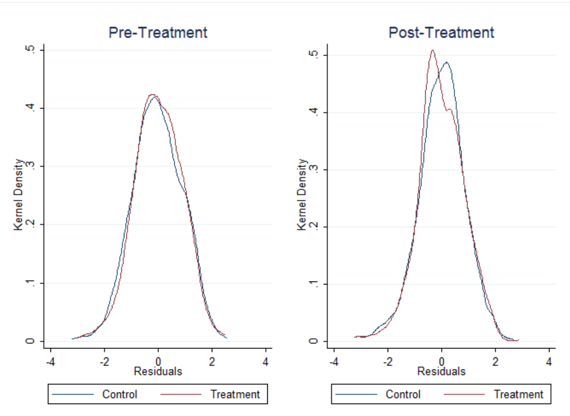


Figure 10: Effect of Treatment on Self-Reported Confidence



Appendix: Educational Material and Tasks

Figure 1: Screenshot of Educational Video Material

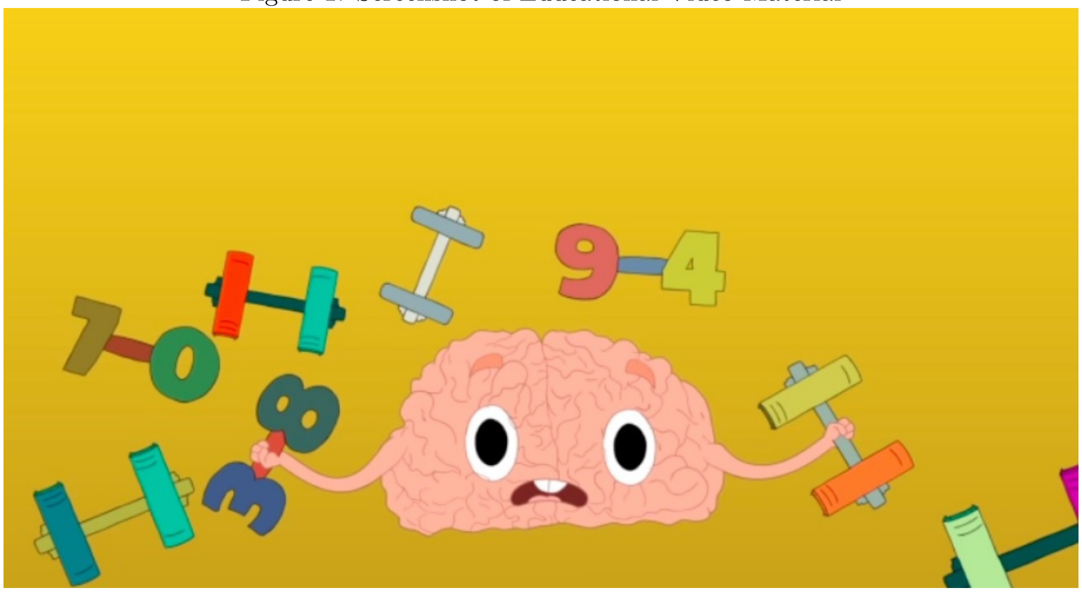


Figure 2: Example of Easy Task

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