Don't Stop Believing:

Empowering Low SES Students Through Cognitive Behavioral Interventions

Juan Pereira[1]

PRELIMINARY DRAFT DO NOT CIRCULATE

Abstract

Unrealized potential among low-income students is a significant loss for both individuals and society. Psychological constraints, such as low aspirations and underconfidence, limit students' ability to pursue opportunities. This paper evaluates a cognitive behavioral therapy (CBT) intervention designed to help students overcome these barriers. Implemented in Uruguay's CECAP program and evaluated through a randomized controlled trial, I find that the intervention improved math scores by 6 percentage points and increased school enrollment by 12 percentage points. Treated students also showed greater resilience, stronger beliefs in effort-driven success, and higher expectations of achieving their educational goals. These findings highlight CBT's potential to foster behavioral change and inform policies when there is a concrete belief attached to the therapy.

[1] This Project has received generous funding from ANII Urguay, Bravo Research Center at Brown University, and the Watson Institute for Public Policy. I would like to thank María José Soler from Universidad de Montevideo, Cecilia Pereira from Instituto CEDIIAP, Inés Arrieta from NGO Jóvenes Fuertes, and María Tarigo from Instituto CEDIIAP for designing the intervention and doing the teachers' training. I thank Elisa Macchi, Andrew Foster, Peter Hull, Joaquin Paseyro, and participants of the 2024 Russel Sage Behavioral Economics Summer Camp for their helpful comments.

1. Introduction

Unrealized potential among low-income students represents a critical loss—not just for the individuals affected but for society as a whole. Recent research highlights how psychological constraints, such as low aspirations, underconfidence, and low self-esteem, prevent these students from seizing educational and economic opportunities (Guyon and Huliery 2022; Alan 2017; Deaming 2021; Appadurai 2017; Ridley et al. 2020; Mani et al 2013). However, there is limited evidence on effective tools that can empower them to make better decisions in the classroom, and even less is known about the mechanisms through which these interventions operate.

In this paper, I examine how a cognitive behavioral therapy (CBT) curriculum designed to address students' limiting beliefs about their potential for success can enhance their aspirations and learning outcomes. Specifically, I analyze behavioral differences between students in the treatment and control groups to better understand the mechanisms driving the results.

I partnered with the Ministry of Education in Uruguay and the NGO Jóvenes Fuertes to design and offer CBT classroom intervention to students in a government program that assists failing secondary school students. We trained teachers to deliver CBT to the students. Then, each of these teachers conducted 10 workshops with the same classroom of students. The intervention helps students identify and challenge negative beliefs and emotions about the role of effort in achieving success. By disrupting the automatic flow of thoughts and emotions triggered by specific situations, it prevents unconscious patterns of behavior that may limit student's potential.

The intervention unfolds within a program offered by the MoE to help teenagers who have discontinued high school reintegrate into the formal education system. Students in this program have typically tried and failed the first year of high school. They belong to the poorest families in the country and are attending CECAP to look for a last chance to re-enroll in the formal education system. The goal is to support their completion of the first three years of high school as a way of reintegrating them into high school and with the hope that they may pursue further education beyond high school. Against this background, a targeted CBT intervention to help students make better decisions at a very crucial time could have lasting impacts.

I experimentally evaluate the causal impact of the intervention by carrying out a randomization at the classroom level. There are two arms in the experiment: the treatment group, which includes students from classrooms that receive the intervention, and the control group, where students from these classrooms carry out regular lectures. Each group has a designated attached teacher who holds two hours of lecture every week, they use this space to carry out the weekly meetings. In the absence of

the intervention, students typically discuss salient situations in the neighborhood and at educational centers. Control students keep discussing these issues while treated students are instructed with our curriculum.

I examine the impact of the intervention on math and language test scores, as well as students' enrollment in the next semester and high school. To explore the mechanisms driving these outcomes, I conducted three incentivized games. The first game tests how students' behavior changes when allowed to practice and improve on a new task. The second game examines their response to failure, while the third evaluates the effects of labeling by assessing performance in a math game after students are primed with a message suggesting it is designed for "math geniuses." Finally, I investigate changes in students' aspirations—measured by the number of years of high school they aim to complete and the role they wish to have at CECAP—as well as their expectations, captured through their perceived likelihood of achieving these aspirations

I find that students who participated in the program performed better on math tests by 6 percentage points and were, on average, 12 percentage points more likely to enroll in secondary school courses in CECAP during the semester immediately following the intervention. The incentivized games reflect that students who participated in the intervention believe, on average, that effort leads to better outcomes and are more resilient in the face of failure. Additionally, these students reported higher expectations of achieving the number of years of education they aspire to complete. The results suggest that the intervention influenced the students through the expected mechanisms, emphasizing the importance of the CBT component of the intervention.

This study contributes to three main strands of the economics literature. First, it adds to the evidence looking at CBT as a feasible public policy. Heller et al 2017, Blattman et al 2017, 2023 found significant effects of CBT interventions on impulsive reactions, and Baker et al 2024 found significant results of economic measures. In this paper, I intend to open the black box of the effect of CBT on an individual’s behavior by providing a CBT intervention attached to a belief and test the main hypothesis through incentivized tasks.

Second, this research contributes to a literature that looks at the effect of soft skills training on economics and educational outcomes. Alan et al 2019, Avitabile 2019, Ganimian 2020, and Outes-Leon et al 2020 among others, point out at the importance of teaching soft skills, especially Fixed and Growth Mindset, for educational outcomes. Finally I contribute to a research agenda that looks into ways to help individuals learn he ‘capacity to aspire’, the idea that the existence of opportunities does not guarantee their realization and people need realistic means of identifying (Dalton 2016, Rao and Walton 2004, Genicto and Ray 2017, Appadurai et al 2004). Without this realistic means, internal

constraints shape behavior in ways that perpetuate poverty. Finally, this paper contributes to a wider audience in psychology that studies Agency and Mindset (Dweck et al 2019, Yager et al 2013, Bandura 1987)

1. Intervention

The primary objective of the intervention is to help CECAP students identify and challenge—if they deem it appropriate—negative beliefs. The intervention assists students in recognizing negative beliefs and emotions about their efforts' role in achieving outcomes. In other words, it helps students interrupt the automatic flow of thoughts and emotions triggered by certain situations, unconsciously leading to behaviors that may hinder their potential.

For instance, when a student believes they are not good at math, this thought triggers a cascade of emotions—nervousness, anxiety, and insecurity—that undermine their confidence. These emotions lead to self-doubt and hesitancy during testing, which disrupts their focus and performance. Consequently, they might make more mistakes, reinforcing their initial belief that they are incapable. Importantly, this cycle might not be reflective of their actual ability. In many cases, the student can succeed at the task but is held back by the self-defeating belief that they cannot. These are the situations we are aiming to change with this intervention.

It is important to emphasize that this intervention aims to help students identify automatic thoughts, but once identified, it is up to the student to decide whether these thoughts are well- or poorly-calibrated. In other words, this is not a “you can do it” intervention designed to recalibrate students' expectations. Instead, it encourages students to replace fast, automatic thought processes with slow, deliberate ones, enabling them to critically evaluate their thoughts—either challenging or accepting them.

The development of the intervention began with training 51 teachers from Uruguay’s Ministry of Education CECAP program. Psychologists and educational psychologists from the NGO Jóvenes Fuertes, the University of Montevideo, and the CEDIIAP Institute conducted the training. The first part consisted of a two-day in-person module introducing teachers to cognitive behavioral therapy (CBT) principles. The second part involved six remote sessions, each lasting two hours and held every two weeks. In these sessions, teachers shared their experiences and, together with a facilitator, prepared for upcoming workshops.

Following the training, each teacher conducted ten workshops with their students. Teachers received a detailed book guide for each workshop, while students received exercise booklets. The intervention

occurred during students' weekly meetings with their assistant teachers. In the absence of the workshops, these meetings usually revolve around current issues in the neighborhood or the CECAP center. Students in the control group continued to participate in these regular meetings. The curriculum is heavily based on CBT principles, described by the Beck Institute as "a structured form of talk therapy that helps individuals recognize and change negative patterns of thinking and behavior, enabling them to handle challenging situations better and improve overall quality of life."

Each of the ten workshops followed the same structure. They began with a video that introduced the day's theoretical component. For instance, Lesson 3 presented the three core beliefs addressed in the workshops and discussed scenarios where students commonly encounter them. After the video and discussion, each lesson included exercises to practice the concepts introduced in the video. Finally, each session concluded with a practical exercise to uncover negative beliefs, understand the emotions they evoke, and test strategies to identify and manage them. According to cognitive behavioral therapy (CBT) principles, practicing is a fundamental component for changing behavior, as it helps individuals internalize new patterns of thought and action through repeated application

For example, in Workshop 4, two students were asked to role-play a situation where they faced an exam. All students were presented with the same test and asked to attempt the initial questions. This activity aimed to detect negative beliefs—when and where they emerged, how they shaped internal dialogues, and what strategies students could use to mitigate their harmful effects.

The intervention curriculum is divided into four modules. The first focuses on the basics of thoughts, emotions, and behaviors and the concept of neuroplasticity. The second module, the core of the intervention, addresses three limiting beliefs. The first belief is tied to absolute judgments, such as "I'm dumb," while the other two relate to how students perceive the relationship between effort and failure in the learning process. The third module teaches students to understand, accept, and manage emotions, while the fourth helps them plan for situations likely to trigger negative thoughts and emotions. Hence, the intervention may influence how students approach situations that trigger limiting thoughts in two ways. First, in the moment, such as during a challenging task, students may identify and address negative thoughts as they arise. Second, through prior planning, students can anticipate scenarios likely to evoke negative mental models. For instance, when preparing for an exam, they can recognize beforehand that such thoughts may occur and decide on a technique to challenge them.

While the intervention is designed to support students in addressing a wide range of thoughts, particular emphasis is placed on three key beliefs that serve as the foundation for practicing and reinforcing ideas and examples. These are the three core beliefs:

All-or-Nothing (Black-and-White) Thinking: This occurs when individuals believe they lack the ability to complete a task, leading to entirely negative judgments such as “I’m not smart,” “I’m not good enough,” or “I’m a failure.” For example, a student facing a challenging math problem within their skill level might think, “I’m terrible at math; I can’t do this.”

Returns on effort: Relates to the belief that effort is a sign of lower ability. Associated thoughts are, for example, “This is too hard” and “This shouldn’t be such a struggle.” For instance, when faced with a difficult task, students may see effort as futile if they believe success is unattainable.

Magnifying mistakes: This refers to the belief that mistakes signal incompetence, leading students to make extreme predictions about the consequences of errors. For example, a mistake on a test might convince a student that they cannot perform well overall.

An essential component of the intervention is helping students experience these thoughts and the emotions and reactions they trigger. At the end of each workshop, students engage in exercises designed to elicit these thoughts and practice identifying them. Teachers then provide "antidotes" to counteract these limiting beliefs, such as the idea that "failure is an essential part of the learning process." It’s important to note that these "antidotes" are not presented as the behaviors students should adopt but as potential solutions for those who wish to challenge their negative thoughts.

The intervention consisted of two key components: CBT techniques and informational content. While both elements potentially played an important role, determining which channel was primarily responsible for the observed effects falls outside the scope of this evaluation. This study focuses on understanding the intervention's overall impact without isolating each component's specific contribution.

1. Context

The intervention unfolds within a non-formal education program administered by the Ministry of Education of Uruguay called CECAP (Centers for Education, Training, Art, and Production). This program targets adolescents who have dropped out of high school, intending to reintegrate them into the formal education system. CECAP has two main objectives: to reintegrate students into the formal education system and help them complete the first three years of high school while attending the program. The program is structured into four semesters. The first semester is an orientation period, during which students are introduced to the program and their future educational options.

A unique feature of CECAP is that at the beginning of each semester, students can attend CECAP in the morning and take high school courses in the afternoon, earning credits recognized by ANEP (National Administration of Public Education). If students complete three semesters of high school courses in parallel with CECAP, the formal education authority (ANEP) awards them a diploma certifying the completion of the first three years of high school.

CECAP operates 22 centers nationwide, and the intervention was carried out in 13 of them. However, it is worth noting that not all centers offer the option to take high school courses concurrently with CECAP. Eight of the thirteen centers included in this study provided this option.

The students participating in CECAP have unique characteristics that make them a particularly relevant population for this study. Table 1 shows that the CECAP student population comes from poor families where students have repeatedly failed to pass the first year of high school. On average, these students are 17 years old, with a majority being male (58%). Most have completed primary school, and 3.5% work while attending CECAP. About 25% live with both parents, 50% live with their mother 8% only with their father. 12% are institutionalized in foster care. These students come from households with very low education levels— parents and guardians have an average of 6.7 years of education, approximately two years less than the national average.

Table 1: Summary Statistics of Students' Characteristics

	Mean	Sd	Max	Min
Male	58%	0.49	1	0
Age	16.95	1.53	22	13
Years of Schooling	6.90	0.73	9	5
Worked before	22%	0.41	1	0
Age when started working	14.10	2.13	18	7
Works now	3%	0.18	1	0
Lives with both parents	25%	0.43	1	0
Lives only with the mother	50%	0.50	1	0
Lives only with the father	8%	0.27	1	0
Lives in foster Home	12%	0.32	1	0
Mother's education	6.66	1.73	16	0
Father's education	6.66	1.56	11	0

Note: This table shows the average characteristics of students in CECAP.

2. Evaluation Design and Data

Evaluation Design

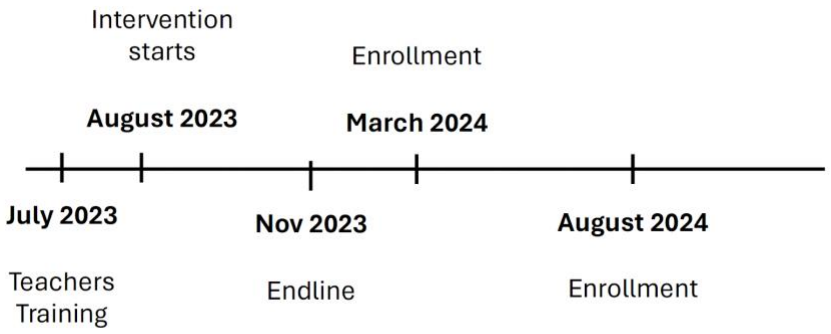
We randomly assigned 78 groups, comprising 703 students, to either the treatment or control group (315 students in 30 treated groups and 388 students in 48 untreated groups) across 13 educational centers. In each group, a teacher trained in the curriculum carried out the intervention. While treated students participated in workshops, control students followed their usual routine, meeting with a teacher to discuss key topics related to their neighborhood and the center.

The program's size constrained the experimental design, limiting the scope of randomization and the populations we could include. However, we prioritized conducting the intervention with a population most suited to its goals: low-income students who had experienced repeated academic failure. This decision meant that we randomized at the level of groups and could not randomize entire schools to create pure control institutions. These design choices introduce concerns regarding potential spillovers, which we will address in detail in the next section.

To minimize attrition, we selected a sample of students to address the high absenteeism and dropout rates. We anticipated an attrition rate consistent with the typical semester-to-semester dropout rate of around 30%. To address this, we remained at each center for approximately two weeks, trying to reach as many active students as possible. This setup, however, comes with a cost. Students who are frequently absent are less likely to participate actively in the program. Nevertheless, the fact that our sample includes approximately 70% of the total enrollment in each center indicates that we are successfully recruiting the most engaged students.

We evaluated the students at two points: before the teachers' training began and at the end, approximately a month and a half after the intervention was implemented. In addition to baseline and end-line, we use administrative data for enrollment outcomes as described in Figure 2 as the timeline. We administered psychometric instruments during baseline and end-line and conducted a survey to assess students' self-perception of their educational aspirations. Students took math and language tests on the same day as the survey. The main difference between the baseline and final evaluations was that the baseline included a Ravens Matrices test, while the final evaluation incorporated three incentivized tasks to understand mechanisms.

Figure 2: Intervention and Evaluation Timeline



Note: This figure shows the timeline of the intervention implementation.

Data

Expectations and Aspirations Survey

We collected a set of survey questions to understand students' self-perceptions, beliefs, and aspirations regarding their educational trajectories. These questions followed a consistent format, first asking about desired outcomes (e.g., "How many years of education do you wish to achieve?") and then about the perceived ability to achieve those outcomes (e.g., "How many years of education do you think you can achieve?"). Responses were recorded as either years of education or on a scale from 1 to 4, indicating agreement with a statement. The baseline and final surveys are included in the appendix for reference.

Math and Language Tests

The Language Test (TECLE) and Math Test (TECA) are brief assessments designed and pre-validated at the national level by the Catholic University of Uruguay.

The language test comprises 64 "fill-in-the-blank" questions to be answered in 5 minutes. For example:

Ana puso la ...

☐ mella ☐ mesa ☐ mefa ☐ meva

The math test (TECA) is an algebra assessment divided into three parts. The first part consists of 60 addition problems to be solved in 5 minutes, the second part of 60 subtraction problems to be solved in 5 minutes, and the final part of 40 multiplication problems to be solved in 3 minutes. This is an example of an addition question:

4-1 = | ☐ 4 ☐ 2 ☐ 3 ☐ 5

Students rarely complete all the questions within the given time, so confidence might manifest in their willingness to attempt answers. These tests were validated for students who have completed primary education, making them ideal for this population.

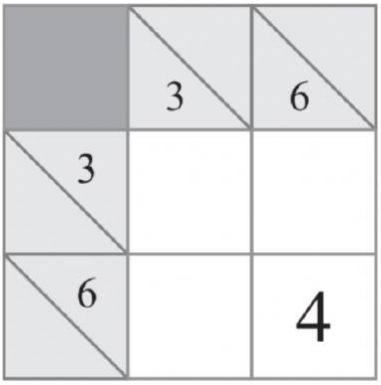
Incentivized Tasks – Games

In addition to evaluating educational outcomes and beliefs through surveys, we conducted three incentivized tasks to deepen our understanding of the mechanisms driving the results. These tasks were designed to examine whether students' behavior aligns with the beliefs the curriculum seeks to modify.

Incentivized task 1

This game tested whether treated students were more likely to believe that effort is important when attempting a new and challenging task: solving as many Kakuro puzzles as possible.

Example of a Kakuro Puzzle



Note: This figure shows a sample Kakuro presented to students to explain the task. Cells must be completed to match the sums at the end of rows and columns.

The task is designed in the following way:

Students first familiarized themselves with Kakuros by attempting a practice example. Then, they were asked:

1- "How many Kakuros can you solve correctly in 4 minutes?"

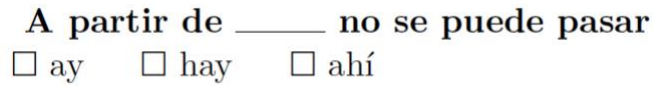
2- "How many do you think you can solve correctly if you have the opportunity to practice first?"

Students were then randomized into groups that practiced or didn't practice. We randomized this to understand further how accurate students' beliefs were. Among those randomized into practice, they first did 4 minutes of practice and then attempted to solve six Kakuros in 4 minutes. Those randomized to no practice did two 4-minute attempts to solve six kakuros. We compare the score from the second set of Kakuros for those who practice and the first set for those who do not.

Incentives are set up as follows: students get paid (either candies or education goods) 2A - A being the number answered in question (2), if they complete A correctly. If they do B<A and reply A to question (2), they get B, and if they do C>A but reply A, they are paid 2A+C. Hence, if students want to say a higher number than they think they can complete, they will always get the same or more by telling the truth. If they report a smaller number of what they believe they could achieve, they will get a smaller payoff than truth-telling.

Incentivized Task 2

This game tested the belief that "mistakes confirm my lack of ability." Students first answered six grammar questions in 30 seconds.

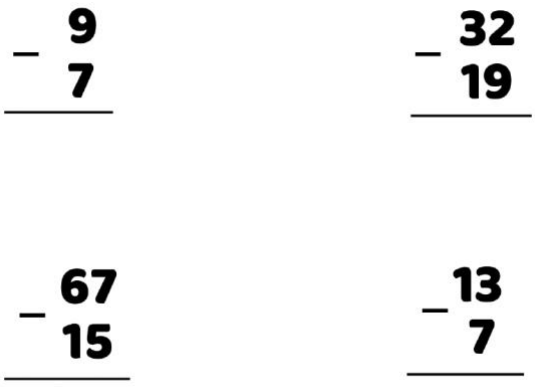


Students were randomized into two groups with differing pass thresholds (e.g., 3/6 correct vs. 5/6 correct). After the first test, facilitators immediately grade the first set, report students' grades and pass/fail. Then, students are invited to attempt a new set of questions. We look at the score and attempts in this second set of questions to estimate the intervention's effect on students' willingness to try after failing. In this game, students receive a prize for passing the tests and for each correct answer.

Incentivized task 3

This game explored how treated and control students reacted to a task labeled 'for math geniuses.' The idea was that the label might trigger limiting thoughts that could undermine the students' effort. Labeling experiments, common in behavioral economics, often examine how external cues, such as titles or descriptions, influence individuals' behavior by shaping their self-perception or expectations. In this case, the label was designed to test whether it would activate self-doubt or a fixed mindset, particularly among students who may already harbor beliefs about their limited mathematical abilities.

Example of subtraction in the first set of questions:



Students had two chances to complete subtraction exercises within 30 seconds in this task. Students are randomized into groups between the two sets of questions with a label saying, 'The next set of questions is for math genius.' The other group does not receive any label. Students receive a prize for each question answered correctly.

3. Estimation of Treatment Effects

We estimated the effects of the intervention (intention-to-treat (ITT)) using ordinary least squares, defined by the following equation:

$$Y_{i,g,endline} - Y_{i,g,baseline} = \alpha + \gamma\, CBT_g + X'\pi + \epsilon_g \qquad (1)$$

In this equation, “i” refers to the student, “g” to the group, and "endline" and "baseline" refer to whether the outcome “Y” was measured at baseline or end-line. The dependent variable is the difference for each student between the baseline and end-line outcomes. CBT is a group-level variable that equals one if the group was randomly assigned to the treatment and zero otherwise. Thus, the coefficient γ identifies the treatment effect. To improve precision, we included baseline characteristics (gender, age, schooling level, and results from the Ravens Progressive Matrices). Standard errors are clustered at the group level. We present a graphical version of this equation for most effects. We display bar graphs showing the averages for treated and control students, with the difference between the two bars representing the parameter of interest in the equation, γ.

Attrition was 18%, but it does not impose threats to the identification of the key parameter. Table A1 shows that the rate of attrition does not differ significantly between treatment and control groups; there is no differential attrition, and Table A2, further depicts that any observable student characteristics do not significantly predict attrition and that there are no material differences between the average characteristics of those whose attired in the control and the treatment group. In addition, attrition is also random at the grade level. Attrition is, however, concentrated in one of the CECAP centers that has 30% level of attraction, 12pp above the average. we run the analysis excluding this center for robustness, and results remain the same. Although this level of attrition is high, it was expected given the studied population and a cost worth assuming.

The two main assumptions required for causal identification are that students’ characteristics are balanced at baseline and the Stable Unit Treatment Value Assumption (SUTVA). Balance at baseline is shown in Table 2, where both groups have similar baseline characteristics.

Table 2: Balance Table

	Treated	Non-treated	Difference	P-value
How many years of High School do you want to complete? (0-4 scale)	3.03	2.92	0.10	0.19
How many years of High School you can complete? (0-4 scale)	2.83	2.73	0.11	0.15
Raven's Matrices (0-15 scale)	5.34	5.14	0.19	0.32
Language Test Scores (%)	49%	46%	2%	0.18
Math Test Scores (%)	44%	44%	0%	0.91
Male	60%	56%	4%	0.27
Years of Schooling	6.83	6.73	0.10	0.26
Age	16.24	16.22	0.02	0.88

Note: This table shows the average value by treated students in the first column, by non-treated students in the second column, and the difference in the third column. The 4th and last column shows the p-value of the difference in column 3.

The SUTVA assumption has two components: no interference between units and no variation in treatment across units. The first component, no interference between units, implies that a student's treatment status does not influence the status of another student with a different status. Given the experimental design, there are two concerns regarding this first component: spillovers from treated students to students in the control group and that all teachers delivering the workshops also taught untreated students. Despite emphasizing to these teachers not to use the intervention content with untreated groups, the intervention might influence how teachers interact with students. If treated students or teachers' behavior affects non-treated students, we expect this effect would bias our results downward. We do two analyses to understand the degree to which the first component might be violated; first, in the end-line survey, we ask students to complete the following sentence: "_____ Mindset", with one of the four options (a) worker's, (b) student's, (c) fixed, (d) men and women's. Table A3 shows that students in the control group respond (c) 65% of the time, while those in the control 15%. Second, we run the analysis controlling for teachers' fixed effects. Table A4 re-runs the analysis, adding teacher's fixed effect to show that the estimated coefficients do not change, implying that teachers' behavior does not seem to confound the intervention effect.

The second component of the SUTVA implies no variation in treatment across units; in other words, the treatment is delivered consistently to all individuals in the treatment group. The main difference in how the intervention is delivered depends on how teachers implement the curriculum; the fact that

adding teachers' fixed effects does not change the estimated parameter reinforces the idea that this assumption holds.

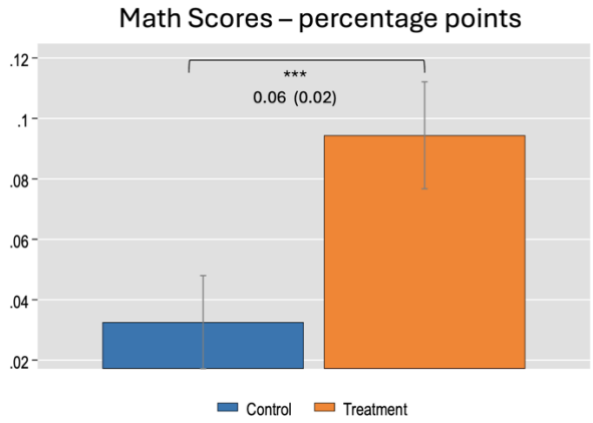
The effect of the intervention on psychometrics, survey measures of beliefs, dropouts, enrollments, and test scores are estimated based on equation (1). The three incentivized tasks cross-randomize different arms; hence, we estimate different models that add new assumptions for each task. Concretely, every time we cross-randomize, we show that students' characteristics are balanced across the cross-randomized variables.

4. Results

Math and Language

The intervention significantly increased the math arithmetic test scores but did not affect literacy test scores. Figure 3 illustrates the intervention's effect on math test scores. Between baseline and the end line, treated and control students increased their math scores: treated students by 0.9 percentage points and control students by 0.03 percentage points. This 0.06 percentage point difference is statistically significant at 99% confidence.

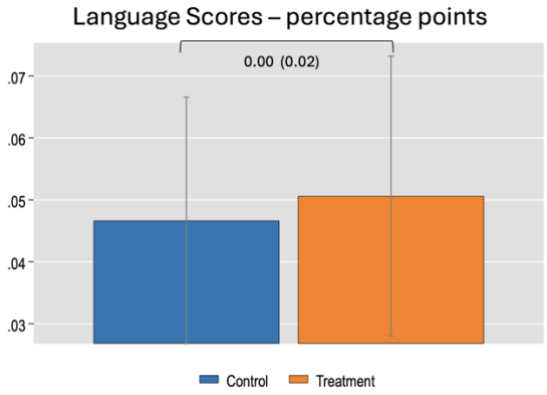
Figure 3: Percentage points difference in Math test scores



Note: This figure shows the average difference between baseline and endline for treated and control students. The orange column shows the average percentage point difference for treated students and the blue one for control students. The top part of the plot shows the difference between both columns and shows with starts the significance level. One * corresponds to 10% significance, ** 5%, and *** 1%. Standard errors are clustered at the group level.

In the literacy test (Figure 4), students in the treatment and control groups improved their number of correct answers at the end line by approximately five percentage points compared to the baseline. This uniform improvement across both groups implies that the intervention did not affect literacy test outcomes.

Figure 4: Percentage points difference in Language test scores

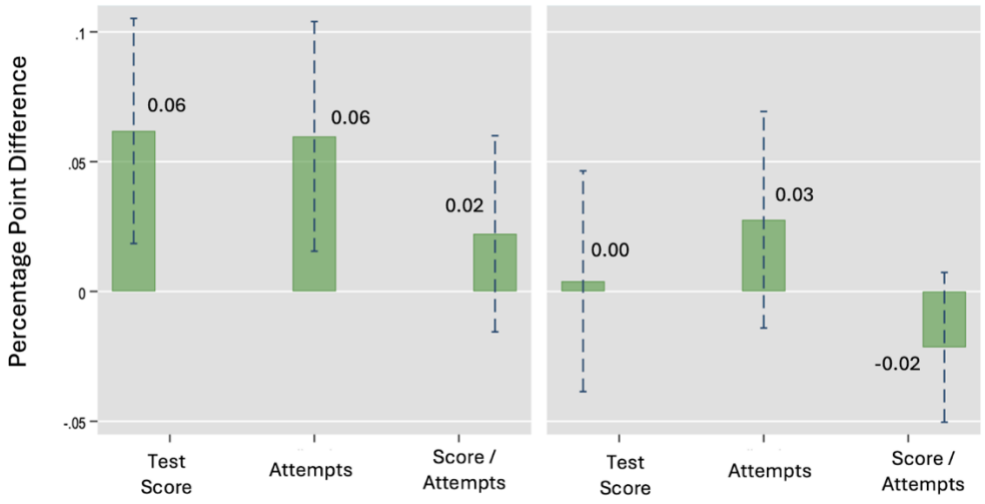


Note: This figure shows the average difference between baseline and enlinde for treated and control students. The orange column shows the average percentage point difference for treated students and the blue one for control students. The top part of the plot shows the difference between both columns and shows with starts the significance level. One * corresponds to 10% significance, ** 5%, and *** 1%. Standard errors are cluster at the group level.

To determine whether students increase the number of correct answers on the tests by arbitrarily raising the number of questions they complete, we examine the relationship between scores and attempts. The left panel of Figure 5 illustrates, in the first column, a 0.06 effect size on math scores, while the second column shows that treated students also attempted 0.06 percentage points more. Column 3 in Figure 3 presents the effect of the intervention on the ratio of correctly completed puzzles to attempted puzzles. If the effect in the third column is positive, it suggests that students were more accurate and converted fewer questions into correct answers. Conversely, if the coefficient is negative, it would imply that students increased the number of questions attempted without maintaining accuracy on each question, suggesting that they were attempting randomly. We estimate a non-significant result, indicating that students maintained the same level of accuracy as in the baseline. Therefore, the intervention did not arbitrarily increase the number of attempts.

The right side of Figure 5 shows that for the language test, all estimated coefficients are zero, indicating no differences between treated and control students.

Figure 5: Scores, Attempts, and Accuracy: Are studying arbitrarily trying?



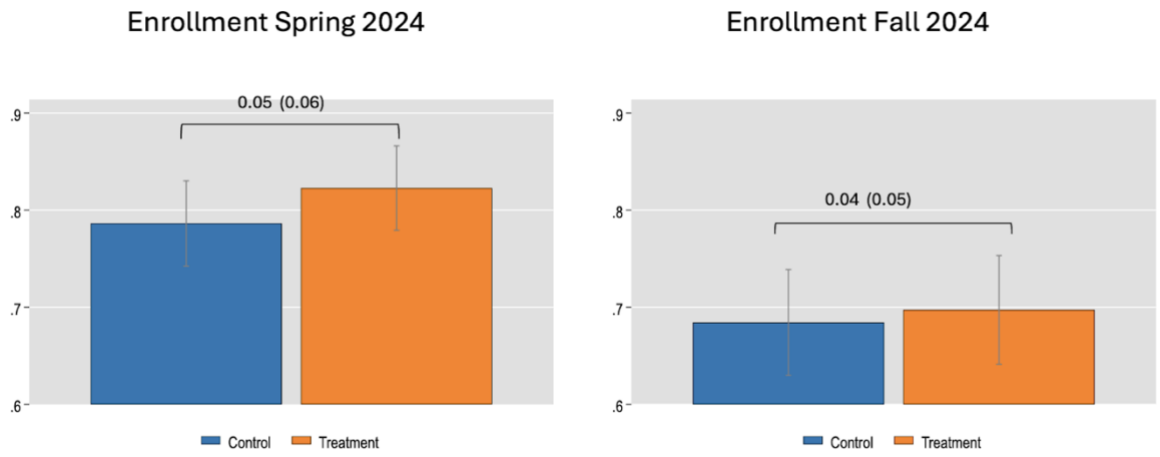
Note: This figure shows the math and the language test scores, the results on the scores, the attempts, and the accuracy. The right-hand side shows results on math, and the left-hand-side shows results on language.

Enrollment

Next Semester enrollment

Re-enrollment for the first semester of 2024 was 5% higher among students in the treatment group, but this result is not statistically significant. Enrollment for the second semester of 2024 also shows no significant difference between treated and control students. Figure 6 illustrates these results. The lack of significance stems from the limited statistical power of this hypothesis test. Considering the sample size, standard deviation, and evaluation design, the minimum detectable effect is 9%.

Figure 6: Next Semester Enrollment



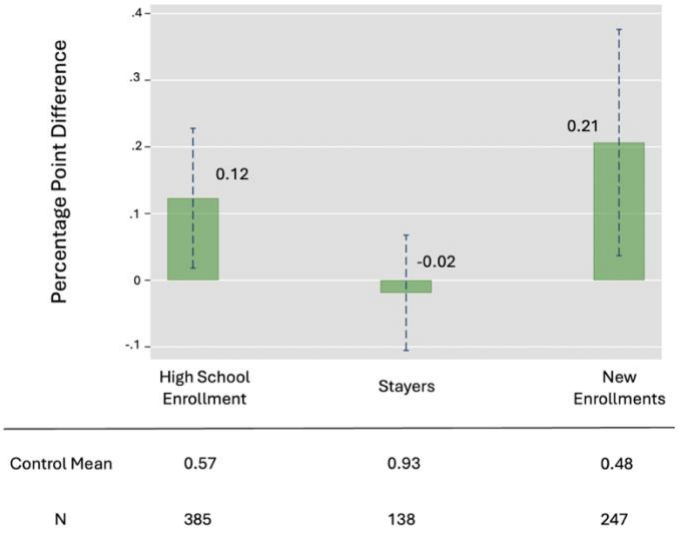
Note: This figure shows the next semester's enrollment by treatment status. Spring and Fall refer to the northern hemisphere.

High School Enrollment

Students participating in 8 out of the 13 centers that were part of the analysis can opt to take parallel high school courses for credit to obtain a high school diploma. Treated students enrolled in secondary school courses 12% more, on average, than control students, as shown in Column 1 of Figure 7.

Two different groups of students can explain this result. On one hand, students who were already taking high school courses might now decide to drop out. On the other hand, students who have decided to join this semester for the first. Indeed, those who were in the first semester when the intervention took place were not allowed to enroll in high school, as the first semester serves as an orientation period. Columns 2 and 3 of Figure 7 show that the 12% effect is attributed to the group of students enrolling in high school courses for the first time. The effect on new enrollments is 21 percentage points, whereas the effect on pre-enrolled students is not significantly different.

Figure 7: High School Enrollment



Note: This figure shows the average enrollment for the treatment and control groups in the first semester of 2024. The second column shows the effect conditioning on those who were already enrolled, and the third column show the result among those who are enrolling for the first time.

5. Mechanisms

There is substantial evidence in economics highlighting the impact of CBT on reducing crime (Heller et al. 2017, Blattman et al. 2017, and 2023), decreasing maternal depression, and increasing mothers' labor force participation (Baranova 2020), as well as raising income (Baker et al. 2024). Despite this evidence, it remains unclear what CBT is truly accomplishing: is it influencing behavior through the anticipated channels? Is it enhancing aspirations? Is it raising individuals' expectations? Or are there other factors unrelated to CBT, such as teachers' efforts in this context, that could be explaining the results?

In this section, we take a further step to address these questions. We examine how students' aspirations, expectations, and behaviors differ between treated and controlled students. Most importantly, I will discuss how these changes relate to the CBT curriculum we implemented.

Game 1: Returns on effort

The first incentivized game consists of three steps. In the first step, students are introduced to the puzzles and taught how to solve them. In the second step, students are asked two questions: (i) How

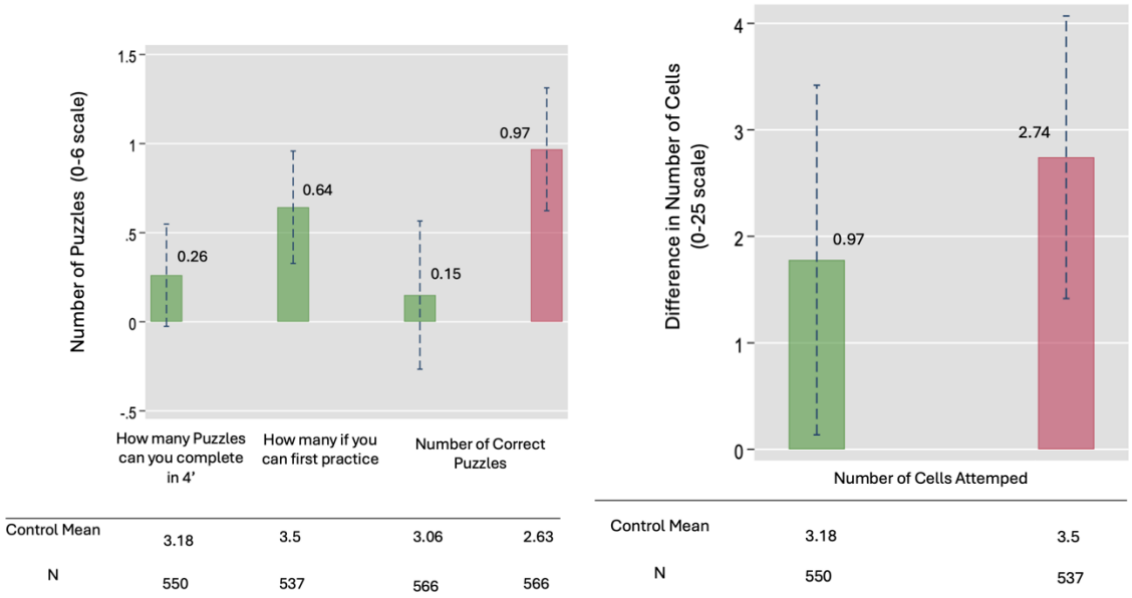
many puzzles do you think you can complete in four minutes? and (ii) How many puzzles do you think you can complete if you first practice for four minutes? In the third step, students are randomized into two groups: the first group has the chance to practice with the help of an instructor before attempting to solve six puzzles in four minutes, while the second group receives six puzzles to complete on their own.

According to the CBT curriculum, we expect to find that treated students believe they could do more puzzles, and we would expect an even larger effect for the returns on practice. The idea is that if a student holds the belief that she can't be successful, then there is no point in exerting effort because it will not enhance her performance. We would also expect that the difference in students' beliefs translates into a higher number of puzzles cells completed and higher correct puzzles. Finally, we would expect that the effect of practice complements the effect of treatment on a number of attempts and correct puzzles. Hence, we would expect that the number of attempted and correct puzzles is higher among treated students who practice, e.i., the interaction effect of practice and treatment is positive and significant.

The first column of Figure 8 shows that both treated and control students had similar expectations regarding the number of correctly completed puzzles. However, column 2 on Figure 8 indicates that treated students believed effort could help them complete more puzzles. This difference in perceived effort returns is statistically significant, reflecting that treated students have significantly different expectations only if they have the chance to practice. The third column of Figure 8 shows that treated students did not complete more correct puzzles than control students. Column 4 indicates that practice pays off; students who were assigned to practice completed one more correct puzzle than those who did not practice.

Figure 8 right-hand-side panel shows the effect of treatment and practice on the number of cells attempted. Both treatment and practice increase the number of effort students exert which in this context translates to the number of cells attempted. Treated students attempt one cell more than control students, and students who practice attempt almost three cells more.

Figure 8: Treatment and Practice effects on expectation, attempts and correct puzzles



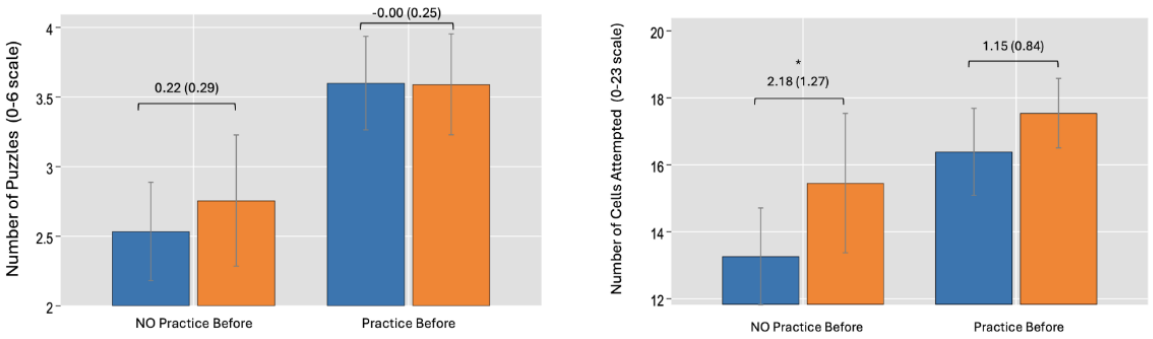
Note: This figure shows the effect of the intervention on the first incentivized task. The left-hand side is the effect on the number of puzzles completed; the right-hand side shows the effect on the number of cells completed correctly.

Figure 9 shows the interaction effect between treatment and practice. The left panel of Figure 9 shows treatment and control effects on number correct puzzles distinguishing between treated and control groups. There is no significant difference between the groups; treated and control students that did not practice have statistically the same score, and that is also true for those who did practice. As shown in the left panel of Figure 8, those who practiced performed better, but as we can see from Figure 9, the effect is the same for treated and control students.

The right-hand-side panel of Figure 9 shows the number of cells attempted by treated and control groups under the practice or no practice condition. For those who did not practice, treated students attempted a significantly higher number of cells, this is consistent with the fact that the intervention increases the willingness to practice and learn even when students are not helped to do so. Among those who did practice, treated students attempted 1.15 cells more, this difference is not statistically different.

To summarize, consistent with the intervention, treated students believe that effort can enhance higher results; they try more, but they do not get more puzzles correct, a results that was not expected.

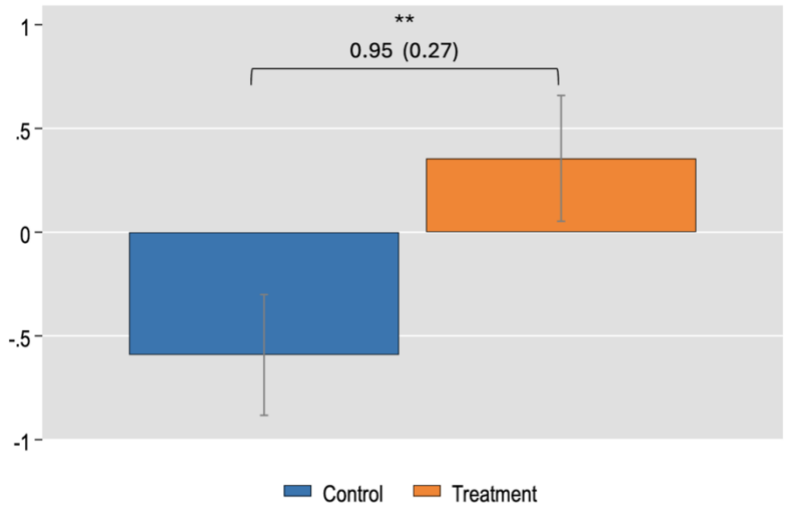
Figure 9: Practice and treatment effect interaction



Note: This figure shows the effect of the intervention on the number of puzzles completed and the number of cells attempted.

Finally, given that treated students have higher expectations of their performance after practicing, we test if this raise in expectation is well calibrated. Figure 10 shows the difference for those who practice between students' expected number of correct puzzles and the number of puzzles they got correct. Among treated students, the difference is positive, pointing out that treated students were overconfident in their performance; for students in the control group, the difference is negative, meaning that they are underconfident. Between treated and control students, treated students have expectations that are closer to zero, meaning that they were, in fact closer to the value that they achieved. These results implied that treated students won better prizes.

Figure 10: Expected number of Correct Puzzles = Actual Number of Correct Puzzles | Practice=1



Note: This figure shows the treatment effect over measures of confidence.

Game 2: Magnifying mistakes

This game examined whether failing a test discouraged students from trying again. The game consists of two steps. First, students complete two grammar tests, earning rewards for passing and for the number of correct answers. The passing score for the first test is randomized: some students need three out of six correct answers, while others require five out of six. In the second step, students get the chance to try a new set of six grammar questions again.

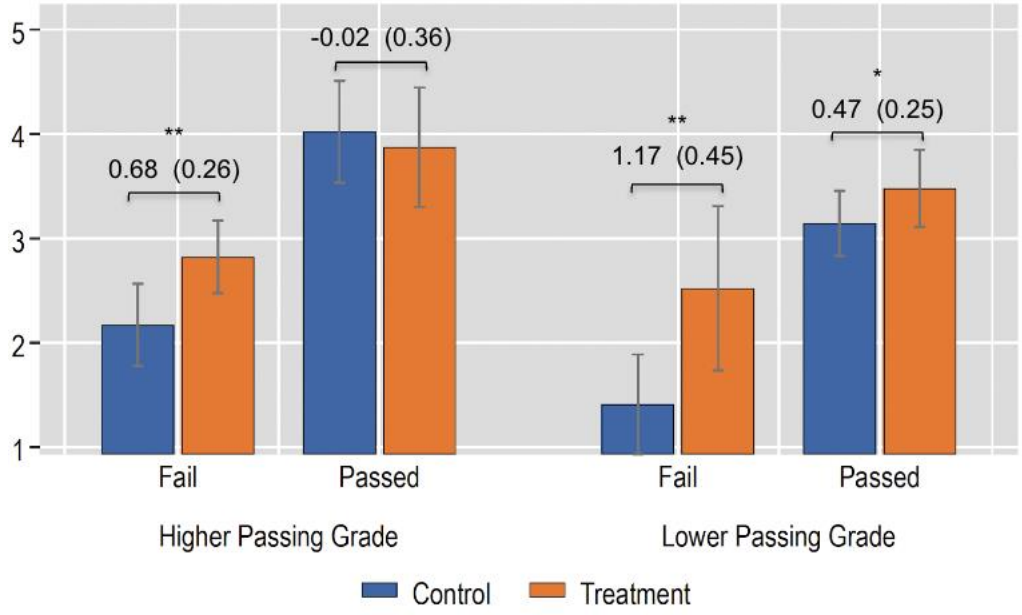
The scores in the first set of exercises are balanced between groups, and those randomized to a higher passing grade are 41% less likely to pass the test (Table A4). Students in the treatment group performed significantly more in the second set of questions; on average, treated students increased their scores by 0.6 more correct answers.

Figure 11 attempts to disentangle the 0.6 effect. In the left side panel, I present the effect among those who were randomized into a higher passing grade. The first two columns compare the average scores between the treated and control groups for those who failed the test with the higher passing grade. Treated students perform significantly better, answering 0.68 more questions correctly than the students in the control group. The next two columns display the results among those who were randomized to the higher passing grade and passed. The results are statistically insignificant. The

right side of Figure 11 illustrates the effect among those who were randomized to a lower passing grade. The first two columns indicate that treated students answered 1.17 more questions correctly than the students in the control group. The final two columns show that treated students who passed the lower passing grade also performed better, but by the smallest margin, answering 0.47 more questions correctly.

These results suggest that the increase in the number of correct answers by the students in the treatment group is almost exclusively explained by students who fail the first test.

Figure 11: Treatment effect by students who failed and passed the first set



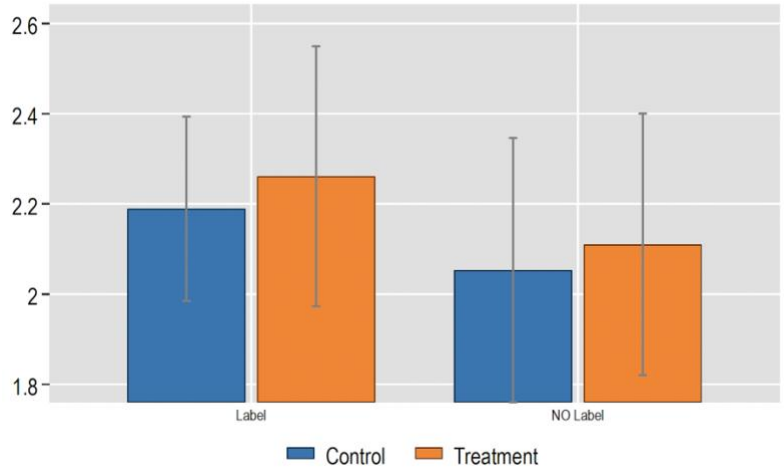
Note: This table shows the treatment effect on game 2. On the left-hand side panel, we show the results conditioning on students randomized to a higher passing grande in the first test. The side shows the results for those randomized to lower passing grades. Then, within each randomization, we distinguish between those who passed and those who failed.

Game 3: Label

In the third and last game, I prime students with a label to see if students who receive a label saying "attention: the future games are for math genius" prevent students from trying. The idea is that the label might trigger negative thoughts about the possibility of performing well. The game has three steps. In the first one, students have to do four substructions; in the second step, students are randomized to receive or not the label, and in the last step, students do a set of four substructions again.

The results suggest that students do not respond negatively to labels that imply limited ability. Instead, such labels motivate them to try harder. Figure 12 shows that both students in the treated and control groups improved their scores and performed better when received the label.

Figure 13: Treatment effects by label condition in the third game



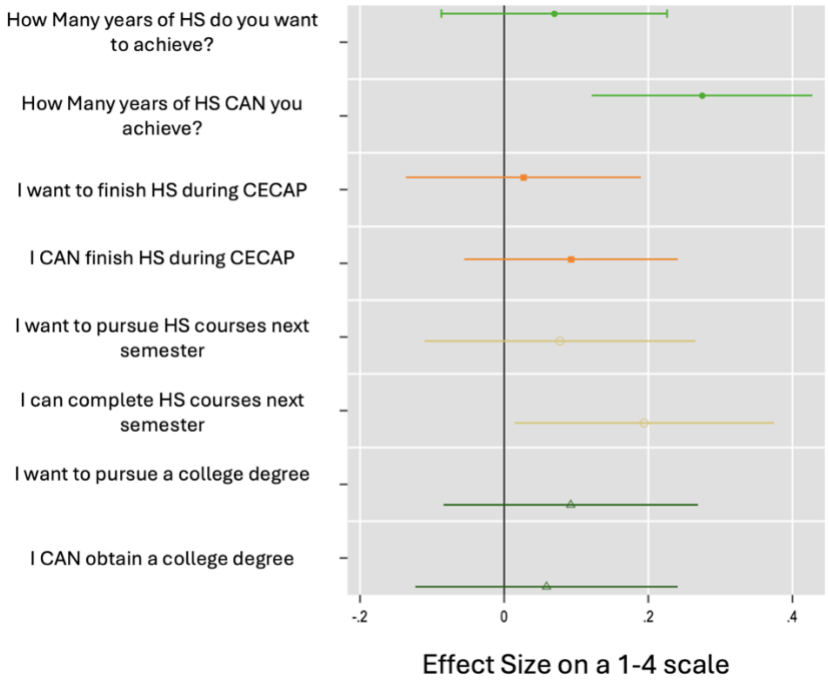
Note: This figures shows the treatment effect by label or no label cross-randomized group.

Aspirations and Expectations

We examine students' aspirations and expectations regarding their educational attainment. To measure aspirations, we asked students how many years of high school they aimed to complete, whether they intended to finish high school while enrolled in the program, if they planned to pursue high school courses in the following semester, and whether they aspired to obtain a college degree. Expectations, on the other hand, were assessed by asking students how confident they were in their ability to achieve the goals they set in their responses to the aspiration questions. This distinction allows us to analyze not only students' desired educational outcomes but also their perceived likelihood of achieving them.

Our findings (Figure 14) indicate that students in the treatment group did not experience an increase in their aspirations between baseline and endline. However, they did exhibit a significant rise in expectations regarding their educational trajectory. Specifically, they reported greater confidence in their ability to complete more years of schooling and a higher perceived likelihood of passing their high school courses in the following semester. These results suggest that while the intervention did not shift students' long-term educational goals, it did enhance their belief in their capacity to achieve their existing aspirations.

Figure 14: Treatment effect on Aspirations and Expectations



Note: This figure shows the treatment effect for aspiration and expectation measurements.

6. Discussion

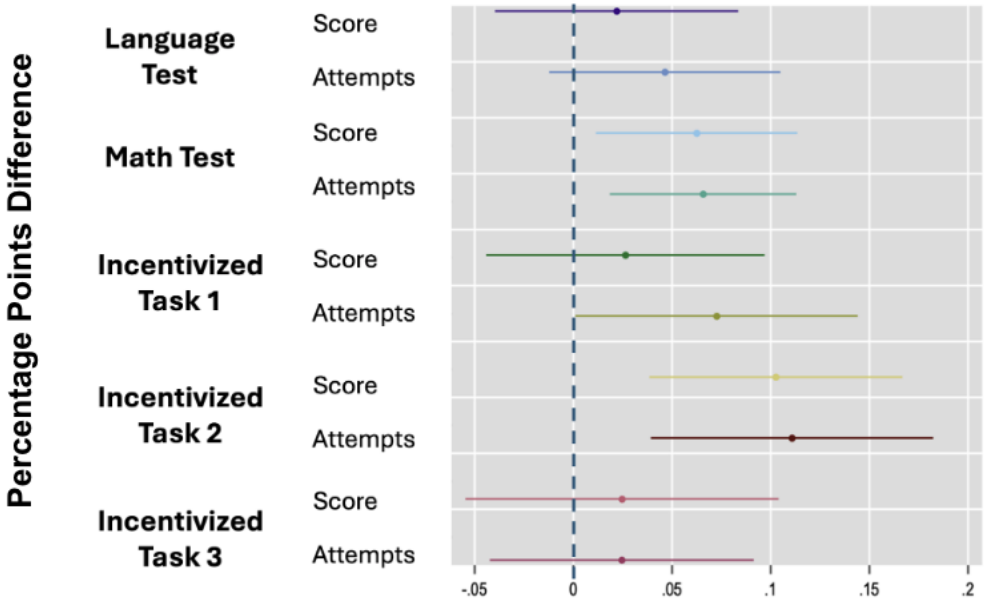
The results show that students who participated in the program performed better on arithmetic tests and, on average, were more inclined to enroll in secondary school courses in CECAP during the semester immediately following the intervention.

The incentivized games reflect that students who participated in the intervention believe, on average, that effort (practice) leads to better outcomes and are more resilient in the face of failure. Additionally, these students reported higher expectations of achieving the number of years of education they aspire to complete

The intervention aimed to counteract negative thoughts leading to pessimistic mental models regarding how effort translates into results. Observing these outcomes collectively, we see that students who participated in the program changed their behavior in ways consistent with a different perspective on how effort drives outcomes. Figure 14 demonstrates that, in almost all the activities proposed to students, those in the treatment group consistently exerted greater effort. While it is beyond the scope of this intervention to observe the students' thoughts as they occur or where they

manifest in their minds, the results suggest that the intervention influenced the students through the expected mechanisms.

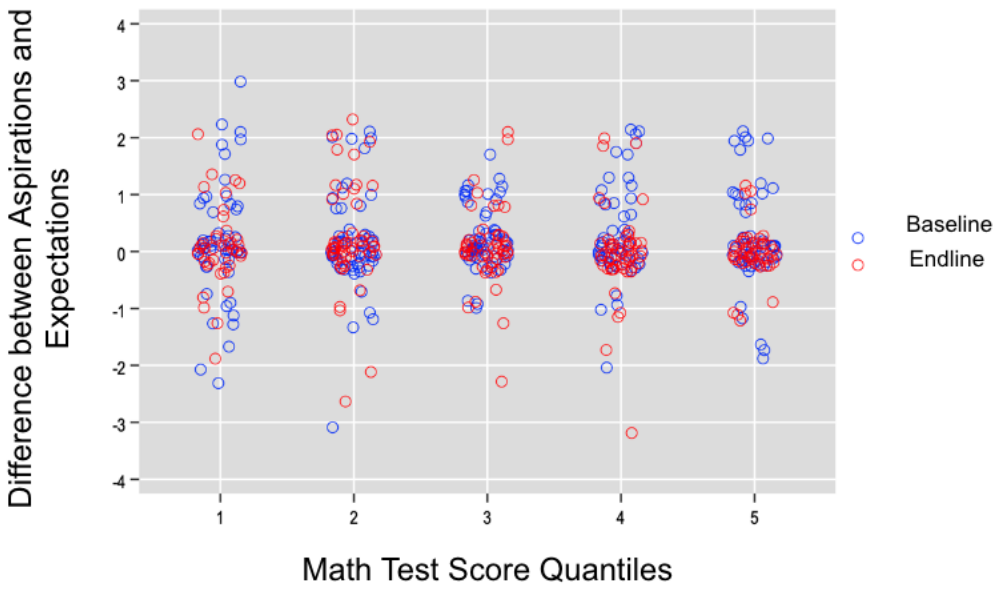
Figure 14: Treated students exert more effort across almost all tasks



Note: This figure shows the level of effort (attempts) and the results obtained for each activity in which the students participated. Each point refers to the parameter estimated based on the main equation, where the dependent variable is the final performance. The horizontal bars indicate 95% confidence intervals.

Finally, Figure 15 intends to understand who are the students changing due to the intervention. Figure 15 shows how students shorten or increase their aspiration and expectation gap between baseline and end line by the math test score quantiles at baseline. The y-axis is the difference between the score on a 1-4 scale to the questions: How many years of high school do you want to achieve? – How many years of school do you think you can achieve?, Positive results reflect students who have higher aspirations to what they think they can achieve. The red bubbles show the same measurement at endline. The figure shows that the number of red bubbles in the upper red corner diminishes from endline to baseline. This implies that students who performed better at baseline, those who are potentially the highest ability students, are the ones who shorten the aspiration-expectation gap.

Figure 15: Heterogeneous impact of the intervention



Note: This figure shows the heterogeneous effect based on baseline mathematics performance. The blue scatter plots represent baseline data showing the relationship between students' math levels and differences between expectations and aspirations (outcome variable). In red, the same relationship is shown at the endline.

7. Conclusion

This study demonstrates that an intervention focused on changing negative thoughts and pessimistic mental models regarding the production function of effort successfully increases students' effort. This increase in effort translated into improved performance on math tests and in the games designed to understand mechanisms. Students raised their expectations regarding the years of secondary education they expect to complete and significantly increased their participation in parallel secondary courses. Finally, the incentivized games suggest that the behavior of students who participated in the program changed consistently with a mental model that reflects a new way of perceiving the relationship between effort and results.

Future work should emphasize understanding how much of this result stems from an on-the-spot change in thinking versus a purely planning effect. Future research should also attempt to understand the origins of pessimistic mental models, including the roles of context, teachers, the institutions students frequent, and parents in shaping absolutist beliefs (e.g., "you're not smart enough").

In a context where dropout, low aspirations, and hopelessness are the norm, this intervention provides promising evidence that addressing internal constraints can play a crucial role in improving student outcomes. The observed behavioral changes align with the principles of

cognitive behavioral therapy (CBT), suggesting that CBT can be a powerful tool for reshaping students' beliefs and actions. Importantly, the findings highlight the need for future interventions to integrate CBT with concrete, goal-oriented thinking to maximize its impact.

Reference

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

Almlund, Mathilde, Angela L. Duckworth, James J. Heckman, and Tim D. Kautz, "Personality Psychology and Economics," In E. Hanushek, S. Machin, and L. Woessman, eds., Handbook of the Economics of Education, (2011), 1–181.

Castillo, Marco, Paul J. Ferraro, Jeffrey L. Jordan, and Ragan Petrie, "The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children," Journal of Public Economics, 95 (2011), 1377–1385.

Dalton, P. S., Ghosal, S., & Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, *126*(590), 165-188.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly, "Grit: Perseverance and Passion for Long-Term Goals," Journal of Personality and Social Psychology, 92 (2007), 1087–1101.

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

Dweck, Carol, Mindset: The New Psychology of Success. (New York, NY: Random House 2006)

Duflo, E., Glennerster, R. & Kremer, M. Chapter 61 Using Randomization in Development Economics Research: A Toolkit. *Handb. Dev. Econ.* 4, 3895–3962 (2007).

Duflo, E., Kremer, M. & Robinson, J. Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *Am. Econ. Rev.* 101, (2011).

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, J¨urgen Schupp, and Gert G. Wagner, "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences," Journal of the European Economic Association, 9 (2011), 522–550.

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

Genicot, G., & Ray, D. (2017). Aspirations and inequality. *Econometrica*, *85*(2), 489-519.

Guyon, N., & Huillery, E. (2021). Biased aspirations and social inequality at school: Evidence from french teenagers. *The Economic Journal*, *131*(634), 745-796.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," Journal of Labor Economics, 24 (2006), 411–482.

Hoxby, C. M., & Avery, C. (2012). *The missing" one-offs": The hidden supply of high-achieving, low income students* (No. w18586). National Bureau of Economic Research.

Jensen, R. The (Perceived) Returns to Education and the Demand for Schooling. *Q. J. Econ.* 125, 515–548 (2010).

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, *5*(10), 434-442.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans, "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success," (Paris, France: Organization for Economic Co-operation and Development 2014).

Langener, A. M., Kramer, A. W., van den Bos, W., & Huizenga, H. M. (2022). A shortened version of Raven's standard progressive matrices for children and adolescents. *British Journal of Developmental Psychology*, *40*(1), 35-45.

Outes-Leon, I., Sánchez, A., & Vakis, R. (2020). The power of believing you can get smarter: The impact of a growth-mindset intervention on academic achievement in Peru. *World Bank Policy Research Working Paper*, (9141).

Raven, John, Jean Raven, and John H. Court, "Manual for Raven's Progressive Matrices and Vocabulary Scales," San Antonio, TX: Harcourt Assessment, (2004).

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, *370*(6522), eaay0214.

Schofield, H., & Mullainathan, S. (2008). The psychology of nutrition messages. In *Beyond Health Insurance: Public Policy to Improve Health* (Vol. 19, pp. 145-172). Emerald Group Publishing Limited.

Schilbach, F., Schofield, H., & Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review*, *106*(5), 435-440.

Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological science*, *26*(4), 402-412.

Yeager, D.S., and C.S. Dweck, "Mindsets that Promote Resilience: When Students Believe that Personal Characteristics Can Be Developed," Educational Psychologist, 47 (2012), 302–314.

Yeager, D.S., P. Hanselman, G.M. Walton, R. Crosnoe, C. Muller, E. Tipton, B. Schneider, C. Hulleman et al., "Where and For Whom Can a Brief, Scalable Mindset Intervention Improve Adolescents' Educational Trajectories?" Unpublished Manuscript 2018

Yeager, David S., Rebecca Johnson, Brian J. Spitzer, Kali H. Trzesniewski, Joseph Powers, and Carol S. Dweck, "The Far-Reaching Effects of Believing People Can Change: Implicit Theories of Personality Shape Stress, Health, and Achievement During Adolescence," Journal of Personality and Social Psychology, 106 (2014), 867–884.

Yeager, David S., Hae Y. Lee, and Jeremy Jamieson, "How to Improve Adolescent Stress Responses: Insights From an Integration of Implicit Theories of Biopsychosocial Models," Psychological Science, 27 (2016a), 1078–1091.

Yeager, David S., Carissa Romero, Dave Paunesku, Christopher S. Hulleman, Barbara Schneider, Cintia Hinojosa, Hae Y. Lee, Joseph O'Brien et al., "Using Design Thinking to Improve Psychological Interventions: The Case of the Growth Mindset During the Transition to High School," Journal of Educational Psychology, 108 (2016b), 374–391.

Yeager, David S., and Gregory M. Walton, "Social-Psychological Interventions in Education: They're Not Magic," Review of Educational Research, 81 (2011), 267–301.

Appendix Tables and Figures

Table A1: Attrition

	attrition			attrition		
	coefficeint	se	p-value	coefficeint	se	p-value
treat	0.0203731	0.02925368	0.48639133			
p8_treat				-0.1294	0.0557	0.0208
p9_treat				0.1209	0.0620	0.0522
p10_treat				-0.1004	0.0653	0.1251
p11_treat				0.0259	0.0706	0.7140
p12_treat				0.0354	0.0694	0.6103
p13_treat				0.0043	0.0704	0.9511
p14_treat				-0.0142	0.0718	0.8431
p15_treat				0.0593	0.0662	0.3707
p16_treat				-0.0576	0.0630	0.3609
p17_treat				-0.0229	0.0767	0.7661
p18_treat				0.1005	0.0637	0.1156
p19_treat				0.0024	0.0647	0.9710
self_efficacy_base2_treat				-0.0114	0.0155	0.4611
mindset_base_treat				0.0748	0.0549	0.1736
grit_base_treat				0.0182	0.1229	0.8823
matrices_score_treat				0.0012	0.0187	0.9489
matrices_cantidades_treat				-0.0237	0.0220	0.2821
tecle_score_treat				0.0109	0.0084	0.1973
tecle_cantidades_treat				-0.0127	0.0089	0.1556
sumas_score_treat				0.0078	0.0178	0.6629
sumas_cantidades_treat				-0.0091	0.0177	0.6094
restas_score_treat				-0.0064	0.0060	0.2885
restas_cantidades_treat				0.0063	0.0063	0.3198
multi_score_treat				0.0052	0.0113	0.6473
multi_cant_treat				-0.0101	0.0103	0.3281
male_treat				0.0555	0.0915	0.5444
schooling_treat				0.0323	0.0398	0.4167
age_treat				0.0135	0.0249	0.5877
p8				0.0340	0.0419	0.4175
p9				0.0233	0.0445	0.6001
p10				0.0848	0.0376	0.0251
p11				-0.0397	0.0496	0.4241
p12				-0.0540	0.0518	0.2978
p13				0.0051	0.0516	0.9220
p14				-0.0305	0.0535	0.5690
p15				-0.0041	0.0485	0.9328

p16		0.0263	0.0409	0.5200
p17		-0.0060	0.0482	0.9006
p18		-0.1197	0.0452	0.0086
p19		0.1108	0.0470	0.0191
self_efficacy_base2		0.0010	0.0106	0.9256
mindset_base		-0.0392	0.0364	0.2816
grit_base		0.0227	0.0785	0.7724
matrices_score		0.0074	0.0134	0.5808
matrices_cantidades		-0.0115	0.0138	0.4039
tecle_score		-0.0004	0.0059	0.9414
tecle_cantidades		0.0017	0.0061	0.7818
sumas_score		0.0028	0.0109	0.7957
sumas_cantidades		-0.0024	0.0116	0.8333
restas_score		0.0060	0.0048	0.2074
restas_cantidades		-0.0029	0.0042	0.4987
multi_score		-0.0029	0.0080	0.7126
multi_cant		0.0024	0.0074	0.7396
male		-0.0140	0.0605	0.8173
schooling		0.0060	0.0276	0.8285
age		-0.0063	0.0185	0.7326
N	703	703		

Table A4: game 2 results

	Score 1st set	Score 2nd set	Pass 1st set	Score 2nd set
Treat	0.14	0.60***		
	(0.16)	(0.19)		
Higher Passing Grade			-0.41***	
			(0.04)	
Pass 1st set				1.21***
				(0.14)
Control mean	3.64		0.76	2.34
N	566		566	566

A5: Returns on effort Table

	How many you think you can complete in 4 minute?	How many you think you can complete in 4 minute if you have the change to practice first?	Number of Correct Sudokus	Number Attempted	Number of Correct Sudokus	Number Attempted
Treat	0.26	0.64***	0.150	1.78**		
	0.15	0.15	0.21	0.82		
Practice					0.97***	2.74***
					0.17	0.67
Control mean	3.18	3.50	3.06	14.81	2.63	14.18
N	550	537	566	567	566	567

	No Practice		**Practice**	
	Number of Correct Sudokus	Number Attempted	Number of Correct Sudokus	Number Attempted
Treat	0.22	2.2*	-0.01	1.16
	0.30	1.27	0.25	0.84
Control mean	2.53	13.26	3.60	16.38
N	274	274	292	293