Growth-mindset interventions at Scale:

Experimental Evidence from Argentina*

Alejandro J. Ganimian†

Abstract: This study presents one of the first evaluations of a "growth mindset" intervention at scale in a developing country. I randomly assigned 202 public secondary schools in the Province of Salta, Argentina to a treatment group in which staff from the ministry of education invited grade 12 students to read a text about the malleability of intelligence, write a letter to a classmate about its main lessons, and post their letters in their classroom, or to a business-as-usual control group. I verify that the intervention was implemented as intended in 90% of treatment schools. Yet, I find no evidence that it led students to find challenging tasks less intimidating. It had a null effect on students' perceptions of the difficulty of schoolwork, their self-efficacy in math and

---

* I gratefully acknowledge the funding provided by the Abdul Latif Jameel Poverty Action Post-Primary Education initiative and the Inter-American Development Bank for this study. I thank Andrea Bergamaschi, Analía Berruezo, Loreto Biehl, Gloria Crespo, Elena Duro, Miriam Goldszier, and Gabriela Guerrero for making this study possible. I also thank Hunt Allcott, Felipe Barrera-Osorio, Clancy Blair, Susana Claro, Sean Corcoran, Andy de Barros, Rajeev Dehejia, Bill Easterly, Raquel Fernández, Jill Gandhi, Isaac Mbiti, Karthik Muralidharan, Dick Murnane, Cybele Raver, Martin Rotemberg, David Yeager, Hiro Yoshikawa, Tyler Watts, Marty West, seminar participants at NYU, Paco Martorell and two anonymous reviewers for comments that informed this draft. Aditi Bhowmick, Nicolás Buchbinder, and María Cortelezzi provided excellent research assistance. This study was registered with the AEA Trial Registry (RCT ID: AEARCTR-0002970). It was approved by the Institutional Review Boards of the Massachusetts Institute of Technology. All views expressed are my own and not of any institution with which I am affiliated.

† Assistant Professor of Applied Psychology and Economics, New York University Steinhardt School of Culture, Education, and Human Development. E-mail: alejandro.ganimian@nyu.edu.

language, and their views on the usefulness of classroom tests. It did not increase student effort in tasks related to school (e.g., attendance), personal development (e.g., reading books), or work. It did not improve school climate, including relationships between peers, bullying, or student vandalism. Consistent with these results, the intervention had no impact on students' performance in school (e.g., passing, repetition, or dropout rates), achievement in the national assessment, or plans to pursue post-secondary education. In nearly all outcomes, I can rule out even small effects and find little evidence of heterogeneity by student characteristics (sex, socio-economic status, or prior grade repetition) and school characteristics (prior achievement, resources, and supports for low-performing students). This study suggests that the benefits of growth-mindset interventions may be more challenging to replicate and scale in developing countries than anticipated.

1. Introduction

There is mounting evidence indicating that the expectations of children and youths and their parents about the payoff from schooling influences their educational investments. These beliefs affect whether (and for how long) children go to school, the type of schools that they choose (e.g., public or private and academic or vocational), whether they invest in complements to schooling (e.g., tuition), and how much effort they exert in school (see Banerjee, Glewwe, Powers, & Wasserman, 2013).

Experiments in developing countries have found that low-income families often hold beliefs that lead them to underinvest in schooling, but that they adjust their behavior when provided with information. Most studies have explored the effects of information on returns to schooling (Avitabile & de Hoyos, 2014; Berniell, 2014; Bonilla, Bottan, & Ham, 2016; Nguyen, 2009) and school quality (Andrabi, Das, & Khwaja, 2017; Camargo, Camelo, Firpo, & Ponczek, 2018; Loyalka et al., 2013).

Providing information on child ability may have a larger effect on human capital accumulation by affecting not only demand for schooling, but potentially also student motivation and effort. It is also more likely to impact equity by correcting parental biases (e.g., about boys and girls). The few studies that provided this type of information have used objective measures of ability (i.e., test scores) (see Barrera-Osorio, Deming, González, & Lagos, 2020; Bobba & Frisancho, 2016; Dizon-Ross, 2019). The main advantage of this approach is that it conveys individual-level information, making the intensity of the treatment inversely proportional to the gap between expected and actual child ability, increasing its chances of affecting those who need it most. Its main drawback, however, is that it does not account for the fact that these ability measures are partly a function of past educational investments (which may themselves be based on incorrect beliefs), so that the information could reinforce the inefficient and inequitable investments it seeks to address.

An alternative is to inform children of their *potential*—rather than their current—ability. A team of psychologists in the United States has designed an intervention with this objective. It asks students to read a short passage that synthesizes research showing that exposure to stimulating environments and practice at challenging tasks can help develop one's intelligence, much like setting ambitious exercise goals and working out at the gym can grow one's muscles. The reading is followed by a brief exercise so that students can internalize this main message. The intervention is based on a large body of research indicating that individuals' beliefs about whether intelligence is fixed or malleable influence their effort, and in turn, their performance (for reviews of this literature, see Dweck & Leggett, 1988; Dweck, Walton, & Cohen, 2014; Dweck & Yeager, 2019). Variations of this "growth mindset" intervention have improved self-beliefs (Aronson, Fried, & Good, 2002), school performance (Good, Aronson, & Inzlicht, 2003; Paunesku, Yeager, Romero, & Walton, 2015), achievement (Blackwell, Trzesniewski, & Dweck, 2007), health (Yeager et al., 2014), and peer relations (Yeager, Miu, Powers, & Dweck, 2013; Yeager, Trzesniewski, Tirri, Nokelainen, & Dweck, 2011).

This paper presents one of the first studies of this intervention at scale in a developing country. I randomly assigned 202 public secondary schools in the Province of Salta, Argentina to a "treatment" group, in which representatives from the ministry of education visited schools, invited grade 12 students to read the passage described above, write a letter to a classmate about how to apply its lessons to their own lives, and put up their letters next to a poster on one of the classroom walls, or to a "control" group that did not implement the intervention. The intervention was conducted during a non-academic period in which students discuss school-related matters with their teacher, so this study assesses whether using this time for this activity had an effect on students' beliefs, effort, performance in school, and achievement. I can verify that the intervention was implemented as intended in 90% of the treatment group using either pictures taken by implementers (83%) or confirmations from principals (7%).

I report five sets of results. First, I find no evidence that the intervention led students to find challenging tasks less intimidating. I show that the intervention had a precisely estimated null effect on students' perceptions of the difficulty of schoolwork, their self-efficacy, and the usefulness of classroom tests. In fact, I find that it may have had a negative effect on female students (increasing their perception of schoolwork as difficult and decreasing their self-efficacy), students from low-income families (decreasing their self-efficacy), and those who had repeated a grade (lowering their propensity to see classroom assessments as useful).

Second, I find no evidence that the intervention increased student effort in school-related tasks (e.g., going to school, attending private tuition), personal development (e.g., reading books, learning languages, playing sports), or existing obligations (i.e., work at or outside the home). I can even rule out small effects in all of these outcomes and find no heterogeneous effects.

Third, I find no evidence that the intervention improved school climate, including relationships between peers, bullying, or student vandalism (i.e., stealing and damaging of school property). In fact, some evidence suggests that it might have had a negative effect on female students (decreasing their propensity to get along with peers).

Fourth, consistent with these null results, I find that the intervention had no impact on students' school performance (e.g., passing, repetition, and dropout rates), their achievement in the national assessment of math and language, or plans to pursue post-secondary education. I even find some indication that it had a negative effect on students' aspirations in schools with higher levels of achievement, resources, and supports for low-performing students.

Finally, I find that the intervention is relatively inexpensive at a cost of USD 2.82 per student, but that it is considerably costlier than suggested by prior studies in developing countries. The main reason for the discrepancy stems from including the cost of training implementers, which had been avoided in a prior study by directly shipping intervention packages to schools. Most school systems are unlikely to deliver an intervention without training for implementers, so my figures seem to be more representative of the actual cost of this intervention at scale.

This study makes several key contributions to research on the growth-mindset intervention. To put these contributions in context, I conducted a detailed review of prior randomized evaluations of this intervention in both developed and developing countries (see Appendix B).

My review indicates that this is the first study that can rule out small positive effects from the intervention on mechanisms (e.g., student beliefs and effort) and outcomes (e.g., achievement). Several impact evaluations had previously found that the intervention had null or mixed effects (see, for example, Burnette, Russell, Hoyt, Orvidas, & Widman, 2018; Dommett, Devonshire, Sewter, & Greenfield, 2013; Gandhi, Watts, Masucci, & Raver, 2019; Sriram, 2014). Yet, none of them was designed to distinguish between precisely estimated null effects and statistically insignificant but imprecise results. This is a major contribution of the present study because it demonstrates that the intervention does not always have the effects seen in efficacy trials, even if its materials are standardized and it can be implemented in one brief session with relatively little adult supervision.

According to my review, this is also one of the first evaluations of the intervention at scale. Until recently, it had been assessed through efficacy trials with small convenience samples. This approach has been instrumental in ensuring the intervention was implemented faithfully, establishing its proof of concept, and carefully measuring its potential mechanisms of impact, but it has been less helpful in understanding its effectiveness at scale within the school system. In recent years, there have been two large-scale randomized evaluations of this intervention. Outes, Sánchez, and Vakis (2020) evaluated it in 800 public secondary schools in three regions of Peru and Yeager et al. (2019) in 65 public secondary schools across the United States. The differences in the context, implementation, and measurement between these studies raise useful questions about the intervention that can inform future research and policy decisions.

Lastly, my review indicates that the sampling, randomization, and data collection strategies in this study are uniquely positioned to assess the effectiveness of this intervention at scale. First, its sample included nearly all secondary public schools of a (sub-national) school system. This approach circumvents the problems of site selection bias present in most prior studies and allows me to understand the effect of the intervention where it is not necessarily welcomed. Second, its school-level randomization avoids the spillovers of student-level randomization and allows me to estimate the impact of the intervention when it is conducted by an entire school. Third, its reliance on administrative data collected by the school system (mainly, through the annual census of schools and national student assessment) has multiple benefits: it minimizes the risk of differential attrition in data collection across experimental groups, it reduces the risk of social-desirability bias that may emerge when implementers collect data on the measures that their own intervention is designed to influence, and it enables me to measure its effect on a wide array of outcomes that it had previously been found to affect.

The rest of this paper is structured as follows. Section 2 presents the context, study design, and intervention. Section 3 describes the data. Section 4 discusses the empirical strategy. Section 5 reports the results. Section 6 discusses implications for research and policy.

2. Experiment

2.1 Context

Schooling in Argentina is compulsory and free from age 4 until the end of secondary school. In 12 out the 24 provinces including Salta, primary education runs from grades 1 to 7 and secondary education from grades 8 to 12 (DiNIECE, 2013). The Argentine school system serves 11.4 million students: 1.8 million in pre-school, 4.8 million in primary school, and 3.7 million in secondary school (DiEE, 2016). The school year runs from February to December.

Argentina enrolls a larger share of youths in secondary school than most Latin American countries: by the late 2000s, 75% of its youths had started secondary school at the appropriate age, compared to 59% in the average country in the region (Bassi, Busso, & Munoz, 2013). Yet, its graduation rate at this level lags behind those of its upper-middle income neighbors: in 2016, it stood at 63%, compared to 65% in Brazil, 91% in Chile, and 77% in Colombia (OECD, 2018). Further, the relative standing of its students in the region has deteriorated: in 2012, Argentina was among the eight lowest performing school systems in all three subjects of the Program for International Student Assessment (PISA), while countries like Brazil, Chile, and Peru had improved and had either caught up with it or surpassed it (OECD, 2013). Many of its students fail to meet national standards. In 2017, 69% of grade 12 students scored in the lowest two of the four levels of the national assessment in math (below basic, basic, satisfactory, and advanced) and 45% did so in reading (SEE-MEDN, 2018a).

The Province of Salta is the eighth-largest sub-national secondary school system in Argentina: in 2016, it served 125,207 students across 394 schools at that level (DiEE, 2016). It is also one of the lower-performing systems: in 2017, 73% of grade 12 students scored in the lowest two levels of the national test in math and 45% did so in reading (SEE-MEDN, 2018c).[1]

2.2 Sample

The sample for the study includes 202 public secondary schools in urban and semi-urban areas of the Province of Salta. I arrived at this sample as follows. First, of the 334 secondary schools in the province, I excluded all 94 private schools because I was interested in the potential of the intervention to impact public schools. Then, I dropped all 26 schools in rural areas because they are spread across the province, which would have limited the capacity of the local ministry of education to implement the interventions. (Note, however, that while rural schools account for 7.8% of all public schools in Salta, they only serve 1.2% of students in the province). Finally, I excluded 12 public schools in urban and semi-urban areas with fewer than 10 students in grade 12 (the target grade of the intervention) to minimize sampling error from small schools.

The schools in the sample are different from out-of-sample schools, regardless of whether I compare them to all out-of-sample schools, public out-of-sample schools, or public and urban or semi-urban out-of-sample schools (Table A.1, appendix A). Specifically, in-sample schools are larger and have higher repetition rates than all three groups of out-of-sample schools. They also have slightly higher dropout rates across secondary school than the first two groups of out-of-sample schools and slightly lower dropout rates in grade 12 than the last group.

In-sample schools had lower results on the 2016 national student assessment when compared to all out-of-sample schools, but they performed on par with public out-of-sample schools, except in math (Table A.2). The mixed results in the comparisons with public and urban or semi-urban out-of-sample schools may be related to the small number of schools in this group.

2.3 Randomization

I randomly assigned the 202 public secondary schools in the sample to: (a) a "treatment" group that was offered an intervention (described in the next section); or (b) a "control" group that was not offered the intervention. I stratified the randomization by geographic location (i.e., whether schools were urban or semi-urban) and the school type (i.e., whether schools were "common" or "technical") to increase statistical power. This procedure resulted in 102 treatment and 100 control schools.

Control and treatment schools were comparable on all indicators of school performance tracked by the school system (Table A.3). I find no statistically differences on any indicators in grade 12, the target grade for the intervention (Panel B), but when I consider all students enrolled in secondary education at these schools (i.e., grades 8 to 12), treatment schools appear to be smaller and have slightly lower repetition rates (Panel A). I test whether these differences matter by accounting for school-level averages of these indicators in my impact estimation.

2.4 Intervention

The growth-mindset intervention administered in Salta was a single-session adaptation of a multi-session version evaluated in the United States (Blackwell et al., 2007).[2] In September of 2017, schools assigned to the treatment group were visited by a representative from the Ministry of Education, Science, and Technology (MECyT) of Salta (locally known as an *Asistente Técnico Territorial* or ATT).[3] The ATT then visited each grade 12 classroom at the school and proceeded as follows. First, he/she explained the purpose of the activity and sought informed consent from all students (students who chose not to participate were allowed to complete schoolwork in silence). Then, the ATT asked all students who agreed to participate to read a passage on how persisting through difficult challenges can develop the brain and write a letter to a classmate of their choice on the three main lessons from the reading and how they might help him/her.[4] Next, the ATT put up a poster in the classroom with all the letters around it to remind students of the activity for the rest of the school year.[5] Finally, the ATT took a picture of the poster and letters and shared it with the MECyT to verify that the intervention was implemented.

The intervention was scheduled to take place during a non-academic period called *tutorías*, which allow students to bring a wide array of concerns to a designated teacher (*tutor*). It is part of the official curriculum of Salta and of most provinces in Argentina (MECyT, 2012). *Tutorías* cover issues such as student-teacher relations, student body government, or bullying. This study assesses whether using this period for this activity has a positive effect on students. Importantly, *tutores* were not required to be in the classroom during the intervention. The MECyT kindly agreed to purposefully time the delivery of the intervention two months before the national assessment because prior studies had found effects of a similar intervention, also administered in a single session during *tutorías*, over this time frame (see Outes et al., 2020).

The reading consists of three parts. The first part seeks to convey the message that, when individuals practice and learn, their brain grows in a similar fashion to muscles after exercise. It explains that the brain is made up of neurons, that connections between neurons allow for problem solving, and that when individuals learning something these connections multiply. The second part describes research on humans and animals that supports the initial message. It also shows photos of neural connectivity for animals with and without access to stimulating environments and for humans at birth and age 6 to illustrate the point from the prior section. The third part contends that, if intelligence can grow through practice at challenging tasks, it makes little or no sense to categorize individuals using labels such as “dumb” or “smart”. Then, it concludes by encouraging the reader to engage in practice, even when it seems hard. The reading had been developed for grade 7 students in New York City. I conducted a pilot in August of 2016 with 15 out-of-sample grade 12 students to check that they could understand the text. I did not make context-specific adaptations, as the developers of the intervention have done in new settings (e.g., Bettinger, Ludvigsen, Rege, Solli, & Yeager, 2018), to prevent any adjustments I introduced from dampening the effect of an otherwise seemingly effective intervention.

Table 1 shows the theory of change of the intervention, which outlines the hypothesized causal chain linking the intervention to its expected effect. The need that the intervention aims to address is that many students believe that intelligence is static, which leads them to want to look smart and thus engage in a series of counterproductive behaviors that ultimate confirm their deterministic worldview (Dweck & Leggett, 1988). The prevalence of this belief and its association with student achievement has been documented in a variety of settings, including many similar to the one that I study (Chaia et al., 2017; Claro, Paunesku, & Dweck, 2016).

I had hypothesized that the intervention would have five main effects. First, students would feel less intimidated by challenging tasks. Students could start perceiving challenging tasks as less difficult (because they anticipated the cognitive gains to be derived from attempting them), they could feel more capable of tackling these tasks (because they believed that, if they persisted, they would eventually solve them), or they could perceive the tasks as a formative experience (as part of the learning process).[6] If students felt less intimidated by challenges, they would exert more effort.[7] This increase in effort could manifest itself in schoolwork, but it could also emerge in other aspects of students’ lives, such as their personal development and even existing obligations. Third, the change in mindset could lead students to improve their relationships with peers (by decreasing the threat that they had previously felt from the success of others).[8] Fourth, these changes would lead to improved school performance and achievement, in turn raising students' aspirations to pursue post-secondary education.[9] And ultimately, these improvements could lead students to want to pursue post-secondary education.[10]

Prior theoretical and empirical work also suggested that the effects of the intervention could differ based on student- and school-level characteristics. First, the effects on the outcomes above could be larger for students who are more likely to be the subject of stereotypes, including students who are female, from low-income families, and/or who struggle at school.[11] Second, based on recent experiments, the effect of the intervention could vary across schools.[12] Yet, it is not clear which school characteristics would predict treatment heterogeneity. I hypothesized that the effect of the intervention may differ by schools' prior achievement, instructional resources, and supports for low-performing students.[13]

2.5 Costs

Part of the increasing enthusiasm for growth-mindset interventions stems from the fact that they can be administered in one session, with little supervision, and are relatively inexpensive. This is certainly the case in developed countries like the U.S., where it can be delivered online (see, for example, Gandhi et al., 2019; Paunesku et al., 2015; Yeager et al., 2019; Yeager et al., 2016). The potential for deploying these interventions at a low cost is arguably even larger in higher education, where students already regularly interact with instructors online (Oreopoulos, Brown, & Lavecchia, 2017; Oreopoulos, Patterson, Petronijevic, & Pope, 2018; Oreopoulos & Petronijevic, 2018).

Yet, there is little information about the costs of this intervention in developing countries, where schools lack computers and internet and the intervention must be conducted on paper. In

Peru, where the ministry of education simply shipped the intervention packets to schools, Outes et al. (2020) estimated intervention costs to be only USD 0.2 per student. However, this setup is unlikely to lead schools to implement the intervention in other settings, where teachers are less willing or able to follow instructions without any training or support.

I calculated the costs of administering the intervention by training ministry staff, a model that is more likely to be accepted by education authorities and teachers in developing countries. Specifically, I did so using the ingredients method explained in detail in Dhaliwal, Duflo, Glennerster, and Tulloch (2012). According to those calculations, the total cost of the intervention in Salta was USD 15,632. These include implementation and materials costs (one hour of salary for the ministry staff in charge of delivering the intervention per classroom and printing costs for the instructions for implementers, instructions for students, and posters for classrooms), which accounted for 72% of the total, and training costs (two hours of salary for the ministry staff who participated in the training session), which accounted for 28% of the total. Considering that it reached an estimated 5,535 students, it cost about USD 2.82 per student, and the marginal cost of adding a classroom of 25 students to the intervention was USD 135. Therefore, the intervention is inexpensive compared to other education interventions (see, for example, EEF, 2018), but its cost is higher than previously suggested.

3. Data

As Table 2 shows, I collected data on: (a) implementation fidelity (in 2017); (b) students' beliefs, effort, school climate, and plans after secondary school, from surveys in the national assessment (in 2017); (c) schools' resources and supports, from principal surveys in the national assessment (in 2016); (d) students' performance in school, from the census of schools (in 2016 and 2017); and (e) students' achievement from the national assessment (in 2016 and 2017).

3.1 Implementation fidelity

The MECyT of Salta provided me with the number of pictures submitted by each ATT at each school as proof for implementing the intervention and with the actual pictures.[14] I use these data to confirm that the intervention was implemented as intended in the vast majority of treatment schools and to estimate the effect of receiving the intervention. To my knowledge, this is the first study of a growth-mindset intervention at scale that can verify its implementation.[15]

3.2 Students' beliefs, effort, school climate, and plans for the future

The MECyT also provided me with the responses of all grade 12 students in Salta to a survey administered as part of the national assessment, roughly two months after the intervention. There are three aspects of this survey worth highlighting. First, it includes multiple questions on behaviors that ought to be affected by the intervention, allowing me to examine each step of its hypothesized causal chain, instead of relying on proxies.[16] Second, the survey was conducted independently from the intervention, which minimizes both the possibility of non-random attrition due to the intervention and of social desirability bias. Third, the survey is census-based, which means that it seeks to cover all grade 12 students.[17]

3.3 Schools' resources and supports

The MECyT also shared the responses of all secondary school principals in Salta to a survey administered at the same time as the national assessment.[18] I use responses to questions on school resources (which enquire about basic conditions, such as whether the school has electricity, and about educational resources, such as whether the school has a library) and school supports for low-performing students (e.g., whether the school develops a personalized plan for students who lag behind) from the 2016 survey to construct two indexes that I interact with the treatment indicator variable to explore heterogeneous effects by school characteristics.

3.4 Students' performance in school

The MECyT also granted me access to all data collected on internal efficiency (e.g., passing, repetition, and dropout rates) through the annual census of schools in Argentina. Importantly, these data are available for the year prior to the intervention, which I use to compare in- and out-of-sample schools and to check balance across experimental groups (sections 2.3 and 2.4), and for the year of the intervention, which I use to estimate impact. The data are reported for secondary schools and for grade 12 students, allowing me to test for impacts at both levels.

3.5 Student achievement

Finally, the MECyT provided the scores of all grade 12 students to the national assessment. This assessment evaluates what students know and can do based on the national curriculum. It is administered on an annual basis, but it covers different grades and subjects on each year. In the year prior to the intervention, the grade 12 test covered math, reading, and natural and social sciences, which I use to compare in- and out-of-sample schools and to check balance (sections 2.3 and 2.4). In the year of the intervention, it only focused on math and reading, which I use to estimate impact. The national ministry of education scaled all scores using a two-parameter logistic Item Response Theory (IRT) model (Yen & Fitzpatrick, 2006), which means that all effects in this paper are with respect to the overall national distribution. This feature sets this study apart from most prior evaluations of growth-mindset interventions, which use assessments designed by researchers and administered over a convenience sample.

4. Empirical Strategy

I estimate the effect of the offer of the intervention (i.e., the intent-to-treat or ITT effect) by fitting the following model:

$$Y_{is}^{t} = \alpha_{r(s)} + \gamma \bar{Y}_{s}^{t-1} + \beta T_{s} + \varepsilon_{is}^{t}$$

where Y_{is}^{t} is an outcome for student i in school s and year t, $r(s)$ is the randomization stratum of school s and $\alpha_{r(s)}$ is a stratum fixed effect, $\bar{Y}_{s}^{t-1}$ is the school-level average of the same outcome for year $t-1$, and T_s is an indicator variable for random assignment to treatment. (The census of schools and national assessment are repeated cross-sections of grade 12 students, so I do not observe each student's prior-year outcome). The parameter of interest is β, which captures the causal effect of the intervention. I use cluster-robust standard errors to account for within-school correlations across students in outcomes and include false discovery rate q-values to account for multiple hypothesis testing, using the Simes procedure in the qqvalue program in Stata (Newson, 2009). I also test the sensitivity of my estimates to the inclusion of $\bar{Y}_{s}^{t-1}$. I fit variations of this model that interact the treatment dummy with student characteristics (indicator variables for female students, students from low-income families, and students who had previously repeated a grade) and school characteristics (indexes of prior-year achievement, resources, and supports) to estimate the heterogeneous effects of the intervention on these sub-groups.[19]

5. Results

5.1 Implementation fidelity

The intervention was implemented as intended in the vast majority of treatment schools. In 85 of the 102 schools in this group (83%), the MECyT received pictures from the ATTs verifying that at least one grade 12 section had read the passage, wrote letters, and put them up next to the poster in their classroom. Further, the MECyT received more pictures from schools with more students: the median treatment school had one picture for every 26 students, which is close to the average class size for grade 12. In seven treatment schools, the ATTs did not send a picture, but a representative of the MECyT called the school and confirmed that the intervention was implemented with the principal. Therefore, the MECyT has verification that the intervention was implemented in 89 treatment schools (90.2% of schools in this group).

The intervention was not implemented in 10 of the 102 schools assigned to receive it (9.8%). In three cases, the principals refused to implement it; in four cases, the ATTs could not find a time that was convenient for them and for the school; and in three cases, the schools were located in areas that were difficult to access and the ATTs could not visit them in time.

ATTs did not track of the number of students who did not grant consent for the study, so I do not know the actual share of students in each classroom who participated in the intervention. However, I estimate this share using two different strategies to offer a range of plausible values for students' participation rate in the intervention.

First, I estimate this share by: (a) identifying the maximum number of eligible students at each school from the enrollment figures for grade 12, the target grade for the intervention (using the school performance data described in section 3.4); (b) adjusting this number based on the average number of absences self-reported by grade 12 students (using the student achievement data described in section 3.5);[20] and (c) dividing the result by the number of student letters from the implementation pictures in each school (using the implementation fidelity data described in section 3.1).[21] This approach indicates that 58% of students completed the activity. This estimate, however, is extremely conservative because it does not consider that students tend to under-report absences, that some students in the enrollment registers may have already dropped out when the intervention was implemented (two months before the end of the school year), and that school principals in Argentina face incentives to over-report student enrollment to keep the number of sections (and thus, the number of teachers they are allowed to hire) constant.

Then, I estimate this share by: (a) identifying the likely number of eligible students based on the actual students who took the national assessment in grade 12 (using the student achievement data from section 3.5), which was administered two months after the intervention and thus offers a more realistic proxy for the actual number of students at the time of the study;[22] and (b) dividing the likely number of eligible students by the number of student letters, as above. This approach indicates that 65% of students completed the activity. This estimate, however, is probably still conservative given that ATTs were not instructed to include all letters from students in their implementation-verification pictures, and accordingly, many of these pictures display the edges of other letters, indicating that some letters were out of the picture frame.

Importantly, both of my estimates of student participation rates are above the 56% response rate in the largest evaluation of a growth-mindset intervention in the United States (see Gopalan & Tipton, 2018). Scaling up my intent-to-treat results by my estimates of the student participation rates would make it harder for me to rule out policy relevant positive effect sizes. Yet, given that only 22 of the 102 treatment schools had pictures that were of high enough resolution to allow me to count the number of student letters, and that even among those schools, pictures did not include all the letters completed in a classroom, it is not possible to know whether such an adjustment would be preferred or even warranted.

5.2 Students' beliefs

In spite of having been implemented with fidelity, the intervention had no effect on students' propensity to find challenging tasks less intimidating. I address this question in three ways, based on my theory of change of the intervention (see discussion in section 2.4).

First, I explore whether treatment students perceived school-related tasks as less challenging. I identified several questions in the survey administered as part of the national assessment that asked students about the extent to which they found a set of school-related tasks difficult (e.g., paying attention in class) using a scale that ranged from 1 ("very simple") to 4 ("very difficult"). I coded responses dichotomously, using a 1 for "very difficult" or "difficult" and 0 otherwise and analyzed whether treatment students were less likely to find these tasks challenging. The intervention had a precisely estimated zero effect on all outcomes, ruling out even small effects of 4 percentage points (pp.) or more (Table 3).[23]

Then, I examine whether the intervention improved students' beliefs about their self-efficacy. I identified questions in the survey that asked students to indicate whether they understand and do well in math and language using a scale that ranged from 1 ("always") to 4 ("never"). I coded responses dichotomously, using a 1 for "always" or "most of the time" and 0 otherwise. The intervention had a null effect on all outcomes, ruling out effects larger than 5 pp. (Table 4).

Finally, I consider whether treatment students were more likely to see tests as formative. I used questions that explicitly asked students about the extent to which assessments served formative purposes, which employed the same scale as above. Once again, the intervention had a precisely estimated null effect on all outcomes, ruling out effects larger than 3 pp. (Table 5).

These null effects, however, mask heterogeneous effects for groups of disadvantaged students. I find some evidence that the intervention may have negatively impacted the beliefs of female students, students from low-income families, and those who had previously repeated a grade. I created indexes of students' perceptions of the difficulty of schoolwork, self-efficacy, and perceptions of the usefulness of classroom assessments by conducting principal component analyses of variables in Tables 3-5 and taking the first principal component of each analysis. Then, I estimated the effect of the intervention on the indexes (not on the individual variables) for each group to reduce the probability of false positives due to multiple hypothesis testing. The intervention seems to have *increased* the perceived difficulty of schoolwork among girls, *decreased* the self-efficacy of girls and students from low-income families, and *decreased* the perceived usefulness of assessments among students who had repeated a grade (Table A.5). Surprisingly, even when the coefficient on the interaction term is not statistically significant, its sign is typically the opposite of what I predicted in the theory of change of the intervention.

I do not find any evidence of heterogeneous effects on students' beliefs by school characteristics. I interact the treatment dummy with each school's prior-year average score on the national assessment (across all subjects), an index of school resources, and an index of school supports (see section 4) and find no statistically significant interaction effects along these dimensions. Yet, most interactions are imprecisely estimated, so it is possible that they exist but I lack sufficient statistical power to detect them (Table A.6).

5.3 Student effort

I also estimate the impact of the intervention on three sets of indicators of student effort. I begin by focusing on school-related tasks, the domain in which I most expected to see changes. I examine whether treatment students were more likely to attend school or private tuition.[24] The survey in the national assessment asks how often students missed school during the year using a scale from 1 ("never") to 4 ("more than 24 times"). I coded responses that constituted "chronic absenteeism" (15 absences or more, see Gottfried, 2014) as 1 and 0 otherwise. The question on tuition was a yes/no question, so I coded answers dichotomously. Surprisingly, treatment students were 3.7 pp. *more* likely than their control peers to miss school (Table 6). However, this difference is only marginally statistically significant, and as the q-value indicates, likely to have emerged due to multiple hypothesis testing. I find no effect on the intervention on students' propensity to attend tuition.

I also examine whether treatment students worked harder on their personal development (e.g., read books outside of school, take art lessons, learn a foreign language, or play sports). All of these were yes/no questions, so I coded them dichotomously. Again, I find a precisely estimated zero effect on all outcomes, allowing me to rule out effects larger than 4 pp.

Lastly, I consider whether the intervention increased student effort on existing obligations (e.g., work at or outside of home). Both were yes/no questions and were coded dichotomously. Once again, I find precisely estimated null effects on all outcomes.

I find no evidence of heterogeneous effects on any group of variables measuring student effort. I created indexes of student effort on school-related tasks, personal development, and existing obligations using the first principal component from separate principal component analyses. Then, I estimated the effect of the intervention on these indexes for the same groups as above. The coefficients on the interactions are around zero and statistically insignificant (Table A.7). I find no evidence of heterogeneity by school characteristics; in fact, most interaction effects are estimated around zero, allowing me to rule out small-to-moderate effects (Table A.8).

5.4 School climate

Next, I estimate the effect of the intervention on three measures of school climate. I first focus on a question that asked students whether they get along with their peers using a scale from 1 ("no, I do not get along with anyone") to 5 ("yes, I get along with everyone"). I coded responses as 1 if students indicated that they got along with some, most, or all their peers and 0 if they reported that they did not get along with anyone or with only a few peers. The intervention reduced students' propensity to get along with peers by 1.5 pp., but as the q-value indicates, this effect is likely to have emerged due to multiple hypothesis testing (Table 7).

I also estimate the effect of the intervention on the student-reported prevalence of bullying. The survey in the national assessment asks students how often peers at their school engage in bullying on a number of groups using a scale that ranges from 1 ("always") to 4 ("never"). I coded responses as 1 if students indicated bullying occurred "often" or "always" and 0 otherwise. Consistent with the results above, I find that the intervention increased the prevalence of bullying against female students by 1.5 pp. and had no effects on other types of bullying. Again, however, this effect seems to have emerged due to multiple hypothesis testing.

Finally, I consider whether the intervention had any effect on student-reported vandalism, which was measured using the same scale and which I coded in the same manner as above. I did not find that the intervention affected the incidence of theft or damages to school property.

I find little evidence of heterogeneous effects on any of the variables measuring school climate. I used the first indicator variable in Table 7 by itself and created indexes of bullying and vandalism using the first principal component from separate principal component analyses. I estimated the effect of the intervention on these variables for the same groups as above. The intervention had a *negative*, but marginally statistically significant, effect on the propensity of female students to get along with peers, but all other interaction terms were consistently estimated around zero and statistically insignificant (Table A.9). I do not find any evidence of heterogeneous effects on school climate by school characteristics (Table A.10).

5.5 Students' performance in school

I also estimate the effect of the intervention on students' performance in school, as measured by the number of enrolled students, and the percentage of students who passed, failed, or repeated the grade, or who dropped out of school. I do not find evidence that the intervention had a positive effect on these outcomes, but my estimates are more imprecise than those for other outcomes because these data are collected at the school level. (This is also why I cannot estimate heterogeneous effects on these outcomes by students' characteristics). The results are similar when I account for schools' performance in the year before the intervention (Table 8).

5.6 Student achievement

Then, I estimate the effect of the intervention on student achievement, as measured by the results of the national assessment of math and reading in grade 12. I find no evidence that the intervention improved test scores in either subject, before or after accounting for the schools' performance in the year prior to the intervention (Table 9). I can rule out effects larger than .07 standard deviations in both subjects. In fact, the distribution of student achievement looks nearly identical across the control and treatment groups, two months after the intervention (Figure A.1). Further, I find no evidence of heterogeneous effects by students' sex, socio-economic status, or prior repetition (Table A.11).

Interestingly, all interactions between the treatment and school-level characteristics are negative, suggesting that schools with higher levels of achievement, resources, and supports benefit less from the intervention. The only statistically significant interaction, however, is the one between the treatment and school resources for math (Table A.12).

5.7 Students' plans after secondary education

Finally, I estimate the effect of the intervention on students' post-secondary education plans. Students were asked whether they planned to study, work, or do both, so I coded each option dichotomously. I do not find any indication that the intervention affected the plans of the average student (Table 10) or of the sub-groups of students mentioned above (Table A.13).

I find some evidence of heterogeneity by school characteristics. First, in schools with higher levels of achievement, the intervention increased the share of students who plan to work and study after secondary school. Second, in schools with more resources, the intervention increased the share of students who plan to work. Third, in schools with more supports, the intervention reduced the share of students who plan to study and increased the share of students who plan to work and study by a similar magnitude (Table A.14).

6. Discussion

6.1 Implications for research

The present study highlights the importance of evaluating promising educational interventions at scale to understand their effectiveness when they are implemented within a school system. The null effects that I found differ considerably from the encouraging results of efficacy trials and they are more consistent with the results from two recent large-scale impact evaluations. Outes et al. (2020) evaluated the intervention in 800 secondary schools in Peru. They found that it raised achievement in math (by .05 standard deviations), but not in reading comprehension. Effects were driven by one region; results for the other two were not statistically significant. Yeager et al. (2019) evaluated the intervention in 65 secondary schools in the United States. They found that it had no effects on the grades of or courses taken by the average student, but low-performers improved their grades and high-performers took more challenging classes.

This study, when read alongside the two other effectiveness trials, also suggests that the intervention only improves achievement when it changes students' beliefs about intelligence. In Salta, I found that it had no effect on beliefs (see section 5.2), so it is perhaps not surprising that it had no impact on effort, climate, performance, or achievement (see sections 5.3 to 5.7). In

Peru, Outes et al. (2020) found that the intervention only had a positive impact on math achievement in Ancash, where it also improved students' self-beliefs in math. They found no such effects on beliefs or achievement in Junín or Lima, the two other regions. In the U.S., Yeager et al. (2019) found that the intervention only improved grades among low-performers, who not only changed their mindsets but also had margin for improvement. These studies draw attention to the importance of piloting the intervention to ensure that it changes students' mindsets before evaluating its impacts on school performance or achievement.

The studies in Salta, Peru, and the U.S. also raise important questions about how context may moderate the effects of the intervention. Context may matter for at least four reasons. First, systems, schools, and classrooms may differ in their capacity to implement the intervention. Outes et al. (2020) found that Ancash, the region of Peru that most benefited from the intervention, had implemented it with greater fidelity than the other two regions. Yet, the Salta study shows that the intervention can fail even when it is implemented correctly. Second, systems, schools, and students may differ in their margin for potential improvements.[25] As the authors of the Peru study note, Ancash is also the most rural of the three regions, so it is also possible that it benefited the most because it started from a lower level of performance. This interpretation is consistent with the results of the U.S. study for low-performing students. Third, students may also differ in their baseline beliefs about the malleability of intelligence. It is possible that students in Salta, Junín, and Lima did not change their beliefs or raise their effort because they did not hold a fixed mindset before they participated in the intervention. This would be consistent with the results of some previous studies in the U.S., where the intervention has been impactful among students with fixed mindsets (e.g., Yeager et al., 2014), but Yeager et al. (2019) do not find treatment heterogeneity by students' baseline mindsets. Finally, schools and teachers may differ in their capacity to help students increase their effort. Yeager et al. (2019) see this as the reason why the intervention has larger effects in schools whose students exhibit challenge-seeking behaviors. Yet, the Salta study finds no heterogeneity across schools with different levels of resources or supports for low-performing students.

The differences in the results of these studies also raise questions about how the intervention may change mindsets. The focus has been on the reading that students are asked to complete. Yet, there are at least two important differences in how the intervention was delivered across Salta and Peru and the U.S. that may play a more important role than previously anticipated. One difference is whether students are required to check their understanding of the reading (before they are asked to write a letter to a classmate on the main lessons from the passage). In Peru, students were asked to answer review questions and discuss the reading in groups. In the U.S., students were asked to summarize the findings of the reading in their own words. This step may be especially important in developing countries, where reading skills are low, and it may partly explain why the intervention had no effects in Salta, where it was omitted.[26] Another difference is whether the activity is led by the students' teachers, as it was in Peru. This could potentially both educate teachers and influence their interactions with students (for a broader discussion of this possibility, see Raudenbush, 1984; Yeager & Walton, 2011). The relative importance of this aspect, however, is unclear, as the intervention in the U.S. was effective even if it was delivered online and teachers did not know which students received it.

These studies also offer several lessons for the design of future evaluations of this intervention. First, they highlight the importance of not only evaluating the intervention at scale, but also of having sufficient statistical power to detect heterogeneous treatment effects across sites. Second, they illustrate the usefulness of measuring students' pre-intervention mindsets and their post-intervention understanding of the reading to make sense of potential null results. This may be achieved either by combining lab and field experiments or by embedding the former in the latter to keep data collection costs manageable. Third, they make clear how essential it is to collect information on students' backgrounds and schools' resources and practices to examine heterogeneous effects along these dimensions.

6.2 Implications for policy

The present study draws attention to the importance of context, intervention design, and implementation in taking education initiatives to scale in developing countries. The case of Salta suggests that delivering the growth-mindset intervention using materials and following processes that have yielded positive effects in other settings will not necessarily lead to similar results (see Yeager & Walton, 2011). Further, the costs of implementing it are not trivial and should be compared against those of initiatives with evidence of effectiveness in these settings (see Ganimian & Murnane, 2016).

This study also offers governments interested in implementing the intervention guidance on some of the aspects that they should consider when deciding whether and how to do so. First, they should try to understand whether potential beneficiaries hold a fixed mindset and whether the extent to which they hold such beliefs is related to their academic performance. They should also consider whether schools will seek to implement the intervention with little training or support (as in Peru) or with both (as in Salta).[27] This decision will play an important role in determining the costs of implementation. Finally, if possible, governments should consider using data already collected by their school system to evaluate the impact of the first iteration of the intervention through a randomized rollout. This will reduce costs, avoid bias in responses, and minimize differential participation, and allow the government to understand whether the intervention works for their school system.

7. Conclusion

I present experimental evidence on a growth-mindset intervention implemented at scale in public secondary schools in Salta, Argentina and find it had no effects, either on intermediate outcomes (e.g., students' beliefs, effort, or school climate) or the ultimate outcomes of interest (e.g., students' performance in school, achievement, and post-graduation plans). Nearly all results are precisely estimated and allow me to rule out even small effects. I find little evidence of heterogeneous effects by students' sex, socio-economic status, and prior grade repetition, or by schools' educational resources and support for low-performing students.

This study and my review of the literature seek to raise important questions about the effectiveness of growth-mindset interventions when implemented at scale in developing countries. It does not seek to call into question the efficacy of variations of this intervention when implemented by its developers among small convenience samples of schools and students, let alone the decades of work that developed the theory on which these interventions are based. It simply proposes a way forward for identifying the conditions that would maximize impact.

8. Endnotes

[1] The reason why Salta is one of the lower-performing school systems in the country but has similar percentages of students in the lowest two levels as the national average is that two-thirds of students in Argentina go to school in the Province of Buenos Aires (a single school system), which generally drives the national averages in the national assessment (SEE-MEDN, 2018b).

[2] It should be noted, however, that several studies have found positive effects of similar interventions after two sessions (Good et al., 2003; Yeager et al., 2014) or one session (Mendoza-Denton, Kahn, & Chan, 2008; Paunesku et al., 2015; Yeager et al., 2013, Study 3; Yeager et al., 2011) including, a 15-minute session (Yeager et al., 2013, Study 2).

[3] ATTs have teaching degrees and either serve or have served in the past as teachers. The MECyT trained all ATTs on how to deliver the intervention in August of 2017, using materials I had prepared.

[4] The original English version of the reading can be accessed at: https://bit.ly/2IRAJI5. The Spanish translation used in Salta can be found at: https://bit.ly/2YfL1VS.

[5] This component of the intervention was first used by Outes et al. (2020) in Peru. The original English version of the poster can be accessed at: https://bit.ly/2HWQfQJ. The Spanish translation used in Salta can be found at: https://bit.ly/2Tl0HU9. The poster was translated by Mindset Works, the organization that had developed the original version.

[6] Several studies have examined whether the growth-mindset intervention impacts students' perceived difficulty of school-related tasks (Burnette et al., 2018; Mendoza-Denton et al., 2008).

[7] Prior studies have documented the effect of mindset interventions on motivation (Blackwell et al., 2007; Eccles, Wigfield, & Schiefele, 1998), but few have included actual measures of effort.

[8] This expectation was informed by the evidence on the effect of mindset interventions on stereotype threat (Aronson et al., 2002; Good et al., 2003) and hostile intent, aggression, and desire to seek revenge (Yeager et al., 2013; Yeager et al., 2011).

[9] Multiple evaluations of mindset interventions have found effects on school performance (Blackwell et al., 2007; Paunesku et al., 2015), but only a few have evaluated its effect on achievement on standardized tests (Good et al., 2003).

[10] Several studies have found that mindset interventions can affect students' post-secondary education plans (Outes et al., 2020; Yeager et al., 2019).

[11] Multiple studies have found that the intervention only works or works best for these sub-groups of students (Aronson et al., 2002; Broda et al., 2018; Good et al., 2003; Paunesku et al., 2015; Yeager et al., 2019; Yeager et al., 2016).

[12] The two largest field experiments in this literature document considerable treatment heterogeneity across schools (Outes et al., 2020; Yeager et al., 2019).

[13] To my knowledge, only one study has examined treatment heterogeneity by school characteristics (Yeager et al., 2019).

[14] Unfortunately, the photos are not of high enough quality to allow me to analyze the content of the letters (e.g., to gauge whether students understood or were persuaded by the reading).

[15] In Peru, Outes et al. (2020) also asked schools that were randomly assigned to implement a similar intervention to submit pictures, but they received such pictures for less than half of treatment schools.

[16] The original survey in Spanish can be accessed at: https://bit.ly/2I0C39h.

[17] Salta has traditionally had high participation rates in the national assessment. In 2016, 92% of all public secondary schools and 83% of all students in these schools participated in the

assessment (SEE-MEDN, 2016). In 2017, 97% of public schools and 79% of students at this level participated (SEE-MEDN, 2018c).

[18] The original survey in Spanish can be accessed at: https://bit.ly/2WhogPp.

[19] The index of prior-year achievement is the school-level average score in the 2016 national assessment, which covered math, reading, and natural and social sciences (see section 3.5). The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively.

[20] I imputed the mean absence rate in the treatment group for four schools without absence data.

[21] I imputed the mean number of letters for schools without clear pictures, under the assumption that the resolution of the pictures of student letters (which is largely determined by the quality of the camera of each implementer's smart phone) is orthogonal to actual implementation fidelity.

[22] These assessments are not attached to any stakes and the National Education Law of 2006 expressly prohibits the dissemination of achievement data at the school, teacher, or student level (see Ganimian, 2015), so schools face no incentives to discourage lower-achieving students from taking the exam.

[23] Throughout the manuscript, when I state that I can rule out effects of a given magnitude, I am referring to the upper bound of the 95% confidence interval (see, e.g., Hoxby, 2000). For example, the upper bound of the first estimate in Table 3 is 2.6 pp., so effects above this magnitude are unlikely. When I make this claim and multiple related hypotheses are being tested, I use the largest upper bound that I observe in a family. For example, in Table 3, I state that I can rule out effects larger than 4 pp. because that is the largest upper bound I observe across all outcomes in that table.

[24] In Argentina, the word "tuition" (*apoyo escolar*) refers not only to fee-charging private providers, but also to programs offered by the government and non-profits for free. Therefore, cost is not as much of a barrier as the word may suggest from its use in other developing countries.

[25] A variation of this argument is that grade 12 students, who are about to graduate from secondary school, may have fewer reasons to change their beliefs and mindsets than grade 9 students, who are transitioning into what is known as middle school in the U.S. and as lower secondary school in other countries.

[26] I explored whether the effect of the intervention in Salta varied either by students' self-assessment of their capacity to understand texts or by their schools' prior-year reading levels, but did not find any evidence of heterogeneous effects on mechanisms or outcomes. These results are available upon request.

[27] As mentioned in section 2.5, very few developing countries have the requisite technological infrastructure to deliver this intervention online throughout the school system.

Table 1: Theory of change of the growth-mindset intervention in Salta

(1) **Need**	(2) **Inputs/Activities**	(3) **Outputs**	(4) **Outcomes**	(5) **Impact**
• **Students believe intelligence is fixed**, which leads them to want to look smart and thus to avoid challenges, give up in the face of obstacles, see effort as pointless, ignore useful negative feedback, and feel threatened by the success of others	• **Ministry representatives receive training** on existing evidence on the growth-mindset intervention and on how to deliver it • **Ministry representatives visit schools** and secure permission from the principal to implement the intervention • **Ministry representatives deliver the intervention in grade 12 classrooms** after securing consent from students	• **Students read a passage** on the malleability of intelligence • **Students write a letter to a classmate** on the three main lessons from the reading • **Students post their letters on the classroom**, next to a poster reminding them of the key messages of the intervention for the rest of the school year	• **Students are less intimidated by challenging tasks** because they see them as less difficult, they feel more capable of doing hard work, or they see them as part of a learning process • **Students exert more effort** in school-related tasks, tasks related to personal development, or existing obligations • **Students get along better with peers**, either by improving their existing relationships or at least engaging in fewer acts of hostility of vandalism towards each other	• **Students' performance in school improves** as a result from a shift in their approach to challenges, greater effort, or better relationships with peers • **Student learning increases** as a result from doing better in school • **Students plan to pursue post-secondary education** as a result from their improvement in school performance and learning
Assumptions:	• Availability of non-academic period at the school (i.e., *tutorías*) • No opposition from principals, teachers, or students	• Students can read and comprehend the text • Students can write a letter • Students do not mind posting their letters on the classroom	• Teachers do not foster a fixed mindset • Outside factors do not hamper student effort • Other students do not encourage bullying, violence, or vandalism	• School work is attainable for students • Student assessments measure what students learn at school • Students do not need to work to support their families

Source: Author's elaboration.

Table 2: Timeline of the study

(1)	(2)	(3)	(4)
		School participation rates	
Month	Event	Control schools	Treatment schools
Panel A. 2016			
February	School year starts		
November	National assessment of grade 12 students (tests of math, reading, natural and social sciences and principal survey)	96%	93%
December	School year ends		
Panel B. 2017			
February	School year starts		
April	MECyT shares data from national census of schools (2016 school year)	100%	100%
August	MECyT holds training for ATTs	-	100%
September	ATTs deliver the intervention	-	100%
November	National assessment of grade 12 students (tests of math and reading and student survey)	99%	95%
December	School year ends		
Panel C. 2018			
February	School year starts		
April	MECyT shares data from national census of schools (2017 school year)	100%	100%
December	School year ends		

Notes: (1) The table shows the timeline for the interventions and rounds of data collection for the study, including the month in which each event occurred (column 1), a brief description of the event (column 2), and the percentage of schools that participated in each event by experimental group (columns 3-4). (2) MECyT refers to the Ministry of Education, Science, and Technology of Salta. ATTs refers to the *Asistentes Técnicos Territoriales* (ATTs), the MECyT staffers who delivered the intervention.

Table 3: ITT effect on students' perceptions of difficulty of schoolwork (2017)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Understanding texts	Writing texts	Speaking in public	Learning new concepts	Paying attention in class	Working in groups	Participating in class	Solving problems
Treatment	0.0049	-0.0056	0.0117	0.0003	0.0089	0.0089	0.0027	-0.0057
	(0.0112)	(0.0122)	(0.0153)	(0.0127)	(0.0110)	(0.0102)	(0.0136)	(0.0141)
Observations	9372	9372	9372	9372	9372	9372	9372	9372
R^2	0.002	0.001	0.001	0.002	0.005	0.001	0.006	0.003
Control mean	0.200	0.254	0.398	0.307	0.196	0.193	0.309	0.438
FDR q-value	0.916	0.916	0.916	0.984	0.916	0.916	0.962	0.916

Notes: This table shows the intent-to-treat (ITT) effect of the intervention on students' perceived difficulty of tasks related to schoolwork. (2) Students were asked to indicate how difficult they found the activities listed above using a scale ranging from 1 ("very simple") to 4 ("very difficult"). The dependent variables in this table are dummies that equal one for students who indicated the task was difficult or very difficult and zero otherwise. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: ITT effect on students' self-efficacy, by subject (2017)

	Math		Language	
	(1)	(2)	(3)	(4)
	I understand it quickly	I do well in it	I understand it quickly	I do well in it
Treatment	0.00004	0.01094	-0.01531	-0.00638
	(0.01630)	(0.01581)	(0.01928)	(0.01869)
Observations	9372	9372	9372	9372
R^2	0.001	0.002	0.003	0.004
Control mean	0.348	0.424	0.563	0.581
FDR q-value	0.998	0.978	0.978	0.978

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on students' perceived performance on math and language. (2) Students were asked to indicate how often they agreed with the statements listed above using a scale ranging from 1 ("always") to 4 ("never"). The dependent variables in this table are dummies that equal one for students who indicated they agreed always or most of the times and zero otherwise. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: ITT effect on students' perceptions of classroom assessments (2017)

	(1) Tests help me improve	(2) Tests help me identify errors	(3) Tests check if I understood what I was taught
Treatment	-0.0041	-0.0011	0.0029
	(0.0107)	(0.0079)	(0.0130)
Observations	9112	9372	9372
R^2	0.008	0.000	0.001
Control mean	0.875	0.150	0.507
FDR q-value	0.894	0.894	0.894

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on students' perceived usefulness of classroom assessments. (2) Students were asked to indicate how often they agreed with the statements listed above using a scale ranging from 1 ("always") to 4 ("never"). The dependent variables in this table are dummies that equal one for students who indicated they agreed always or most of the times and zero otherwise. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: ITT effect on student effort (2017)

	School-related tasks		Personal development				Existing obligations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Chronically absent to school	Attends private tuition	Reads books outside of school	Takes artistic lessons	Learns a foreign language	Plays sports	Works at home	Works outside of home
Treatment	0.0374*	-0.0070	0.0065	-0.0046	0.0090	0.0084	-0.0080	0.0084
	(0.0206)	(0.0229)	(0.0137)	(0.0110)	(0.0091)	(0.0136)	(0.0172)	(0.0149)
Observations	9372	9002	8236	8243	8156	8558	8969	8946
R^2	0.017	0.013	0.005	0.003	0.003	0.009	0.023	0.002
Control mean	0.288	0.233	0.343	0.166	0.083	0.617	0.459	0.264
FDR q-value	0.572	0.761	0.761	0.761	0.761	0.761	0.761	0.761

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on student effort. (2) Students were asked to indicate how many schooldays they had missed during the year using a scale ranging from 1 ("never") to 4 ("more than 24 days"). The dependent variable on absenteeism is a dummy that equal one for students who reported to have missed 15 or more days and zero otherwise. All questions on personal development and existing obligations were yes/no questions and were coded dichotomously. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: ITT effect on school climate (2017)

		Bullying				Vandalism	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Gets along with peers	Bullying on students with good grades	Bullying on students who repeat grades	Bullying on students because of their personal or family characteristics	Bullying on female students	Stealing	Damaging school property
Treatment	-0.0149**	0.0074	-0.0016	0.0126	0.0153**	0.0164	0.0106
	(0.0072)	(0.0133)	(0.0119)	(0.0146)	(0.0074)	(0.0167)	(0.0182)
Observations	9150	9372	9372	9372	9372	9372	9372
R^2	0.002	0.001	0.002	0.004	0.002	0.006	0.011
Control mean	0.904	0.192	0.197	0.252	0.079	0.160	0.314
FDR q-value	0.141	0.672	0.894	0.672	0.141	0.672	0.672

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on students' perceived school climate. (2) Students were asked to indicate how frequently other students at their school engaged in the activities listed above using a scale ranging from 1 ("always") to 4 ("never"). The dependent variables in this table are dummies that equal one for students who indicated that the activities occurred always or many times and zero otherwise. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8: ITT effect on students' performance in school (2017)

	Number of students enrolled		Percentage of students who passed the grade		Percentage of students who failed the grade		Percentage of students who dropped out of school		Percentage of students who repeated the grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment	-10.35	-10.46	-2.28	-2.21	1.93	1.86	-0.29	-0.27	0.01	0.01
	(6.58)	(6.59)	(2.64)	(2.63)	(2.48)	(2.47)	(0.95)	(0.95)	(0.51)	(0.51)
Prior-year school index		1.51		-1.43*		1.27		-0.32		0.01
		(2.08)		(0.83)		(0.78)		(0.30)		(0.16)
Observations	189	189	195	195	195	195	199	199	195	195
R^2	0.331	0.332	0.255	0.267	0.292	0.302	0.033	0.039	0.007	0.007
Control mean	68.88		72.15		25.18		3.39		2.46	
FDR q-value	0.588	0.588	0.753	0.753	0.753	0.753	0.970	0.970	0.985	0.985

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on students' performance in school. (2) This information is collected at the school level through the national census of schools. (3) The prior-year school index is the first principal component from a principal component analysis that included the enrollment, passing, failure, repetition, and dropout rates for grades 8 to 12 in all schools in the sample for the 2016 school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 9: ITT effect on student achievement (2017)

	Math (IRT-scaled score)		Reading (IRT-scaled score)	
	(1)	(2)	(3)	(4)
Treatment	-0.020	0.015	-0.051	-0.008
	(0.049)	(0.035)	(0.063)	(0.043)
Prior-year school index		0.641***		0.801***
		(0.107)		(0.117)
Observations	8814	8814	8865	8865
R^2	0.025	0.076	0.018	0.067
Control mean	-0.259		-0.055	
FDR q-value	0.853	0.853	0.853	0.853

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on student achievement. (2) All scores have been scaled using a two-parameter logistic Item Response Theory (IRT) model with respect to the national distribution to have a mean of zero and a standard deviation of one. (3) Prior-year school achievement refers to the first principal component from a principal component analysis that included school-level average scores in assessments of math, reading, natural, and social sciences in grade 12 during the 2016 school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 10: ITT effect on plans after secondary education (2017)

	(1) Plans to work	(2) Plans to study	(3) Plans to do both
Treatment	-0.001	-0.004	0.013
	(0.006)	(0.020)	(0.016)
Observations	9377	9377	9377
R^2	0.004	0.006	0.001
Control mean	0.043	0.453	0.360
FDR q-value	0.903	0.903	0.903

Notes: (1) This table shows the intent-to-treat (ITT) effect of the intervention on students' post-secondary plans. (2) The dependent variables in the regressions are indicator variables for students who indicated that they plan to work, study, or do both after secondary education. (3) All estimations include randomization strata fixed effects. (4) * significant at 10%; ** significant at 5%; *** significant at 1%.

9. References

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review, 107*(6), 1535-1563. doi:10.2139/ssrn.2538624

Aronson, J., Fried, C. B., & 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*, 113-125.

Avitabile, C., & de Hoyos, R. (2014). *The heterogeneous effects of information about the returns to schooling on student learning: Evidence from a randomized control trial in Mexico*. Retrieved from Washington, DC:

Banerjee, A. V., Glewwe, P., Powers, S., & Wasserman, M. (2013). *Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence*. Unpublished manuscript. Abdul Latif Jameel Poverty Action Lab (J-PAL). Cambridge, MA.

Barrera-Osorio, F., Deming, D., González, K., & Lagos, F. (2020). *Providing performance information in education: An experimental evaluation in Colombia*. Unpublished manuscript. Harvard Graduate School of Education (HGSE). Cambridge, MA.

Bassi, M., Busso, M., & Munoz, J. S. (2013). *Is the glass half empty or half full? School enrollment, graduation, and dropout rates in Latin America*. (IDB Working Paper Series No. IDB-WP-462). Inter-American Development Bank. Washington, DC.

Berniell, L. (2014). *Condicionantes de la entrada de los jóvenes al mundo laboral: Educación, información y oportunidades laborales de calidad*. AAEP 2014. Banco de Desarrollo de América Latina (CAF). Posadas, Argentina.

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. S. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization, 146*, 1-15.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development, 78*(1), 246-263.

Bobba, M., & Frisancho, V. (2016). *Learning about oneself: The effects of signaling academic ability on school choice*. Unpublished manuscripts. Inter-American Development Bank. Washington, DC.

Bonilla, L., Bottan, N. L., & Ham, A. (2016). *Information policies and higher education choices: Experimental evidence from Colombia*. Unpublished manuscript. University of Illinois. Urbana, IL.

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., & Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness, 11*(3), 317-338.

Burnette, J. L., Russell, M. V., Hoyt, C. L., Orvidas, K., & Widman, L. (2018). An online growth mindset intervention in a sample of rural adolescent girls. *British Journal of Educational Psychology, 88*(3), 428–445.

Camargo, B., Camelo, R., Firpo, S., & Ponczek, V. (2018). Information, market incentives, and student performance: Evidence from a regression discontinuity design in Brazil. *The Journal of Human Resources, 53*(2), 414-444.

Chaia, A., Child, F., Dorn, E., Frank, M., Krawitz, M., & Mourshed, M. (2017). *Drivers of student performance: Latin America insights*. McKinsey & Company. London, UK.

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences, 113*(31), 8664–8668.

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2012). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. In P. Glewwe (Ed.), *Education policy in developing countries*. Chicago, IL and London, UK: The University of Chicago Press: Abdul Latif Jameel Poverty Action Lab (J-PAL).

DiEE. (2016). *Anuario estadístico 2016*.

DiNIECE. (2013). *Redefiniciones normativas y desafíos de la educación secundaria en Argentina. Acuerdos federales en un sistema descentralizado*. La educación en debate. Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE). Buenos Aires, Argentina.

Dizon-Ross, R. (2019). Parents' beliefs about their children’s academic ability: Implications for educational investments. *American Economic Review*.

Dommett, E. J., Devonshire, I. M., Sewter, E., & Greenfield, S. A. (2013). The impact of participation in a neuroscience course on motivational measures and academic performance. *Trends in Neuroscience and Education*, 122–138.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological review, 95*(2), 256.

Dweck, C. S., Walton, G. M., & Cohen, G. L. (2014). *Academic tenacity: Mindsets and skills that promote long-term learning*. Bill & Melinda Gates Foundation. Seattle, WA.

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481-496.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In N. Eisenberg (Ed.), *Handbook of child psychology, Vol. 3: Social, emotional, and personality development (5th ed.)*. New York, NY: Wiley.

EEF. (2018). *Teaching and learning toolkit*. Education Endowment Foundation. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Toolkit/complete/EEF-Teaching-Learning-Toolkit-October-2018.pdf

Gandhi, J., Watts, T. W., Masucci, M. D., & Raver, C. C. (2019). *The effects of two mindset interventions on low-income students' academic and psychological outcomes*. Unpublished manuscript. New York University. New York, NY.

Ganimian, A. J. (2015). *El termómetro educativo: Informe sobre el desempeño de Argentina en los Operativos Nacionales de Evaluación (ONE) 2005-2013*. Proyecto Educar 2050. Buenos Aires, Argentina.

Ganimian, A. J., & Murnane, R. J. (2016). Improving educational outcomes in developing countries: Lessons from rigorous evaluations. *Review of Educational Research, 86*(3), 719-755.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*(645-662).

Gopalan, M., & Tipton, E. (2018). *Is the National Study of Learning Mindsets nationally representative?* Unpublished manuscript. Pennsylvania State University. State College, PA.

Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk (JESPAR), 19*(2), 53-75.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics, 115*(4), 1239-1285.

Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., . . . Rozelle, S. (2013). Can information and counseling help students from poor rural areas go to high school? Evidence from China. *Journal of Comparative Economics, 41*, 1012–1025. doi:10.1016/j.jce.2013.06.004

MECyT. (2012). *Diseño curricular para educación secundaria*. Ministerio de Educación, Ciencia y Tecnología, Gobierno de la Provincia de Salta. Salta, Argentina.

Mendoza-Denton, R., Kahn, K., & Chan, W. (2008). Can fixed views of ability boost performance in the context of favorable stereotypes? *Journal of Experimental Social Psychology, 44*(4), 1187–1193.

Newson, R. (2009). *QQVALUE: Stata module to generate quasi-q-values by inverting multiple-test procedures*.

Nguyen, T. (2009). *Information, role models and perceived returns to education: Experimental evidence from Madagascar*. Unpublished manuscript. Massachusetts Institute of Technology. Cambridge, MA.

OECD. (2013). *PISA 2012 results: What students know and can do. Student performance in mathematics, reading and science (Vol. I)*. Retrieved from Paris, France:

OECD. (2018). *Education at a glance 2018: OECD indicators*. Organisation for Economic Co-operation and Development (OECD). Paris, France.

Oreopoulos, P., Brown, R. S., & Lavecchia, A. M. (2017). Pathways to education: An integrated approach to helping at-risk high school students. *Journal of Political Economy, 125*(4), 947-984.

Oreopoulos, P., Patterson, R. W., Petronijevic, U., & Pope, N. G. (2018). *Lack of study time is the problem, but what is the solution? Unsuccessful attempts to help traditional and online college students*. (NBER Working Paper No. 25036). National Bureau of Economic Research (NBER). Cambridge, MA.

Oreopoulos, P., & Petronijevic, U. (2018). Student coaching: How far can technology go? *Journal of Human Resources, 53*(2), 299-329.

Outes, 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*. (Policy Research Working Paper No. 9141). The World Bank. Washington, DC.

Paunesku, D., Yeager, D. S., Romero, C., & Walton, G. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, 1-10.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85.

SEE-MEDN. (2016). *Aprender 2016: Informe de resultados*. Secretaría de Evaluación Educativa, Ministerio de Educación y Deportes de la Nación. Ciudad Autónoma de Buenos Aires, Argentina.

SEE-MEDN. (2018a). *Aprender 2017: Informe de resultados de secundaria*. Buenos Aires, Argentina: Ministerio de Educación, Cultura, Ciencia y Tecnología, Secretaría de Evaluación Educativa.

SEE-MEDN. (2018b). *Aprender 2017: Informe de resultados, Buenos Aires, 6to año secundaria*. Ciudad Autónoma de Buenos Aires, Argentina: Secretaría de Evaluación Educativa, Ministerio de Educación, Cultura, Ciencia y Tecnología.

SEE-MEDN. (2018c). *Aprender 2017: Informe de resultados, Salta, 5to año secundaria*. Secretaría de Evaluación Educativa, Ministerio de Educación y Deportes de la Nación. Ciudad Autónoma de Buenos Aires, Argentina.

Sriram, R. (2014). Rethinking intelligence: The role of mindset in promoting success for academically high-risk students. *Journal of College Student Retention, 15*(4), 515–536.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J., Crosnoe, R., Muller, C., . . . Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature Human Behaviour, 573*, 364-369.

Yeager, D. S., Johnson, R., Spitzer, B., Trzesniewski, K., Powers, J., & Dweck, C. S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of personality and social psychology, 106*, 867-884.

Yeager, D. S., Miu, A., Powers, J., & Dweck, C. S. (2013). Implicit theories of personality and attributions of hostile intent: A meta-analysis, an experiment, and a longitudinal intervention. *Child development, 84*, 1651-1667.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Scheinder, B., Hinojosa, C., . . . Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374-391.

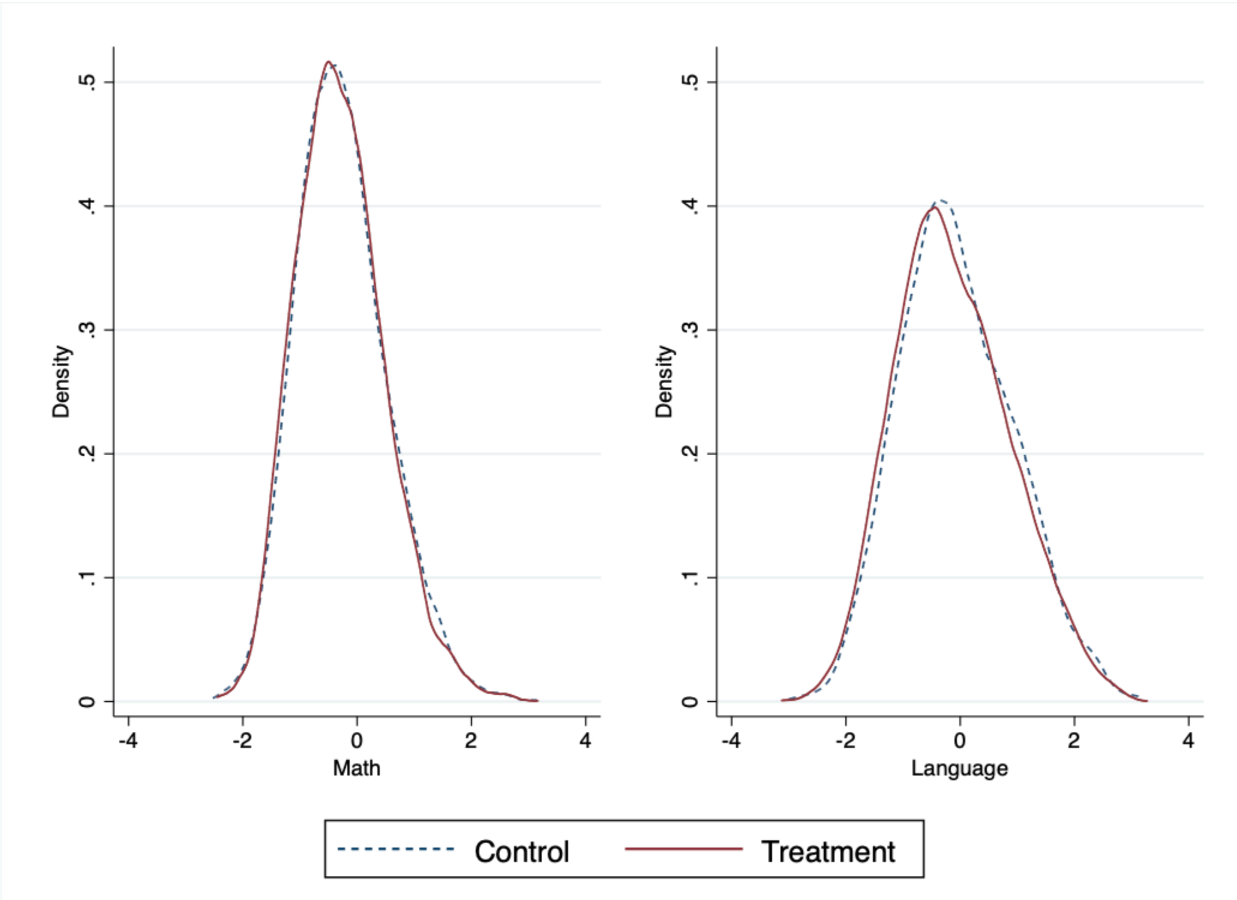
Yeager, D. S., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C. S. (2011). Adolescents' implicit theories predict desire for vengeance after peer conflicts: Correlational and experimental evidence. *Developmental Psychology, 47*(4), 1090-1107.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education They're not magic. *Review of Educational Research, 81*(2), 267-301.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement (4th ed.)*. Westport, CT: American Council on Education and Praeger Publishers.

Appendix A: Additional graphs and tables

Figure A.1: Endline test score distribution on the national student assessment (2017)



Notes: (1) The figure shows the distribution of scaled scores on math and language in the national student assessment, roughly two months after the intervention, by experimental group. (2) Test scores are scaled with respect to the national student achievement distribution to have a mean of 0 and a standard deviation of 1.

Table A.1: Comparison between in- and out-of-sample schools on school performance (2017)

		Out-of-sample schools						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All schools	All	Public	Public and non-rural	In-sample schools	Col. (5)- Col. (2)	Col (5)- Col. (3)	Col. (5)- Col. (4)
Panel A. Secondary school								
Number of students enrolled	323.252	163.116	74.363	270.75	474.58	311.464**	400.217***	326.681***
	(330.579)	(185.151)	(159.281)	(308.904)	(365.127)	(29.612)	(39.926)	(93.557)
Percentage of students who passed the grade	77.676	78.089	76.572	81.499	77.258	-.831	.686	-4.012
	(16.77)	(15.768)	(16.537)	(15.416)	(17.754)	(1.674)	(2.097)	(4.483)
Percentage of students who failed the grade	16.493	16.963	18.525	15.438	16.019	-.944	-2.506	.465
	(13.107)	(13.043)	(14.36)	(14.705)	(13.188)	(1.308)	(1.644)	(3.309)
Percentage of students who dropped out of school	5.831	4.948	4.903	3.063	6.723	1.775**	1.82*	3.548
	(7.803)	(6.307)	(6.417)	(4.456)	(8.995)	(.774)	(.992)	(2.228)
Percentage of students who repeated the grade	10.466	6.031	6.423	8.702	14.656	8.624***	8.232***	6.883**
	(11.606)	(11.118)	(13.811)	(12.691)	(10.465)	(1.094)	(1.468)	(2.744)
N (schools)	1127	925	802	288	202	1127	1004	490
Panel B. Grade 12								
Number of students enrolled	48.521	32.978	16.814	40.636	59.221	26.243***	42.407***	33.427*
	(47.233)	(30.467)	(26.294)	(45.007)	(53.429)	(5.051)	(8.368)	(17.306)
Percentage of students who passed the grade	75.347	76.513	76.909	83.844	74.129	-2.384	-2.78	-10.122
	(19.964)	(18.574)	(18.389)	(16.835)	(21.321)	(2.413)	(3.019)	(6.592)
Percentage of students who failed the grade	22.326	21.094	20.609	14.809	23.612	2.518	3.003	9.442
	(19.136)	(18.137)	(17.838)	(16.758)	(20.116)	(2.312)	(2.871)	(6.207)
Percentage of students who dropped out of school	2.327	2.393	2.482	1.347	2.259	-.134**	-.223*	.68
	(4.363)	(4.418)	(5.199)	(2.274)	(4.321)	(.528)	(.687)	(1.358)
Percentage of students who repeated the grade	2.545	1.573	1.126	.224	3.214	1.641***	2.089**	3.349*
	(5.739)	(4.649)	(4.23)	(.741)	(6.306)	(.632)	(1.008)	(1.929)
N (schools)	1127	925	802	288	202	1127	1004	490

Notes: (1) The table shows the means and standard deviations of all secondary schools in Salta (column 1), non-RCT schools (columns 2-4), and RCT schools (column 5). It also tests for differences between each group of non-RCT schools and RCT schools (columns 6-8). Panel A shows results for all secondary school students and Panel B for grade 12 students. (2) Dropout rates should be interpreted as an upper-bound estimate, as they actually refer to the percentage of students who leave their schools without asking for a pass to another school. (3) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.2: Comparison between in- and out-of-sample schools on student achievement (2016)

		Out-of-sample schools						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All schools	All	Public	Public and non-rural	In-sample schools	Col. (5)-Col. (2)	Col. (5)-Col. (3)	Col. (5)-Col. (4)
Math (IRT-scaled score)	-.085	.325	-.348	-.276	-.263	-.588***	.085**	.041
	(.917)	(1.059)	(.814)	(.816)	(.784)	(.017)	(.036)	(.045)
Language (IRT-scaled score)	.017	.371	-.169	-.005	-.136	-.507***	.033	-.104**
	(.945)	(.992)	(.891)	(.914)	(.882)	(.018)	(.04)	(.05)
Social Sciences (IRT-scaled score)	-.094	.295	-.264	-.262	-.257	-.552***	.007	.023
	(.897)	(.964)	(.825)	(.79)	(.814)	(.017)	(.038)	(.046)
Natural Sciences (IRT-scaled score)	-.034	.328	-.186	-.054	-.183	-.51***	.004	-.109**
	(.91)	(.951)	(.818)	(.826)	(.849)	(.018)	(.04)	(.049)
N (students)	14700	4895	1333	645	9805	14700	11138	10450

Notes: (1) The table shows the means and standard deviations of all secondary schools in Salta (column 1), non-RCT schools (columns 2-4), and RCT schools (column 5). It also tests for differences between each group of non-RCT schools and RCT schools (columns 6-8). The table shows results for grade 12 students. (2) All scores are standardized with respect to the national distribution to have a mean of zero and a standard deviation of one. (3) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.3: Balancing checks between experimental groups on school performance (2017)

	(1) In-sample schools	(2) Control schools	(3) Treatment schools	(4) Col. (3)- Col. (2)
Panel A. Secondary school				
Number of students enrolled	474.58	511.576	438.317	-69.947*
	(365.127)	(400.361)	(324.864)	(41.57)
Percentage of students who passed the grade	77.258	77.58	76.942	-.588
	(17.754)	(16.509)	(18.973)	(2.532)
Percentage of students who failed the grade	16.019	16.623	15.426	-1.244
	(13.188)	(13.483)	(12.933)	(1.876)
Percentage of students who dropped out of school	6.723	5.797	7.632	1.832
	(8.995)	(7.358)	(10.308)	(1.278)
Percentage of students who repeated the grade	14.656	16.812	12.542	-4.226***
	(10.465)	(10.898)	(9.615)	(1.435)
N (schools)	202	100	102	202
Panel B. Grade 12				
Number of students enrolled	59.221	64.434	54.06	-9.968
	(53.429)	(60.029)	(45.698)	(6.271)
Percentage of students who passed the grade	74.129	72.162	75.928	3.613
	(21.321)	(24.316)	(18.152)	(3.732)
Percentage of students who failed the grade	23.612	25.442	21.939	-3.189
	(20.116)	(22.837)	(17.258)	(3.514)
Percentage of students who dropped out of school	2.259	2.396	2.133	-.424
	(4.321)	(4.273)	(4.391)	(.746)
Percentage of students who repeated the grade	3.214	3.673	2.76	-.922
	(6.306)	(6.82)	(5.751)	(.864)
N (schools)	202	100	102	202

Notes: (1) The table shows the means and standard deviations of all schools in the sample (column 1), control schools (column 2), and treatment schools (column 3). It also tests for differences between control and treatment schools, using randomization fixed effects (column 4). Panel A shows results for all secondary school students and Panel B for grade 12 students. (2) Dropout rates should be interpreted as an upper-bound estimate, as they actually refer to the percentage of students who leave their schools without asking for a pass to another school. (3) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.4: Balancing checks between experimental groups on student achievement (2016)

	(1) In-sample schools	(2) Control schools	(3) Treatment schools	(4) Col. (3)-Col. (2)
Math (IRT-scaled score)	-.263	-.237	-.292	-.056
	(.784)	(.796)	(.77)	(.06)
Language (IRT-scaled score)	-.136	-.116	-.158	-.039
	(.882)	(.881)	(.882)	(.056)
Social sciences (IRT-scaled score)	-.257	-.224	-.295	-.064
	(.814)	(.822)	(.802)	(.051)
Natural sciences (IRT-scaled score)	-.183	-.162	-.206	-.05
	(.849)	(.864)	(.832)	(.055)
N (students)	9805	5215	4590	9805

Notes: (1) The table shows the means and standard deviations of all secondary schools in Salta (column 1), non-RCT schools (columns 2-4), and RCT schools (column 5). It also tests for differences between each group of non-RCT schools and RCT schools (columns 6-8). The table shows results for grade 12 students. (2) All scores are standardized with respect to the national distribution to have a mean of zero and a standard deviation of one. (3) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Heterogeneous ITT effects on students' beliefs by student characteristics (2017)

	Perceived difficulty index			Self-efficacy index			Usefulness of assessments index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.044	-0.003	0.000	0.064	0.029	0.001	-0.008	-0.004	0.048
	(0.053)	(0.046)	(0.050)	(0.056)	(0.056)	(0.061)	(0.046)	(0.039)	(0.039)
Female	0.146***			0.105***			0.183***		
	(0.041)			(0.040)			(0.035)		
Treatment x Female	0.118**			-0.118**			0.036		
	(0.059)			(0.057)			(0.056)		
Low SES		0.329***			-0.220***			0.092**	
		(0.041)			(0.044)			(0.041)	
Treatment x Low SES		0.060			-0.104*			0.031	
		(0.067)			(0.062)			(0.056)	
Repeated a grade			0.154***			-0.326***			0.038
			(0.043)			(0.044)			(0.038)
Treatment x Repeated			0.054			-0.018			-0.105**
			(0.061)			(0.070)			(0.051)
Observations	9148	8812	9186	9148	8812	9186	8994	8676	9053
R^2	0.007	0.016	0.005	0.005	0.015	0.021	0.008	0.004	0.002
Control mean	0.055	0.055	0.055	-0.110	-0.110	-0.110	0.059	0.059	0.059

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on students' perceived difficulty of school-related tasks, self-efficacy in math and language, and usefulness of classroom assessments, by sex, socio-economic status, and prior grade repetition. (2) The dependent variables in all regressions are the first principal component from principal component analyses of all variables in Tables 3-5, so they can be interpreted in standard deviation units. (3) Students' socio-economic status is calculated by an index developed by the national ministry of education based on students' household assets. Students' prior repetition status indicates whether they repeated a grade in primary or secondary school prior to the current school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.6: Heterogeneous ITT effects on students' beliefs by school characteristics (2017)

	Perceived difficulty index			Self-efficacy index			Usefulness of assessments index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.015	-0.043	0.012	0.045	-0.257*	0.000	0.002	0.173*	0.008
	(0.059)	(0.242)	(0.047)	(0.061)	(0.133)	(0.057)	(0.044)	(0.088)	(0.033)
Prior-year achievement	0.055			0.209**			-0.045		
	(0.081)			(0.101)			(0.084)		
Treatment x Achievement	-0.108			0.165			0.003		
	(0.149)			(0.160)			(0.117)		
Prior-year resources		0.286			0.186**			-0.065	
		(0.172)			(0.084)			(0.059)	
Treatment x Resources		-0.293			-0.139			0.087	
		(0.175)			(0.087)			(0.062)	
Prior-year supports			-0.030			-0.026			-0.020
			(0.022)			(0.030)			(0.017)
Treatment x Supports			-0.044			0.021			0.003
			(0.035)			(0.045)			(0.028)
Observations	9228	1706	8220	9228	1706	8220	8969	1640	7999
R^2	0.002	0.011	0.003	0.008	0.015	0.005	0.002	0.006	0.002
Control mean	0.055	0.055	0.055	-0.110	-0.110	-0.110	0.059	0.059	0.059

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on students' perceived difficulty of school-related tasks, self-efficacy in math and language, and usefulness of classroom assessments, by schools' prior-year achievement, resources, and supports. (2) The dependent variables in all regressions are the first principal component from principal component analyses of all variables in Tables 3-5, so they can be interpreted in standard deviation units. (3) Prior-year school achievement refers to the school-level average score across the four subjects assessed in the year prior to the intervention. The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.7: Heterogeneous ITT effects on student effort by student characteristics (2017)

	School tasks index			Personal development index			Existing obligations index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.031	-0.080	-0.069	0.039	0.009	-0.001	0.034	-0.002	-0.007
	(0.056)	(0.056)	(0.058)	(0.038)	(0.049)	(0.043)	(0.054)	(0.039)	(0.038)
Female	0.085**			0.096**			-0.368***		
	(0.036)			(0.039)			(0.039)		
Treatment x Female	-0.077			-0.033			-0.069		
	(0.050)			(0.056)			(0.056)		
Low SES		0.047			-0.306***			0.307***	
		(0.040)			(0.037)			(0.045)	
Treatment x Low SES		0.015			0.047			0.002	
		(0.061)			(0.059)			(0.063)	
Repeated a grade			-0.190***			-0.249***			0.306***
			(0.029)			(0.036)			(0.039)
Treatment x Repeated			-0.006			0.054			0.024
			(0.058)			(0.054)			(0.056)
Observations	8890	8578	8983	7865	7609	7906	8803	8510	8861
R^2	0.009	0.008	0.016	0.003	0.016	0.012	0.042	0.030	0.032
Control mean	-0.002	-0.002	-0.002	-0.087	-0.087	-0.087	0.099	0.099	0.099

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on students' effort on school-related tasks, personal development, or existing obligations, by sex, socio-economic status, and prior grade repetition. (2) The dependent variables in all regressions are the first principal component from principal component analyses of all variables in Table 6, so they can be interpreted in standard deviation units. (3) Students' socio-economic status is calculated by an index developed by the national ministry of education based on students' household assets. Students' prior repetition status indicates whether they repeated a grade in primary or secondary school prior to the current school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.8: Heterogeneous ITT effects on student effort by school characteristics (2017)

	School tasks index			Personal development index			Existing obligations index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.079	-0.226	-0.067	-0.004	-0.142**	-0.002	0.012	-0.094	0.015
	(0.061)	(0.198)	(0.054)	(0.046)	(0.066)	(0.036)	(0.052)	(0.147)	(0.044)
Prior-year achievement	0.042			0.040			-0.250***		
	(0.090)			(0.065)			(0.061)		
Treatment x Achievement	0.002			-0.110			0.108		
	(0.163)			(0.106)			(0.140)		
Prior-year resources		0.211			0.069			-0.151	
		(0.132)			(0.061)			(0.103)	
Treatment x Resources		-0.166			-0.033			0.105	
		(0.135)			(0.062)			(0.106)	
Prior-year supports			-0.040**			-0.055***			-0.006
			(0.020)			(0.017)			(0.026)
Treatment x Supports			0.024			0.034			0.046
			(0.045)			(0.032)			(0.038)
Observations	8863	1628	7893	7849	1426	7017	8769	1622	7808
R^2	0.009	0.014	0.011	0.001	0.009	0.004	0.017	0.017	0.014
Control mean	-0.002	-0.002	-0.002	-0.087	-0.087	-0.087	0.099	0.099	0.099

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on student achievement by students' sex, socio-economic status, and prior grade repetition. (2) The dependent variables in all regressions are the first principal component from principal component analyses of all variables in Table 6, so they can be interpreted in standard deviation units. (3) Prior-year school achievement refers to the school-level average score across the four subjects assessed in the year prior to the intervention. The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.9: Heterogeneous ITT effects on school climate by student characteristics (2017)

	Gets along with peers			Bullying index			Vandalism index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.004	-0.012	-0.007	0.035	0.047	0.054	0.044	0.068	0.058
	(0.009)	(0.008)	(0.008)	(0.051)	(0.057)	(0.059)	(0.058)	(0.062)	(0.065)
Female	-0.031***			-0.003			0.025		
	(0.009)			(0.045)			(0.038)		
Treatment x Female	-0.022*			0.019			0.006		
	(0.012)			(0.066)			(0.063)		
Low SES		-0.011			-0.013			-0.118***	
		(0.009)			(0.041)			(0.040)	
Treatment x Low SES		-0.012			0.011			-0.032	
		(0.014)			(0.068)			(0.061)	
Repeated a grade			-0.001			0.031			-0.063
			(0.010)			(0.045)			(0.041)
Treatment x Repeated			-0.022			-0.014			-0.022
			(0.015)			(0.069)			(0.062)
Observations	9032	8719	9117	9148	8812	9186	9148	8812	9186
R^2	0.007	0.003	0.003	0.004	0.003	0.004	0.013	0.016	0.014
Control mean	0.904	0.904	0.904	-0.034	-0.034	-0.034	-0.015	-0.015	-0.015

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on student achievement by students' sex, socio-economic status, and prior grade repetition. (2) The first dependent variable is the indicator variable from column 1 in Table 7. The second and third dependent variables in all regressions are the first principal component from principal component analyses of variables in cols. 2-7 in that table, so they can be interpreted in standard deviation units. (3) Students' socio-economic status is calculated by an index developed by the national ministry of education based on students' household assets. Students' prior repetition status indicates whether they repeated a grade in primary or secondary school prior to the current school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.10: Heterogeneous ITT effects on school climate by school characteristics (2017)

	Gets along with peers			Bullying index			Vandalism index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.022***	-0.038**	-0.014*	0.092	0.259*	0.063	0.032	0.327**	0.041
	(0.008)	(0.017)	(0.008)	(0.059)	(0.147)	(0.054)	(0.061)	(0.145)	(0.058)
Prior-year achievement	0.026**			-0.183*			-0.093		
	(0.013)			(0.095)			(0.107)		
Treatment x Achievement	-0.037			0.209			-0.044		
	(0.023)			(0.157)			(0.176)		
Prior-year resources		0.018			-0.115			-0.090	
		(0.028)			(0.135)			(0.095)	
Treatment x Resources		-0.020			0.176			0.097	
		(0.028)			(0.142)			(0.103)	
Prior-year supports			-0.002			0.001			0.035
			(0.006)			(0.025)			(0.026)
Treatment x Supports			-0.004			0.025			-0.056
			(0.008)			(0.044)			(0.046)
Observations	9011	1656	8029	9228	1706	8220	9228	1706	8220
R^2	0.002	0.007	0.002	0.004	0.012	0.004	0.011	0.026	0.010
Control mean	0.904	0.904	0.904	-0.034	-0.034	-0.034	-0.015	-0.015	-0.015

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on student achievement by students' sex, socio-economic status, and prior grade repetition. (2) The first dependent variable is the indicator variable from column1 in Table 7. The second and third dependent variables in all regressions are the first principal component from principal component analyses of variables in cols. 2-7 in that table, so they can be interpreted in standard deviation units. (3) Prior-year school achievement refers to the school-level average score across the four subjects assessed in the year prior to the intervention. The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.11: Heterogeneous ITT effects on student achievement by student characteristics (2017)

	Math (IRT-scaled score)			Reading (IRT-scaled score)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.016	-0.020	-0.042	-0.023	-0.038	-0.081
	(0.058)	(0.054)	(0.058)	(0.067)	(0.063)	(0.073)
Female	-0.204***			0.025		
	(0.031)			(0.038)		
Treatment x Female	-0.020			-0.056		
	(0.043)			(0.053)		
Low SES		-0.193***			-0.347***	
		(0.044)			(0.048)	
Treatment x Low SES		0.001			-0.023	
		(0.052)			(0.060)	
Repeated a grade			-0.243***			-0.398***
			(0.032)			(0.045)
Treatment x Repeated			0.058			0.076
			(0.045)			(0.063)
Observations	8787	8469	8825	8831	8508	8869
R^2	0.043	0.039	0.044	0.019	0.046	0.050
Control mean	-0.259	-0.259	-0.259	-0.055	-0.055	-0.055

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on student achievement by students' sex, socio-economic status, and prior grade repetition. (2) All scores have been scaled using a two-parameter logistic Item Response Theory (IRT) model with respect to the national distribution to have a mean of zero and a standard deviation of one. (3) Students' socio-economic status is calculated by an index developed by the national ministry of education based on students' household assets. Students' prior repetition status indicates whether they repeated a grade in primary or secondary school prior to the current school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.12: Heterogeneous ITT effects on student achievement by school characteristics (2017)

	Math (IRT-scaled score)			Reading (IRT-scaled score)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.036	-0.166*	-0.021	-0.062	-0.180	-0.053
	(0.065)	(0.087)	(0.054)	(0.076)	(0.114)	(0.069)
Prior-year achievement	0.728***			0.895***		
	(0.096)			(0.085)		
Treatment x Achievement	-0.228			-0.240		
	(0.197)			(0.229)		
Prior-year resources		0.166***			0.140	
		(0.060)			(0.097)	
Treatment x Resources		-0.134**			-0.081	
		(0.066)			(0.103)	
Prior-year supports			0.011			0.020
			(0.031)			(0.039)
Treatment x Supports			-0.015			-0.015
			(0.039)			(0.053)
Observations	8814	1611	7865	8865	1640	7910
R^2	0.078	0.070	0.022	0.068	0.060	0.014
Control mean	-0.259	-0.259	-0.259	-0.055	-0.055	-0.055

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on student achievement by students' sex, socio-economic status, and prior grade repetition. (2) All scores have been scaled using a two-parameter logistic Item Response Theory (IRT) model with respect to the national distribution to have a mean of zero and a standard deviation of one. (3) Prior-year school achievement refers to the school-level average score across the four subjects assessed in the year prior to the intervention. The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.13: Heterogeneous ITT effects on plans after secondary education by student characteristics (2017)

	Plans to work			Plans to study			Plans to do both		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.001	0.001	0.001	-0.004	0.005	-0.010	0.014	0.016	0.021
	(0.010)	(0.006)	(0.006)	(0.023)	(0.022)	(0.020)	(0.019)	(0.019)	(0.019)
Female	-0.049***			0.129***			-0.016		
	(0.006)			(0.016)			(0.016)		
Treatment x Female	-0.003			0.007			-0.002		
	(0.009)			(0.025)			(0.024)		
Low SES		0.017***			-0.123***			0.080***	
		(0.006)			(0.018)			(0.015)	
Treatment x Low SES		-0.008			-0.002			-0.022	
		(0.009)			(0.026)			(0.024)	
Repeated a grade			0.046***			-0.169***			0.099***
			(0.007)			(0.015)			(0.016)
Treatment x Repeated			-0.002			0.016			-0.020
			(0.011)			(0.023)			(0.024)
Observations	9148	8812	9186	9148	8812	9186	9148	8812	9186
R^2	0.018	0.005	0.016	0.023	0.019	0.031	0.001	0.005	0.009
Control mean	0.043	0.043	0.043	0.453	0.453	0.453	0.360	0.360	0.360

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on students' post-secondary plans by students' sex, socio-economic status, and prior grade repetition. (2) The dependent variables in the regressions are indicator variables for students who indicated that they plan to work, study, or do both after secondary education. (3) Students' socio-economic status is calculated by an index developed by the national ministry of education based on students' household assets. Students' prior repetition status indicates whether they repeated a grade in primary or secondary school prior to the current school year. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.14: Heterogeneous ITT effects on plans after secondary education by school characteristics (2017)

	Plans to work			Plans to study			Plans to do both		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.001	0.012	0.002	-0.012	-0.083	-0.007	0.039**	0.088	0.020
	(0.007)	(0.011)	(0.006)	(0.022)	(0.072)	(0.019)	(0.018)	(0.065)	(0.015)
Prior-year achievement	-0.020*			0.060**			-0.050**		
	(0.011)			(0.029)			(0.024)		
Treatment x Achievement	0.010			-0.044			0.110***		
	(0.019)			(0.055)			(0.039)		
Prior-year resources		-0.036***			0.061			-0.046	
		(0.011)			(0.044)			(0.045)	
Treatment x Resources		0.036***			-0.045			0.036	
		(0.012)			(0.048)			(0.047)	
Prior-year supports			0.004			-0.014			0.008
			(0.003)			(0.010)			(0.007)
Treatment x Supports			0.000			-0.030*			0.026**
			(0.004)			(0.017)			(0.013)
Observations	9231	1706	8221	9231	1706	8221	9231	1706	8221
R^2	0.005	0.007	0.005	0.007	0.008	0.013	0.002	0.006	0.004
Control mean	0.043	0.043	0.043	0.453	0.453	0.453	0.360	0.360	0.360

Notes: (1) This table shows the heterogeneous intent-to-treat (ITT) effects of the intervention on students' post-secondary plans by students' sex, socio-economic status, and prior grade repetition. (2) The dependent variables in the regressions are indicator variables for students who indicated that they plan to work, study, or do both after secondary education. (3) Prior-year school achievement refers to the school-level average score across the four subjects assessed in the year prior to the intervention. The indexes of school resources and supports are the first principal components from principal component analyses of questions in the 2016 survey of principals on the resources and supports for low-performing students at the school, respectively. (4) All estimations include randomization strata fixed effects. (5) * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix B: Prior research

Table B.1 offers an overview of prior randomized evaluations of growth mindset interventions, presented in chronological order of publication. Papers that included multiple studies are discussed separately, one after the other.

1. Search strategy

I did not follow a systematic procedure for searching for studies for inclusion in this appendix. First, I tracked evaluations of growth mindset interventions cited in prior reviews—specifically, Dweck, Walton and Cohen (2014) and Yeager and Walton (2011). Then, I tracked evaluations cited in those evaluations. Next, I searched for any additional evaluations in the websites of the developers of the growth mindset intervention, including Carol Dweck, Gregory Walton, and David Yeager. Finally, I included evaluations referred to me by colleagues.

2. Inclusion criteria

I included studies conducted from 2000 to 2019 in primary, secondary, or tertiary education, in developed or developing countries. I included evaluations of interventions that conveyed the message that intelligence is malleable in a variety of ways (e.g., through readings, videos, and/or mentors). I did not include, however, descriptive or correlational studies of growth mindset or evaluations of other types of interventions (e.g., purpose for learning interventions). I only included studies that employed randomized experiments, either in the lab or the field. I also omitted studies in the papers cited below that did not meet these criteria. I included both journal articles and working papers, but not students' theses.

3. Reporting effects

I report the direction of statistically significant effects on all outcomes measured in each study in the units used by the authors. Specifically, I report the difference between experimental groups adjusted for baseline performance whenever available. If it is not available, I report the difference between experimental groups without any covariates other than randomization blocks if applicable. If neither is available, I report the difference between experimental groups adjusted for baseline performance and any other covariates that the authors included. Finally, I report effects on both the average study participant and on sub-groups of participants. I only report magnitudes of effects on test scores and percentage points (otherwise, I report the sign of the effect) because effect sizes on scales are rarely comparable across studies.

Table B.1: Impact evaluations of growth mindset interventions

(1) Study	(2) Setting	(3) Sample	(4) Randomization	(6) Intervention	(7) Results
Aronson, Fried, and Good (2002)	Stanford, CA (United States)	79 students in one college (after exclusions)	Student-level: • T: 28 • C1: 28 • C2: 23	• T: Students were invited to three one-hour sessions in a lab. In each session, they were shown a brief video clip on the malleability of intelligence. Then, they were asked to respond to letters from at-risk middle-school students to convince them that intelligence is expandable. • C1: Students were shown a brief video clip on intelligence being composed of many talents. Then, they were asked to respond to similar letters as students in T group. • C2: No intervention.	• T students had lower SAT scores than C students prior to the intervention, so all analyses include baseline SAT as covariate. • Effects are only presented disaggregated by race (blacks v. whites), not overall. • T students reported viewing intelligence as more malleable than C1 students, but there was no difference between C1 and C2 students and no heterogeneous effects by race. • T students were still more likely to view intelligence as malleable nine weeks after the intervention and African American students were more influenced by the intervention. • T students also reported higher enjoyment of the educational process and identification with academic achievement nine weeks after the intervention and African American students were more influenced by the intervention. • T students also reported (marginally statistically significant) lower levels of identification with academic achievement and African American students were more influenced by the intervention. • The intervention did not affect students' perceptions of stereotype threat.

					• T students had higher GPAs than C1 and C2 students. • Stronger endorsement of intelligence as malleable 9 weeks after the intervention was *negatively* associated with higher GPAs. The authors conjecture that this may be due to a protective strategy, small sample size, or restriction of range in the malleability scale.
Good, Aronson, and Inzlicht (2003)	Texas (United States)	138 grade 7 students in 1 junior high school (after exclusions)	Student-level: • T1:~34 • T2:~34 • T3:~34 • C:~32	• All students were assigned to a mentor for two 90-minute sessions and read web content. Mentors communicated with students through text and e-mail between meetings. • T1: Mentors taught students about the expandable nature of the brain. Students read web content on how the brain grows through problem solving. • T2: Mentors taught students that academic setbacks should be attributed to situations and not intelligence. Students read web content on how academic setbacks can be explained by the transition to junior school. • T3: Mentors and web content taught students the messages from T1 and T2. • C: Mentors and web content taught students learned about the harmful effects of drug use on health and academics.	• Students with math and reading scores below 1.5 times the IQR or above 1.5 the IQR were excluded from the analysis. • Effects are only presented disaggregated by sex (males v. females), not overall. • Female students performed below male students in math test scores in the C condition, but there were no statistically significant differences between sexes in the treatment conditions. • Females in the treatment conditions outperformed females in the control condition. • T1 and T2 students had higher test scores in reading than C students, but there was no difference between T3 and C students at the end of the year.
Mendoza-Denton, Kahn, and Chan (2008), Study 1	Western region (United States)	65 students in 1 university (after exclusions)	Student-level: • T1: 15 • T2: 17 • T3: 20	• All students received one of four press releases before completing a math test: - T1: First paragraph confirmed that Asians outperform whites in math	• All analyses include personal investment in performance and ethnicity as covariates.

			• T4: 13	(confirmed stereotype); the second paragraph described innate ability as the most important predictor of math achievement (entity prime). - T2: Same as T1, but second paragraph described effort as most important predictor of math achievement (incremental prime). - T3: Same as T1, but first paragraph disconfirmed the Asian-white achievement gap (disconfirmed stereotype). - T4: First paragraph same as T3 and second paragraph same as T2.	• There were no main effects from the confirmation of the achievement gap or the entity prime by themselves. • When the stereotype was confirmed, students who received the entity prime performed better on a math test (i.e., T1 students outperformed T2 students). • When the stereotype was disconfirmed, students who received the entity and incremental primes performed no differently (i.e., no difference in performance between T3 and T4 students).
Study 2	Western region (United States)	186 students in 1 university (after exclusions)	Student-level: • T1: 53 • T2: 54 • T3: 41 • T4: 38	• All students received one of four press releases before completing a math test: - T1: First paragraph confirmed that males outperform females in math (confirmed stereotype); the second paragraph described innate ability as the most important predictor of math achievement (entity prime). - T2: Same as T1, but second paragraph described effort as most important predictor of math achievement (incremental prime). - T3: Same as T1, but first paragraph disconfirmed the male-female achievement gap (disconfirmed stereotype). - T4: First paragraph same as T3 and second paragraph same as T2.	• All analyses include personal investment in performance and ethnicity as covariates. • When the stereotype was confirmed, male students who received the entity prime performed better on a math test (i.e., among males, T1 students outperformed T2 students) and found the test less difficult. • When the stereotype was disconfirmed, male students who received the entity prime performed better (i.e., among males, T4 students outperformed T3 students), but there were no statistically significant differences in perceived difficulty of the test between these groups. • When the entity prime was provided, male students who received the confirmation of the stereotype performed better (i.e., among males, T1 students outperformed T3 students) and found the test less difficult. • No differences when the incremental prime was provided (i.e., among

					males, T2 students performed on par with T4 students). • When the stereotype was confirmed, female students performed worse (i.e., among females, T3 and T4 students outperformed T1 and T2 students) and found the test more difficult. • No differences between any other two groups of females.
Blackwell, Trzesniewski, and Dweck (2007), Study 2	New York City, NY (United States)	91 grade 7 students in 1 public secondary school (after exclusions)	Student-level: • T: 48 • C: 43	• T: Students read scientific articles and participated in activities demonstrating how the brain responds to learning something new, and discussions on avoiding labels and benefits of learning across eight weekly 25-minute sessions. • C: Students read articles and participated in activities but the content differed (Workshop focus on mnemonic strategies, activities addressed personal and academic issues). The number and duration of sessions were the same as in the T group. • Additionally, both groups received lessons on the brain's functions, on the pitfalls of stereotyping , and study strategies.	• T and C students performed on par on the material taught in the workshop, but T students scored 31 pp. higher than C students on items assessing the incremental theory of intelligence. • T students endorsed an incremental theory of intelligence more strongly after the intervention, but there was no statistically significant difference among C students during the same period. • T students were 18 pp. more likely to be identified as showing positive change in classroom motivation by their teachers than C students. • The math test scores of all students decreased between two time points prior to the intervention; the decline continued among C students after the intervention, but it was reversed (though not offset) among T students.
Yeager, Trzesniewski, Tirri, Nokelainen, and Dweck (2011), Study 3	Multiple regions (Finland)	187 grade 9-10 students in 6 schools (after exclusions)	Student-level; treatment arm size not specified	• T: Students read and summarized stories about bullying as part of an online reading comprehension task emphasizing that people can change. • C: Students also read and summarized stories about bullying, but without the emphasis on malleable personality.	• T students endorsed the belief that people can change more strongly than C students. • T students were also less likely than C students to indicate that people would behave in the same way across different scenarios, consistent with their difference in beliefs that people can change.

					• T students endorsed a lower desire for vengeance than C students, but they were no more likely to choose neutral or prosocial strategies to solve hypothetical situations. • None of these effects varied by sex, age, or pre-intervention frequency of being bullied or desire for vengeance. • T students were 19 pp. less likely to believe bullies were bad people than C students. • T students had lower feelings of hatred towards bullies and of shame towards self than C students, but T and C students did not differ in their feelings of sadness after bullying.
Dommett, Devonshire, Sewter, and Greenfield (2013)	Gloucestershire (South-west England)	383 grade 7 students in 5 schools (after exclusions)	School-level: • ST1xTT1: 28 • ST1xTT2: 29 • ST1xTC: 32 • ST2xTT1: 28 • ST2xTT2: 25 • ST2xTC: 17 • SC1xTT1: 21 • SC1xTT2: 19 • SC1xTC: 28 • SC2xTT1: 19 • SC2xTT2: 25 • SC2xTC: 30 • SC3xTT1: 30 • SC3xTT2: 28 • SC3xTC: 24	• Schools were randomly assigned to one of five conditions: • ST1: Students received material on neuroscience through teacher-directed workshops (teacher, neuroscience). • ST2: Same as ST1, but students received material through an interactive computer software (computer, neuroscience). • SC1: Same as ST1, but students received material on study skills (teacher, study skills). • SC2: Same as ST2, but students received material on study skills (computer, study skills). • SC3: No intervention. • Then, within each school, math teachers were randomly assigned to one of three conditions, all via the same software used for students:	• ST1 and ST2 students (combined, across teacher interventions) had similar levels of beliefs in intelligence, academic performance, and effort as SC3 students, 19 months after the intervention. • ST1 and ST2 students (combined, across teacher interventions) had the same performance on a math assessment as SC3 students, 19 months after the intervention. • The same two patterns above hold for comparisons between SC1 and SC2 students (combined, across teacher interventions) and SC3 students, 19 months after the intervention. • TT1, TT2, and TC (combined, across student interventions) had similar levels of beliefs and performance, 19 months after the intervention. • ST1 and SC1 students (combined, across teacher interventions) had similar levels of beliefs and performance as ST2 and SC2

				- TT1: Professional development workshops on neuroscience (computer, neuroscience). - TT2: Same as TT1, but teachers received PD on study skills (computer, study skills). - TC: No intervention. • All conditions were crossed, such that there were 15 experimental arms in total.	students, 19 months after the intervention.
Yeager, Trzesniewski, and Dweck (2013)	San Francisco, CA (United States)	230 grade 9-10 students in 1 high school (after exclusions)	Student-level: • T: 82 • C1: 82 • C2: 82	• T: Students participated in six 50-minute sessions across 3 weeks: - In sessions 1 and 2, they learned about brain and personality malleability through scientific articles and team-building activities, worksheets, and lectures. - In sessions 3 and 4, they applied the incremental theory to resolve hypothetical conflicts. - In sessions 5 and 6, they learned that motives behind actions can be changed through skit performances, written assignments, and focus groups. • C1: Same sessions as T group, but: - In sessions 1 and 2, students learned about how the brain responds to learning. - In sessions 3 and 4, they used coping strategies to resolve hypothetical conflicts. - In sessions 5 and 6, they learned to practice thinking positively and avoiding "all or nothing" thinking. • C2: No intervention.	• Nearly all analyses combine C1 and C2. • All analyses include sex, grade, and baseline peer nominations for aggressive behavior as covariates. • T students endorsed the entity theory of intelligence less strongly than C students two weeks after the intervention. • T students displayed lower levels of aggression in a game than C students one month after the intervention. • T students displayed higher levels of pro-social behavior in a game than C students one month after the intervention. • T students had reduced their conduct problems more than C students three months according to teacher reports after the intervention (especially, among victims of peer aggression). • T and C1 students had a lower association between victimization and depressive symptoms than C2 students. • T students were also less likely to be absent to school than C students.

Yeager, Miu, Powers, and Dweck (2013), Study 2	Oakland, CA and a medium-sized city in GA (United States)	63 grade 9 students in 2 schools (after exclusions)	Student-level; treatment arm size not specified	• T: 15-minute activity, which had three parts: - Students read a scientific article on how people's behaviors are controlled by their thoughts and feelings in their brains, which have constant potential for plasticity; - They read notes from upperclassmen endorsing a malleable view of personality; and - They wrote notes to future students describing the malleability of people's traits. • C: Same as T, but emphasized the malleability of academic skills such as study skills.	• All analyses combine the CA and GA samples. • All analyses include an indicator variable for one of the samples. • T students endorsed the entity theory of intelligence less strongly than C students. • T students exhibited fewer attributions of hostile intent in an ambiguous provocation scenario than C students. • T students were less likely to respond negatively and more likely to respond positively to the provocation scenario than C students.
Study 3	San Francisco, CA (United States)	78 grade 9 students in 1 secondary school (after exclusions)	Student-level; treatment arm size not specified	• On the first week of school, teachers provided an overview to T and C students of how the brain changes and learns. • T: Two weeks later, students completed the same activity as T students in study 2. • C: Two weeks later, students completed the same activity as C students in study 2, but focusing on the malleability of athletic skills.	• All analyses include race, sex, classroom, and endorsement of entity theory of intelligence at baseline as covariates. • T students endorsed less strongly the entity theory of intelligence than C students. • T students were less likely to attribute a hypothetical peer's negative action to hostile intent than C students, eight months after the intervention. • T students showed lower desire for revenge, eight months after the intervention. • T students were less likely to respond negatively and more likely to respond positively to a provocation scenario than C students.
Sriram (2014)	South-western region (United States)	105 students in 1 university (after exclusions)	Student-level: • T: 60 • C: 45	• Students in a remedial course were asked to complete four 5-minute web-based activities. • T: Each session included a quote that illustrated the growth mindset	• T students endorsed less strongly the entity theory of intelligence than C students after the intervention. • All subsequent analyses include academic discipline, academic self-

				theory, questions preparing students to engage with a movie clip, a clip that portrayed an issue related to a fixed or growth mindset, questions asking students to reflect on the clip, another clip from a lecture on intelligence, the brain, and its malleability, a summary of research on mindset, and teaser questions for the next session. • C: Same as T, but sessions focused on study skills.	confidence, commitment to college, general determination, goal striving, and study skills at baseline as covariates. • T students reported higher levels of academic effort than C students. • T students did not have higher academic achievement than C students (using the sum of students' percentile ranks in the SAT and high school class rank as covariates).
Yeager et al. (2014), Study 2	Northern CA (United States)	78 grade 9 students in 1 high school (after exclusions)	Student-level; treatment arm size not specified	• On the first week of school, teachers provided an overview to T and C students of how the brain changes and learns. • T: Two weeks later, students completed the same activity as T students in studies 2 and 3 in Yeager, Miu, et al. (2013), but the activity lasted 25 minutes (instead of 15 minutes). • C: Two weeks later, students completed the same activity as C students in study 3 in Yeager, Miu, et al. (2013), but the activity lasted 25 minutes (instead of 15 minutes).	• T students endorsed less strongly the entity theory of intelligence than C students, one to two days after the intervention. • T students responded less negatively than C students to exclusion in an online game. • T students reported lower stress than C students, eight months after the intervention. • T students reported fewer symptoms of physical illness than C students, eight months after the intervention. • T students had higher grades than C students by the end of the school year (mostly, by slowing a decline in grades among T students).
Study 3	CA (United States)	131 grade 9 students in 1 high school (after exclusions)	Student-level; treatment arm size not specified	• T: Same as study 2, but materials were on computer (not paper) and students could read it in Spanish. • C: Same as study 2, but materials were on computer and could be read in Spanish.	• T students *did not* endorse the entity theory of intelligence less strongly than C students, immediately after the intervention. • T students responded less negatively than C students to exclusion in an online game. • T students reported lower stress than C students, eight months after the intervention.

					• T students reported fewer symptoms of physical illness than C students, eight months after the intervention. • The average T student did not have higher grades than the average C student by the end of the school year, but among students who endorsed the entity theory of intelligence, T students had higher grades than C students.
Paunesku, Yeager, Romero, and Walton (2015)	East, West and South-west United States	1,594 grade 9-12 students in 13 high schools (after exclusions)	Student-level; treatment arm size not specified	• All teachers created an account in a study website and scheduled two 45-minute sessions, roughly two weeks apart. • T1: A session in which students read an article describing the brain's ability to grow and reorganize based on hard work and good strategies on challenging tasks and participated in two writing exercises: one in which they summarized the article and another one in which they advised a hypothetical student (growth mindset). • T2: Same as T1, but session focused on beyond-the-self-goals (sense of purpose). • T3: Students participated in both sessions above (combined). • C: Same as T3, but sessions focused on transition to high school and economic self-interest.	• T1 students endorsed less strongly the entity theory of intelligence than C students (including baseline beliefs about intelligence as a covariate). • Effects on GPA are only presented disaggregated by students' propensity to drop out of school, not overall. • Among those at risk of dropping out, T1, T2, and T3 students had higher GPAs than C students (using baseline GPA, race, sex, and school as covariates). • Among students not at risk of dropping out, the GPAs of the four groups were comparable. • Among those at risk of dropping out, T1, T2, and T3 students (combined) were more likely to earn satisfactory grades in core courses than C students. (This analysis was not conducted for students not at risk of dropping out).
Yeager, Lee, and Jamieson (2016), Study 2	CA, NY, TX, VA, and NC (United States)	3,276 grade 9 students in 10 schools (after exclusions)	Student-level: • T: 1,630 • C: 1,646	• All students participated in two one-period online sessions (one to four weeks apart): • T: Students read an article on the malleability of intelligence, they are asked to share an example in which practice improved their skills, and	• T students reduced their endorsement of the entity theory of intelligence more than C students. • All subsequent analysis use baseline achievement as covariate. • Students with low baseline achievement (1 SD below mean at

				they were asked to write a letter encouraging a student struggling in school. • Several changes from prior intervention: - Reading included quotes from admired adults and celebrities. - More and more diverse writing exercises. - Focus on purpose for practice. - Use of bullet points instead of paragraphs. - Less information on each page. - Inclusion of data from research. - Use of examples that are more relevant to high school students. • C: Same as T, but activity focused on the transition to high school.	baseline) had a higher GPA in grade 9. • Students with high baseline achievement (1 SD above mean at baseline) did not have a higher GPA in grade 9. • T students had lower courses with failing grades than C students in grade 9. • T students were more likely to choose a more difficult math homework than C students. • T students were less likely to make fixed-trait, person-focused attributions than C students. • T students were less likely to adopt performance avoidance goals.
Ehrlinger, Mitchum, and Dweck (2016), Study 2	Not specified	94 students in 1 university (after exclusions)	Student-level: • T1: 47 • T2: 47	• T: Students read an article that described scientific evidence that intelligence is malleable. • C: Same as T, but article purported to offer scientific support for intelligence being fixed.	• T students spent more attention to difficult problems in a test than C students. • T students were less overconfident about their performance on the test than C students.
Yeager et al. (2016), Study 1	Rochester, NY (United States)	60 grade 9-11 students, number of high schools not specified (after exclusions)	Student-level: • T: 30 • C: 30	• All students completed a 25-minute reading and writing activity. • T: Students read an article that indicated that, if a person is excluded or victimized, it is not because of a fixed personal deficiency, and people who exclude others are not inherently bad. Then, students are asked to write to a future student to persuade them to hold an incremental theory. • C: Same as T, but article focused on adjusting to the physical environment of high school (lockers, hallways, and smells).	• T students reported lower levels of threat appraisals than C students. • T students had lower cortisol levels than C students. • T students performed better on a mental math task than C students.

Study 2	Region not specified (United States)	205 grade 9 students, number of high schools not specified (after exclusions)	Student-level; treatment arm size not specified	• T: Same as study 1. • C: Same as study 1.	• T students had higher GPAs in grade 9 (using prior achievement, advance placement course enrollment, and sex as covariates). • T students did not have lower levels of threat appraisals than C students, but the relationship between daily stressors and threat appraisal was stronger among C students. • T students did not exhibit a lower relationship between daily stressors and cortisol than C students (because this relationship was not observed among C students).
Outes, Sánchez, and Vakis (2017)	Ancash, Junin, and Lima (Peru)	Not specified number of grade 9 students in 800 public high schools	School-level: • T: 400 • C: 400	• T: Schools received an intervention packet: - an article on how individuals can grow their intelligence if they persevere through challenges; - discussion questions that asked students to reflect on examples from their lives in which they could have implemented the lessons from the reading; and - instructions that asked students to write a letter to younger peers on the lessons from the reading. • The packet also included instructions for the principal and the *tutor* in charge of each grade. It was implemented in one 90-minute session during the time allotted for *tutorías*, as in the present study. • C: Schools conducted their regular *tutorías*, as in the present study.	• T students performed .05 SDs better than C students in math, but no better in reading. • Average effects driven entirely by Ancash region (.2 SDs in math, .12 SDs in reading); positive but statistically insignificant effects in Junín and precisely estimated null effects in Lima. • Effect of receiving the intervention was .15 SDs in math (again, entirely driven by Ancash). • T students were 1 pp. more likely to report intending to pursue post-secondary education. • Effect on post-secondary aspirations driven entirely by Ancash region (5 pp.), which also saw an improvement in students' self-beliefs in math. Junín region saw a negative effect on students' self-beliefs in math (8 pp.) • No average effects on students' self-beliefs about math ability, reading ability, teacher effort, or teacher support for students' socio-emotional development.

Bettinger, Ludvigsen, Rege, Solli, and Yeager (2018)	Rogaland county (Norway)	258 grade 8 students in 1 high school (after exclusions)	Student-level: • T: 179 • C: 175	• All students participated in two 45-minute online sessions: • T: In session 1, students read an article about research in neuroscience that demonstrates the brain's potential to grow and change and they were asked to summarize the article and explain how it related to their own lives. In session 2, students read endorsements of the growth mindset. Then, they read about how to use the growth mindset for beyond-the-self goals. • C: Same as T, but sessions focused on the brain's functions and localization.	• T students endorsed an entity theory of intelligence .56 SDs less strongly than C students. • T students were more likely to want to solve more challenging math questions than C students in a "make-a-worksheet" exercise. • Many students only completed the first 10 questions of an algebra test, so all analyses on this test focus on those questions. • T students performed on par with C students on these questions (the results are only statistically significant when using GPA, math grade, vocational track, and indicator variables for female students, students older than 16 years, and baseline fixed mindset score as covariates). • The conditional effect on algebra performance is entirely driven by students who initially had a fixed mindset. • Effects on algebra were larger for T students with an initially low GPA and those in the vocational track.
Broda et al. (2018)	East Lansing, MI (United States)	6,529 students in 1 college (after exclusions)	Student-level (within race/ethnic group blocks): • T1: 2,135 • T2: 2,172 • C: 2,222	• T1: Students read a short article introducing the concept of brain plasticity. Then, they were asked to identify moments in their lives when they have adopted a growth mindset. Finally, they were asked to write a letter to future first-year students based on lessons from the article (growth mindset). • T2: Students read stories ostensibly from upperclassmen on a recent survey on the challenges of starting college. The stories are matched with the reader's gender and	• All effects are presented by sub-group, not overall. • Among Latino/a students, T1 students had higher GPAs in the fall and spring semesters of their freshmen year than C students. • Among African American and white students, no such effects were found. • Even when including baseline belonging, baseline mindset, ACT score, high school GPA, first-generation status, and Pell grant eligibility as covariates, among

				race/ethnicity. Then, were asked to write short reflective responses on the meaning of these stories for their own lives (social belonging). • C: Same as T2, but reading focused on the physical environment of college (e.g., weather, class schedule, navigating the campus, and finding places to eat).	Latino/a students, T1 students did not attempt or complete more credits or were more likely to be enrolled full-time than C students. • No effects were found for T2 students.
Polley (2018)	Dhaka (Bangladesh)	1,016 grade 6-8 students in 2 secondary schools	Student-level (within school/grade/sex/ baseline performance/ network blocks); treatment arm size not specified	• T1: Four one-hour weekly sessions focusing on the relationships between brain strength and learning and between effort and success. • T2: Four one-hour weekly sessions focusing on how information is filtered through the brain and the relationship between effort and success. • C: Free period (supervised study time).	• T1 students were twice as likely as C students to be mentioned by teachers as having increased effort in the semester following the intervention. • T1 students performed .12 SDs higher on math quizzes than C students. The effect on T1 students is lower among already hard-working students and boys. • T1 students in grades 6 and 7 performed .11 SDs higher in math, .15 SDs in science, and .15 SDs in world studies than C students, but no better in English, Bengali, religion, physical education, or home economics. • T1 students in grade 8 performed no better than T2 or C students on any subject of the board exams. • T1 students were less overconfident than C students, but T2 had no effect on overconfidence. • No impact on self-reported effort ranking, study hours, or friend's effort. • No differences between T1 and T2 students on average.
Yeager et al. (2019)	Nationally representative sample (United States)	12,486 grade 9 students in 65 middle schools	Student-level (within race-by-performance blocks); treatment arm size not specified	• All students participated in two self-administered 25-minute online sessions: • T: In session 1, students were asked to read a passage on the malleability	• Among low-performing students, T students endorsed the entity theory of intelligence less strongly than C students right after session 2.

				of intelligence. In session 2, they were asked to apply the ideas from the passage to their own lives. • C: Same as T, but the sessions focused on brain functions.	• Among the same students, T students earned higher GPAs in core classes than C students. • Among high-performing students, T students endorsed the entity theory less strongly than, but had similar GPAs to, C students. • Treatment effects were smaller in schools with higher achievement levels. • Treatment effects were larger in schools with greater challenge-seeking behavior (measured at endline among control students), but did not vary based on baseline beliefs of intelligence. • Among high-performing students, T students were 3 pp. more likely to take Algebra II or higher courses in grade 10 than C students.
Gandhi, Watts, Masucci, and Raver (2019)	Chicago, IL (United States)	404 grade 11 students in 275 high schools (after exclusions)	Student-level: • T: 211 • C: 193	• T: students completed two online sessions: - In year 1, students were asked to write about problems in the world/community they wanted to solve. Then, they read about students working hard because they want to have a positive impact on the world. They were asked to think about their goals and write about how working hard can help them achieve them (purpose for learning). - In year 2, students were asked to elicit what issues in the world mattered to them. They were presented with information and vignettes on the learning mindset. They answered questions about how to use this mindset to strengthen their brain. Finally, students wrote	• T students were less likely to have a single parent and their parents worked more hours per week at baseline. • In year 1, T students were more likely to describe a picture in a way that was more aligned with meaningful goals than C students. • No effect of T on college knowledge, anxiety, or belongingness. • Negative and marginally statistically significant effect of T on GPA. • In year 2, T students' meaning-making of their academic environment was more consistent with a learning mindset than that of C students. • No effect on college knowledge, anxiety, or socio-political motivation.

				about how to use the mindset in their classroom (growth mindset). • C: Same as T, but: - In year 1, students were asked to reflect on how their lives had changed between middle and high school. Then, they read about how the brain learns through classroom assignments. They were asked to write to a hypothetical incoming middle school student. - In year 2, students were given information on brain science and how health behaviors can improve it. Students then wrote about how to keep their brain healthy during the year.	
Rege et al. (2019), Study 1	Nationally representative sample (United States)	14,472 students in 76 high schools (after exclusions)	Student-level; treatment arm size not specified	• T and C students completed two 25-minute online sessions, one to four weeks apart: • T: Students completed a reading that indicated that doing challenging work can strengthen one's abilities, explained how neurons worked and distinguished between strong and weak neural connections, and indicated that hard work makes neural connections more efficient. It presented evidence that during adolescence the brain can learn and grow and explained that stronger brains can be helpful regardless of students' plans. Students were then asked to advise a future struggling ninth-grader based on the reading and to explain how they planned to use their stronger brains to achieve meaningful goals. • C: Same as T, but the reading focused on the transition to high	• T students were less likely to endorse a fixed mindset. • T students were more likely to select challenging math exercises than C students. • T students chose more challenging math exercises than C students, and this effect did not vary by sex, race/ethnicity, parental education, ninth-grade course level, or status as prior low-achieving student.

				school. The reading included stories and opinions from upperclassmen. Students were then asked open questions and provided their reactions to the reading.	
Study 2	Akershus and Rogaland counties (Norway)	6,541 students in 49 high schools (after exclusions)	Student-level; treatment arm size not specified	• T: Same as study 1. • C: Same as study 1.	• T students were less likely to endorse a fixed mindset than C students. • T students were more likely to select challenging math exercises than C students. • T students chose more challenging math exercises than C students, and this effect did not vary by sex, prior math grades, or school type (schools differed in the timing of selection of math courses). • T students were 3 pp. more likely to take a theoretical (more challenging) math class than C students. • When schools allowed students to choose math courses after the intervention, T students were 6 pp. more likely to take a theoretical math class. When schools allowed students to choose math courses before the intervention, T students were 2 pp. more likely to do so

Source: Author’s elaboration.

4. References

Aronson, J., Fried, C. B., & 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*(2), 113-125.

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. S. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization, 146*, 1-15.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development, 78*(1), 246-263.

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., & Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness, 11*(3), 317-338.

Dommett, E. J., Devonshire, I. M., Sewter, E., & Greenfield, S. A. (2013). The impact of participation in a neuroscience course on motivational measures and academic performance. *Trends in Neuroscience and Education*, 122–138.

Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94–100.

Gandhi, J., Watts, T. W., Masucci, M. D., & Raver, C. C. (2019). *The effects of two mindset interventions on low-income students' academic and psychological outcomes*. Unpublished manuscript. New York University. New York, NY.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*(645-662).

Mendoza-Denton, R., Kahn, K., & Chan, W. (2008). Can fixed views of ability boost performance in the context of favorable stereotypes? *Journal of Experimental Social Psychology, 44*(4), 1187–1193.

Outes, I., Sánchez, A., & Vakis, R. (2017). *Cambiando la mentalidad de los estudiantes: Evaluando el impacto de Expande tu Mente sobre el rendimiento académico*. (Documento de Investigación No. 83). Grupo de Análisis para el Desarrollo (GRADE). Lima, Peru.

Paunesku, D., Yeager, D. S., Romero, C., & Walton, G. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, 1-10.

Polley, T. (2018). *Believing is achieving: Targeting beliefs about intelligence to increase learning in Bangladeshi secondary schools*. Unpublished manuscript. Duke University. Durham, NC.

Rege, M., Hanselman, P., Foldøy Solli, I., Dweck, C. S., Ludvigsen, S., Bettinger, E., . . . Yeager, D. S. (2019). *How can we inspire nations of learners? An investigation of growth mindset and challenge-seeking in two countries*. University of Stavanger. Norway.

Sriram, R. (2014). Rethinking intelligence: The role of mindset in promoting success for academically high-risk students. *Journal of College Student Retention, 15*(4), 515–536.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J., Crosnoe, R., Muller, C., . . . Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature Human Behaviour, 573*, 364-369.

Yeager, D. S., Johnson, R., Spitzer, B., Trzesniewski, K., Powers, J., & Dweck, C. S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of personality and social psychology, 106*, 867-884.

Yeager, D. S., Lee, H. Y., & Jamieson, J. P. (2016). How to improve adolescent stress responses: Insights from integrating implicit theories of personality and biopsychosocial models. *Psychological Science, 27*(8), 1078–1091.

Yeager, D. S., Miu, A., Powers, J., & Dweck, C. S. (2013). Implicit theories of personality and attributions of hostile intent: A meta-analysis, an experiment, and a longitudinal intervention. *Child development, 84*, 1651-1667.

Yeager, D. S., Trzesniewski, K. H., & Dweck, C. S. (2013). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child development, 84*, 970–988.

Yeager, D. S., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C. S. (2011). Adolescents' implicit theories predict desire for vengeance after peer conflicts: Correlational and experimental evidence. *Developmental Psychology, 47*(4), 1090-1107.