Policy Research Working Paper 9141

The Power of Believing You Can Get Smarter

The Impact of a Growth-Mindset Intervention on Academic Achievement in Peru

Ingo Outes-León
Alan Sánchez
Renos Vakis

WORLD BANK GROUP
Poverty and Equity Global Practice
February 2020

Policy Research Working Paper 9141

Abstract

This paper evaluates the academic impact of a growth-mindset intervention on students starting the secondary level in public schools in urban Peru. ¡Expande tu Mente! is a 90-minute school session aimed at instilling the notion that a person´s own intelligence is malleable. Students in schools randomly assigned to treatment showed a small improvement in math test scores and educational expectations, with a large and sustained impact in test scores among students outside the capital city. At a cost of $0.20 per pupil, ¡Expande tu Mente! was highly cost-effective. The results show the potential that brief growth-mindset interventions have for developing countries.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at rvakis@worldbank.org.

The Power of Believing You Can Get Smarter: The Impact of a Growth-Mindset Intervention on Academic Achievement in Peru[1]

Ingo Outes-León[2]

Alan Sánchez[3]

Renos Vakis[4]

Keywords: education, disadvantaged students, growth mindset, metacognition, academic support, Peru

JEL codes: D91, I20

[1] A preliminary version of this document was circulated as a working paper in Spanish (Outes, Sánchez and Vakis, 2017). This project was implemented with the support of the Ministry of Education from Peru. We thank Gonzalo Manrique for providing excellent research assistance, and Nicolás Pazos and María Gracia Rodríguez for their support during the implementation of the project. We are also grateful to David Yeager for his advice at an early stage of the project, and to Isy Faingold, Fabiola Cáceres, Andrea Cornejo, Luis Baiocchi, Efraín Rodríguez and the MINEDULab team for their continuous support during the implementation stage.

[2] London School of Economics. Email: i.w.outes@gmail.com

[3] Corresponding author. Email: asanchez@grade.org.pe. *Grupo de Análisis para el Desarrollo*.

[4] The World Bank. Email: rvakis@worldbank.org

1. Introduction

In recent years, substantial evidence has emerged on the impact of both improving school inputs and creating incentives for students and teachers to improve school attendance and achievement in the context of developing countries (for a review, see Glewwe and Muralidharan, 2015). The goal of many of these interventions is to promote skill development, especially for disadvantaged groups in highly unequal societies. Comparatively, the role of socio-emotional interventions to help students succeed at school in low-and-middle-income countries remains less studied, yet this is a promising avenue given the existence of complementarities between cognitive and socio-emotional investments (Cunha and Heckman 2006, 2007).

We seek to generate evidence of a tool designed to improve the growth-mindset of adolescent students who are transitioning from primary to secondary. The intervention is based on the self-theories of intelligence (Bandura and Dweck, 1985; Dweck and Leggett 1988). Individuals tend to think of their own intelligence in two ways: some believe intelligence is something that is fixed (fixed mindset), whereas others believe intelligence is malleable and that it can be improved through effort (growth mindset). Self-theories of intelligence influence the way individuals respond to challenges: students with a growth mindset tend to be more interested in looking for situations in which they can learn and, eventually, improve their intelligence and academic achievement, whereas those with a fixed mindset are less likely to be interested in looking for situations in which they can learn (Dweck et al., 1995; Dweck, 2006). Throughout the late 1990s and early 2000s, studies from the field of psychology show that self-theories of intelligence of students can be modified through short-lived interventions and that this can have an impact on motivation and effort, and, eventually on academic achievement among students from disadvantaged groups.[5] See, for instance, Mueller and Dweck (1998), Aronson (1999), Aronson et al. (2002), Good et al. (2003), Blackwell et al. (2007). More recently, Yeager et al., (2019) show that a 50-minute, online, growth mindset intervention improved GPA scores by 11% of a standard deviation among low-achiever students in a random sample of schools in

[5] In the United States, disadvantaged groups account for students from public schools from poor neighborhoods and with a predominance of Afro-American and Latino students.

the United States[6] (see also Paunesku et al, 2015);[7] Bettinger et al. (2018) adapted the tools from the Yeager et al. study in a small sample in rural Norway, replicating their results.[8] Also, Alan et al. (2019) randomly assigned fourth-grade students to 12 two-hours sessions in Istanbul, Turkey, and show this led to persistent improvements in math test scores by 20% of a standard deviation.[9]

While encouraging, recent experimental evidence equally shows that these interventions can also be ineffective (Foliano et al., 2019; Ganimian, 2019). Furthermore, a recent meta-analysis that analyzes results from 29 studies shows mixed evidence of impacts on academic achievement (Sisk et al., 2018). To our knowledge, no study has demonstrated whether a growth mindset intervention can be successful in the context of low-and-middle-income countries.[10] With millions of students around the globe attending schools of insufficient quality

---

[6] Yeager et al. (2019) randomly assigned a 50-minute (2 sessions, 25 minutes each), online intervention in a sample of low-achiever students from the ninth-grade. Students come from a national sample of 65 regular public secondary schools in the United States.

[7] Mueller and Dweck (1998) show that students whose effort is praised improved their performance in math test scores by 30% compared to students whose intelligence is praised. Blackwell et al. (2007) administered 8-sessions of 25 minutes each (one per week) in which the science of the brain was explained to students, in particular how brain connections grow, and people become more intelligent when faced with challenges. As part of these sessions, students were asked to read a brief text that described the science behind the growth mindset (“You can grow your brain”). Subsequently, students were asked to think about examples in which their performance had improved through practice. Finally, they were asked to write a letter to younger students explaining the concepts learned –this was to help persuade them about the concepts learned through cognitive dissonance (Aronson, 1999). By the end of the academic year, an improvement in math test scores by 0.2 standard deviation was detected. Good et al. (2003) used a mentorship strategy (which included sessions for two months, and email exchanges with the mentor during the school year). They found positive effects on math test scores among women by around 0.2 standard deviation. Aronson et al. (2002) developed a similar intervention in which they worked with mentors that were in the last years of college to teach the growth mindset to first-year college students –and found an impact on final scores of 0.29 standard deviation. Paunesku et al. (2015) randomly assigned a 45-minute online session to students (the sample, 1,594 students in 13 schools, was larger than previous studies). During the session, students were asked to read an article about the growth mindset, and then to write letters to other students about the ideas learned. By the end of the semester, students from the intervention group were 6.4 p.p. more likely to have a satisfactory score in their main course.

[8] Bettinger el al. (2018) randomly assigned a 135-minute, online intervention (3 sessions, 45 minutes each) among 345 students in 1 rural school in Norway.

[9] Alan et al. (2019) randomly assigned the intervention at the school level in a sample of 16 schools in Turkey (Istanbul).

[10] Ganimian (2019) randomly assigned a short-lived intervention at the school level in a sample of 202 urban, public secondary schools in Argentina (province of Salta), with a focus on

in the developing world, the potential of highly low-cost socio-emotional interventions stands as highly policy relevant. We aim to fill this gap.

We focus on Peru, a middle-income country that in recent years benefited from sustained economic growth along with substantial poverty reduction (monetary poverty reduced from 54% to 21% to between 2000 and 2015 according to official statistics), but where schooling achievement remains very low and highly unequal –by 2018, only 14% and 16% of students in the second grade of secondary level (equivalent to the eighth-grade) achieve a satisfactory performance in *Mathematics* and *Reading Comprehension* tests, respectively.[11] Within this context, we evaluate the impact of a growth-mindset intervention in Peru. The intervention, called '*¡Expande tu Mente!*', adapted existing tools (Yeager et al, 2016 and 2019; Paunesku et al., 2015) to a context in which online interventions are not feasible. The intervention, implemented with support from the Ministry of Education, consists of a 90-minute school session led by tutor teachers. To measure the impact of the sessions, 800 public schools were sampled in three regions in Peru. Two equally powered samples were created: a *metropolitan* sample and a *regional* sample. Within each sample, half of the institutions were randomly selected to participate in '*¡Expande tu Mente!*'.

Packages were sent by courier and implementation monitored by phone. Tutor teachers from grades seven and eight (the first two years of the secondary level) were asked to deliver the sessions to all their students. Approximately 60% of all eligible schools delivered the sessions. We used information from the students' census evaluation from 2015 and 2016 to test the academic impact of '*¡Expande tu Mente!*' sessions approximately 2 and 14 months after the intervention took place (respectively). Results after 2 months show that being assigned to the sessions improved *Mathematics* test scores by 5% of a standard deviation (the point estimate for *Reading Comprehension* is similar but statistically insignificant). We also uncover substantial geographical heterogeneity: the main result for *Mathematics* is explained by an ITT impact of 13% in the *regional* sample and virtually a zero impact in the *metropolitan* sample. In the *regional* sample, we also detect an ITT impact of 9% in *Reading Comprehension* test

students in the 12th grade (approximately 9,000 students). The duration of the Ganimian intervention is not reported, but it is based on the adapted 90-minute version reported in Outes et al. (2016). Interestingly, the author found no impact of the intervention on academic achievement, nor indeed on other dimensions.

[11] http://umc.minedu.gob.pe/wp-content/uploads/2019/04/presentacion-web-ECE2018-1.pdf

scores. The results of the intervention are persistent. In the *regional* sample, 14 months after the intervention those students from schools assigned to treatment that took the census evaluation improved their scores by 10%, 12% and 10% in *Mathematics*, *Reading Comprehension*, and *History, Economics and Geography* (compared to the control group). Our results shed light on the potential of growth-mindset interventions. Given its per capita ITT cost, this puts the '*¡Expande tu Mente!*' package among the most highly cost-effective educational interventions in the developing world. At the same time, our results highlight that while growth-mindset interventions work, they do not work in all contexts. In *metropolitan* areas, where schools and classrooms are larger and students have higher scores, our intervention failed to be effective.

The rest of the paper is organized as follows. Sections 2 and 3 describe the intervention and its conceptual pathways (respectively). In section 4, we describe the experimental design and the implementation of the intervention in the field. Sections 5 and 6 present the data and the empirical strategy, and sections 7 and 8 report our main results and robustness checks. Section 8 concludes.

2. The intervention

The intervention, known as '*¡Expande tu Mente!*' in Peru ('Grow Your Mind!'; from here onwards, GYM)) was a collaboration between the authors and the Peru Ministry of Education (MINEDU). Within MINEDU, the research team worked in coordination with *MineduLAB*, the cost-effectiveness laboratory of the Ministry, and the Direction of Secondary Education. The main objective of the intervention was to instill in students the growth-mindset to improve their motivation, perseverance, effort at school, and, ultimately, academic performance (see the theory of change in **Section 3**). The basic education level in Peru is composed of a pre-school level, six grades at the primary level and five grades at the secondary level. The intervention was designed for the early grades of the secondary level, specifically for those in the first and second grade of secondary (equivalent to the seventh and eighth grades, respectively), a transitional period that students often find challenging and which is the beginning of the path to dropping-out from school.

The key component of the intervention is a 90-minute school session designed to introduce students to neuroplasticity and the concept that cognitive abilities are malleable. It was delivered during school hours by local teachers, specifically by the tutor in charge of each classroom.[12] The content of the session is based on previous work by Yeager et al. (2016, 2019), Paunesku et al. (2015) and Blackwell et al. (2007).[13] The session comprises three distinct segments of approximately 30 minutes. In the first segment, students were asked to individually read a text designed to introduce students to the concept of growth-mindset (the 'GYM text'). This was an adaptation of the seminal intervention text in (Blackwell et al., 2007), which explains that the 'brain works like a muscle' and that with effort and practice anyone can improve their intelligence (see **Annex 1**). In the second part of the session, the content of the text is debated by groups of 4 to 5 students (the composition of each group was freely decided by each teacher) and then by the classroom, with the teacher serving as a moderator of the debate. In the last segment of 30 minutes|, students were asked to write individually a reflective letter to a friend/relative describing what they learned.[14]

In a follow-up session, teachers were instructed to carry out two additional activities. First, to choose the best letters (according to their subjective assessment) and to hang all their student letters on the wall, alongside a poster provided by the program and which summarizes the key concepts of the session (see **Annex 1**). Teachers were asked to leave the letters and the poster on the classroom walls for the remainder of the school year. They were to act as a physical reminder of the sessions and increase the long-term salience of their teachings. Second, teachers were asked to take a picture of the students alongside the letters and the poster and to share this picture by email.

From the outset, the intervention was designed to be highly replicable and scalable. All intervention materials reached schools by parcel delivery (the 'GYM package'), and no ad-hoc

[12] In schools in Peru, tutorial sessions take place weekly over the entire school year. The primary aim of these sessions is to provide a space to help students with their emotional and academic development, and to promote a strong relationship and better emotional understanding between students and teachers. While there are guidelines from MINEDU about the content of these sessions, ultimately it is up to each school and each teacher to define the topics to be covered.

[13] Intervention tools were translated into Spanish and adapted to the cultural and curricular idiosyncrasies of the Peruvian context. Instructions for tutor teachers were further developed by the authors. Before the intervention, tools were validated in focus group discussions with local teachers and pupils at GRADE.

[14] The importance of reflective writing has been shown in a multiplicity of psychosocial studies (see, for instance, Countryman, 1992), including growth-mindset interventions (e.g., Paunesku et al., 2015).

school visit or teacher training was required. The GYM package contained 6 elements: (1) a letter for the head teacher signed by the Director of Secondary of MINEDU, explaining the nature of the activity and requesting the school to participate; (2) individual letters to all tutor teachers within the school (who would be in charge of running the sessions); (3) detailed instructions about the structure of the session and activities to be carried out; (4) an annex with additional information about the concepts of growth mindset and examples about how to motivate the students; (5) copies of the GYM text for all students expected to receive the session; and (6) 1 poster for all classrooms to be taking part of the intervention.

Head teachers and tutors were expected to manage independently the implementation process within each school. The remote nature of the intervention raises questions of implementation fidelity and compliance. While aware of these concerns, the design was deliberate. We aim to avoid the pitfalls of small-scale evaluations that do not replicate well when scaled up (Muralidharan and Niehaus, 2017). Strategies to minimize incompliance include a tracking verification system by phone and the request of providing photographic evidence of completed sessions (see **Section 4** for a more detailed explanation).

3. Conceptual framework

Figure 1 exemplifies the theory of change from the intervention. Students are exposed to the concept that, with effort and practice, anyone can improve their intelligence. By appropriating this concept, we expect students with a fixed-mindset to modify their mental model towards a growth-mindset. Because of this new knowledge, students might increase their perseverance and motivation to learn, and adjust their aspirations and expectations, which eventually leads to improvements in the academic effort and learning outcomes. Furthermore, it is expected that there will be a feedback process by which improvements in learning outcomes lead to further improvements in perseverance, motivation, aspirations, and efforts exerted. The intervention is expected to influence the student, however by design the tutor teachers must first expose themselves and learn about the growth mindset. Therefore, through the process of studying the material to implement the GYM session, we expect that teachers either change and/or receive confirmation of their own priors, which could influence the way they interact with academically weak students.

The impact of the intervention on students' skills can be illustrated through a production technology that links school inputs to the formation of cognitive skills and socio-emotional skills (including aspects such as motivation and perseverance) over multiple periods. Specifically, consider the skills formation model (Cunha and Heckman, 2008), enhanced to include the role of school inputs –as in a standard education production function, see Todd and Wolpin (2007). Cognitive and socio-emotional skills in period t (represented by vector S_t^C and S_t^N) depend on skills accumulated up to period $t-1$ (S_{t-1}^C and S_{t-1}^N) and investments in both types of skills in period t. For simplicity, we distinguish between investments at the school level ($I_t^{j,School}$) and other investments ($I_t^{j,Others}$) for $j = (C, N)$). The skills formation process is depicted as follows,

$$S_t^C = f^c\left(S_{t-1}^C, S_{t-1}^N, I_t^{C,School}, I_t^{C,Others}, \mu^C\right)$$
$$S_t^N = f^N\left(S_{t-1}^C, S_{t-1}^N, I_t^{N,School}, I_t^{N,Others}, \mu^N\right)$$

where μ^j represent initial endowments. We conceptualize the intervention as an exogenous increase in school-level investments in socio-emotional skills in period t ($I_t^{N,School}$). Consider a three-period horizon ($t-1$, t and $t+1$). In this case, both a contemporaneous (direct) impact on S_t^N, and a subsequent (indirect) impact on S_{t+1}^C and S_{t+1}^N are expected, the latter occurs due to cross-productivity (impact of socio-emotional skills on cognitive skills) and realignments on other investments (e.g., more hours of study at home). Furthermore, the intervention is likely to modify the way the teacher interacts with students (e.g. changes in pedagogical strategies), which can have an additional direct impact on school-level investments, generating additional improvements in skills accumulation in periods t and $t+1$.

4. Empirical design

(a) Experimental Design

The experiment took place in 2015. Three adjacent regions (of 24 in the country) were selected in coordination with MINEDU: Lima, Junín, and Ancash. These regions capture some of the geographic diversity of Peru –coast and highlands– while their relative accessibility to Lima city (from which parcels were sent) ensured reliable parcel delivery. The study focused on public schools and in urban areas, the latter to reduce logistical complexity of parcel delivery—in the Lima region population is predominantly urban, whereas in Junín and Ancash approximately 2 out of 3 households are located in urban areas. Three types of schools were excluded from the sampling frame: (a) those located in districts in the top quintile of wealth; (b) those that operated night-shifts only; and, (c) those with fewer than 15 pupils in total. These restrictions resulted in a universe of 946 schools in the three regions. To maximize statistical power, it was decided to sample 800 from these 946 schools.

In total, 400 of the 800 schools were randomly selected to receive the 'GYM package'. The study contains two equally powered samples: the *metropolitan* samples with 398 schools from the Lima region, a densely populated area with over 12 million inhabitants, and the *regional* sample which includes 402 urban and peri-urban schools in the regional towns of Ancash and Junín. To ensure balance in the experimental sample and maximize statistical power, treatment status was stratified across poverty quintiles and regional strata (Bruhn and McKenzie, 2008), making 12 strata in total.[15] **Table 1** provides an overview of treatment and control schools, across the pooled and individual experiments.

Treatment schools were targeted to receive the 'GYM package' and to implement the Grow Your Mind session during regular tutorial hours in 2015. In each institution that received the package, tutor teachers were asked to implement the session in every classroom for grades seven and eight. Control schools were not approached, and, thus, continued using the national curriculum in their tutorial sessions. The experiment was designed to assess the impact of the 'GYM package' using information from the students' census evaluation (*Evaluación Censal de Estudiantes*, ECE), administered a few months after the intervention for those in the eighth grade, and more than a year later for those in the seventh grade (see **Section 4e**).

[15] Pre-intervention power calculations of the school-clustered randomized control trial suggested a low minimum detectable effective size (MDES) of 12% for the pooled sample, while individual experiments were powered to detect an MDES of at least 17%. To compute MDES in this level-3 cluster RCT, we assumed 2 sections per school with 25 pupils per section, an intra-class correlation at the school-level of 0.25 and within-school correlation across teachers of 0.15. In reality, pre-treatment power calculations were very conservative: not only average schools turned out to be larger, but the intra-class correlation across schools was closer to 0.15. Ex-post calculation based on the 2015 ECE national data suggests that the pooled and individual studies were powered to detect 8% and 12% MDES respectively.

(b) Implementation

Packages were prepared in July 2015, at GRADE premises (in Lima). Each 'GYM package' was personalized for each school based on the number of classrooms in grades seven and eight, as reported in the 2014 Peru School Census.[16] Once packages were ready, a private courier company was hired to deliver them to all schools in the treated group. Each package contained an official letter from MINEDU (signed by the Director of the Secondary Level) addressing the head teacher of the school. This letter was placed on the outside of the package so that it was visible to the person who received it.

To locate schools, the courier was given the name of the institution, full address and (when available) contact phone numbers, as reported in the publicly available information system of MINEDU called ESCALE. The initial aim was to deliver all packages during August 2015. However, there were problems with some of the deliveries because the information given to the courier company was not detailed enough and/or was not accurate. For this reason, for a group of schools the courier company was given a physical map to locate the district (elaborated with Google Maps, using the geo-referenced location from the school) and the most up to date information about phone numbers (which was not necessarily information that was publicly available at that time). This information was obtained from the data division of the Ministry (*Sistema de Información de Apoyo a la Gestión de la Institución Educativa*, SIAGIE). Based on this strategy, the delivery of packages was completed between August and September 2015. As part of the protocol, each school that received the package signed a delivery receipt.

(c) Monitoring delivery and compliance

A small tracking team, composed of two individuals with experience in telemarketing, was trained to monitor–through phone calls–all schools that had received the 'GYM package' according to the courier. The tracking exercise had multiple purposes. First, to make sure that the 'GYM package' had been received by tutor teachers in each school. Second, to ask if there were any doubts about the nature of the sessions and/or if the material they received was incomplete. Third, to monitor when sessions were going to take place and/or to ask why

[16] If the number of sections was uneven, we rounded it to the next even number (e.g., 4 packages were sent to schools with 3 sections). We assumed each section had 30 students.

sessions were not going to take place. Fourth, to remind schools that tutor teachers had to share by email at least one picture per classroom (with the students standing up next to the letters prepared by them, and next to the GYM poster). For all these reasons (monitoring delivery and monitoring compliance of the sessions) the tracking team was in constant contact with both head teachers and tutor teachers from September and up to December 2015 (the end of the school year). This tracking exercise was complemented with massive email remainders.

The results of the delivery process and compliance are reported in **Table 2**. From the 400 schools assigned to treatment, 340 of them received the 'GYM package' (85%; 79% in the *metropolitan* sample, 91% in the *regional* sample). Those cases where the package did not reach its destination include very remote areas and/or cases in which the courier was not able to locate the school. In addition, there were a handful of cases in which the school did not want to receive the package. Considering all eligible schools (**Table 2**, Panel A), 60% declared having administered at least 1 'GYM session', whereas 55% provided physical evidence (i.e., a picture) of at least 1 'GYM session'. In our analysis, we define a school as a complier if there is evidence (i.e., pictures) that GYM sessions were administered in all classrooms at a given grade. Following that definition, the compliance rate was 45% in the eighth grade and 43% in the seventh grade. A larger compliance rate is observed in the *regional* sample, however, this difference is driven by differences in the package delivery success rate; when looking at the compliance rates among those that received the package (**Table 2**, panel B), the difference (in percentage points) reduces substantially. When sessions did not take place, the tracking team was asked to inquire about the reason. The most common reason was lack of time due to other activities.

(d) <u>Cost</u>

Efforts to develop a scalable intervention resulted in a highly inexpensive tool. Typical school-level interventions incur substantial costs in teacher training and school visits; the 'GYM package' requires neither of them. As a result, the unit cost of "°¡Expande tu mente!" was $55 per treatment school or $0.20 per pupil (on average). Intervention costs were split between (a) courier services for parcel delivery (30%), (b) telephone operators for monitoring (20%) and (c) printing and packing of intervention materials (50%).

5. Data

(a) Students' census evaluation

The experiment was designed to assess the impact of the 'GYM package' using results from the annual students' census evaluation (*Evaluación Censal de Estudiantes*, ECE) administered by MINEDU. The strategy of asking teachers from grades seven and eight to implement the 'GYM session' generated two separate evaluation cohorts: the eighth-grade cohort assessed in the ECE 2015, and the seventh-grade cohort assessed in the ECE 2016. First introduced at the secondary education level in 2015, the ECE administers *Mathematics* and *Reading Comprehension* tests to all students enrolled in grade eight.[17] In 2016, the evaluation included a *History, Geography and Economics* test. These are our key outcomes measures.

The ECE in 2015 and 2016 were administered in mid-November. All pupils were requested to sit their exams on the same dates throughout the entire Peruvian geography. Test content and marking were carried out centrally by MINEDU, while the administration of the tests was under the supervision of a bespoke team of enumerators. The ECE data set provides individual test scores for all participating students. Based on these scores, students are classified into four groups according to their level of learning: 'Before beginning', 'Beginning', 'In process' and 'Satisfactory'. Only those in the 'Satisfactory' group have achieved a learning level consistent with their current grade. In 2015, this accounted for 9.5% in *Mathematics* and 14.7% in *Reading Comprehension* for all the student populations in the eighth grade.

Apart from test scores, the ECE also collects pupil background data via a self-administered questionnaire, which includes standard modules on household, parental and student socio-economic characteristics, one question on education expectations (¿What is the highest education level you expect to achieve?) and, in the ECE 2015 only, two Likert scales to measure students' perceptions about their ability to understand and complete different tasks related to math and reading comprehension (respectively). We test the impact of the intervention on education expectations and students' self-perception of their own abilities in these two subjects to improve our understanding of the mechanisms.

[17] Two other ECE studies take place annually in Peru, a census assessment of students in Year 2 and a sample assessment for Year 4s. For our study, only data from the Year 8 ECE is relevant.

(b) Descriptive statistics

In **Table 3** we report school characteristics by geographical sample. The information corresponds to the 2014 School Census (Panel A to D) and ECE 2015 (Panel E). Differences between both samples (*metropolitan* versus *regional*) are substantial. Schools in the *metropolitan* sample are larger –with more classrooms, and more students per classroom– more likely to have an afternoon shift, and in most cases have better physical infrastructure (access to sewage, existence of a science laboratory) than those in the *regional* sample; they are also less likely to have bilingual schools and to be in districts where the main Peruvian social programs operate. Concerning students' achievement in these schools, attainment is higher in the *metropolitan* sample, but grade attainment is low in general; in *Mathematics* and *Reading Comprehension*, only 6% and 11% attain a satisfactory level in each subject, respectively.[18]

Based on this information, we also report results from the balancing tests (see **Table A1** in the Web Appendix). Treatment and control schools appear to be similar in most observed dimensions. Some differences are observed (by day shift and sex), however, these differences are not substantial in magnitude. Unfortunately, no ECE was available before treatment to confirm balancing in the outcome variables, as 2015 was the first year in which ECE was administered in schools at the secondary level.

Figure 2 presents kernel distributions for *Mathematics* and *Reading Comprehension* standardized test scores in 2015, separately for treatment and control schools, **Figure 3** presents analogous information for 2016, including also scores from *History, Geography, and Economics*. A clear pattern arises whereby improvements are observed in all areas and both years in the *regional* sample. In contrast, no such improvements are observed in the *metropolitan* sample. This graphical analysis strongly mirrors the parametric analysis presented later in the paper.

[18] Results correspond to the control group.

6. Evaluation strategy

To empirically assess the magnitude of the impact of the 'GYM package', we apply the following Intent-to-Treat (ITT) estimation model,

$$Y_{ijs} = \alpha + \beta GYM_j + \mu_s + \varepsilon_{ijs} \qquad (1)$$

where Y_{ijs} is an outcome variable for pupil i, in school j and randomization strata s; GYM_j is a variable that takes the value of 1 if j was assigned to treatment (i.e., to receive the 'GYM package') status, 0 otherwise. Thus, the coefficient β provides an estimate of the intent-to-treat treatment effect. We also include a full set of strata dummies, μ_s, to account for the study design. When drawing inference, here and onwards we correct standard errors for school-level clustering.[19]

Under imperfect compliance, the ITT comparison underestimates the average treatment on the treated (ATT). Using our monitoring data, we classify schools according to whether all GYM sessions were implemented before the ECE— $Treated_j$ is a binary variable that takes the value of 1 in that case and 0 otherwise. In Equation (2), we retrieve ATT estimates by instrumenting $Treated_j$ with the original assignment status, GYM_j,

$$Y_{ijs} = \alpha + \beta_{LATE} \widehat{Treated}_j + \mu_s + \varepsilon_{ijs} \qquad (2)$$

Given the random nature of the instrumental variable, with this specification, we can obtain the ATT for the sub-sample of complying schools. There are many alternative definitions for compliance in our study, ranging from whether schools received the parcel to whether objective evidence for session implementation is available. We focus on the most demanding definition: $Treated_j = 1$ if 'all sections from a certain grade in school j (7th and 8th grade, depending on whether the focus in on medium-term or short-term estimates, respectively) provided photographic evidence of implementation'.

---

[19] Not only is it appropriate, with a random assignment at the school level, but in the knowledge that errors are strongly correlated within schools, failure to account for this source of clustered correlation could invalidate our inferencing tests. With 799 schools in our analysis sample, we can be confident that cluster corrections will not suffer from finite sample biases.

7. Results

Short-term impacts

Results from the impact of the intervention after two months of implementation are reported in **Table 4**, for the full sample and the *metropolitan* and *regional* samples. In this and all other tables from here onwards, test scores are standardized to the mean and standard deviation of the control group. Results from ITT estimates (Equation (1)) are reported in **Panel (A)**. Point estimates suggest that the 'GYM package' had an ITT impact both on *Mathematics* and *Reading Comprehension*, however only the former is significant at standard levels. Specifically, students in schools assigned to receive the 'GYM package' increased their average test score in *Mathematics* by 5% of a standard deviation (the point estimate for *Reading Comprehension* is of a similar magnitude, 4%). Differences in the impact of the intervention between the two geographical samples are striking. Students in the *metropolitan* sample appear to have not benefited from the intervention with point estimates close to zero. In contrast, benefits in the *regional* sample are substantial: students from schools assigned to receive the 'GYM package' in this sample increased their test scores in *Mathematics* and *Reading Comprehension* by 13% and 9% of a standard deviation, respectively.

Under imperfect compliance, ITT estimates provide a lower bound of impact. LATE estimates confirm this. We use 'schools that provided picture-verification for all classes in eighth-grade' as the LATE definition of treatment, and assignment status as the instrumental variable. These results are reported in **Panel B**. In the full sample, the GYM intervention had an impact of 15% in *Mathematics* test scores –the point estimate for *Reading Comprehension* was 11%, also statistically insignificant. The impact observed in the full sample is driven by the impact of the 'GYM package' among complier schools in the *regional* sample, with improvements in these subjects by 34% and 22% (respectively). In the Web Appendix, we replicate ITT and LATE estimates using school-level aggregated outcome measures (see **Table A2**). This specification provides equivalent findings, but with larger standardized impacts.

In **Table 5**, we report results transforming the test scores for *Mathematics* and *Reading Comprehension* into binary variables according to whether students' results are considered to be at the 'Before beginning' level (the lowest level), 'Beginning', 'In process' or 'Satisfactory' level (the highest level), following the standards of the school curriculum in Peru. For

Mathematics (Panel A), improvements are observed across the whole distribution of grades. Specifically, we observe both reductions in the proportion of students in the lowest grade category ('Before beginning') and improvements in the proportion in the 'In process' and 'Satisfactory' categories. As before, results are driven by the *regional* sample. Point estimates obtained for the 'Satisfactory' category (ITT coefficients of 1 p.p. in the full sample and 2.3 p.p. in the regional sample) represent improvement by 17% and 38%, respectively. Similar results are observed for *Reading Comprehension* (Panel B) but in this case, point estimates are smaller and, in some cases, marginally insignificant.

Overall, our estimates show strong evidence of the substantial impact of the intervention in the *regional* sample, but not in the *metropolitan* sample. It is difficult to single out a specific reason that might drive these results. One possible explanation is that "*¡Expande tu Mente!*" is more likely to succeed among students with low-achievement—as evidence from other countries has also shown. At the same time, schools in the *metropolitan* sample are larger and better equipped. It is likely that that the GYM intervention might have been more salient in a regional context, where teachers have fewer teaching tools at their disposal. Besides, the fact that the average classroom size is lower in the *regional* sample might have contributed to the success of the GYM sessions—everything else equal, a session that requires all students to read an article is more likely to be effectively delivered in a smaller classroom.

Robustness checks and heterogeneous estimates

There might be concerns that remote schools in the treatment group were more likely not to receive the 'GYM package', thereby potentially generating sample selectivity. To deal with this, we report results excluding remote schools from the control group. We do this by excluding schools from the control group located in areas the courier was not able to reach. These results largely resemble those obtained from our main sample (see **Table A3** in the Web Appendix). Also, our core analysis results remain robust to controlling for school characteristics that are different between schools assigned to treatment and control.[20]

[20] Though the stratified randomization strategy was designed to minimize systematic imbalances between the treated and control samples, this possibility could not be ruled out. A key balancing test would have been to assess whether school attainment between treatment and control schools differed before treatment. Unfortunately, no school assessment data were available before 2015. Instead, we carried out a balancing test using administrative school-level data obtained from the 2014 Annual School Census. Treatment and control schools do

We also test for heterogeneous results by gender and ethnicity. Students in the full sample are evenly distributed between males and females, and a minority (6%) has a language different than Spanish as the native tongue (11% in the *regional* sample). Spanish native speakers have substantially higher test scores than non-Spanish native speakers, whereas males have higher scores in mathematics and females in reading comprehension—in all cases mirroring patterns observed at the national level. In **Table 6**, we report estimates of the impact of the intervention by gender and ethnicity (*proxied* by the native tongue). We found no evidence of differential impacts between males and females, and, similarly, no evidence of differential impacts between Spanish and non-Spanish native speakers at standard levels of significance. Looking at the point estimates by the native tongue, point estimates are substantially larger for Spanish native speakers in the *regional* sample. At face value, this result might come as a surprise since other studies have found that growth mindset interventions tend to favor minorities that might be afflicted by racial stereotypes. We speculate that one reason that might explain why the impact of the intervention could be larger for Spanish native speakers is that a minimum level of reading comprehension is required to read and understand the Grow Your Mind article that is part of the intervention.

<u>*Medium-term impacts*</u>

Previous results reflect the short-term impact of the GYM intervention, measured approximately two months after the GYM session took place. While the impact on that cohort is only observed once, our design allows us to measure the impact 14 months after the intervention, on those students who were in the seventh grade in 2015 and eighth grade in 2016. These results are reported in **Table 7**. Point estimates for the full sample are like those observed the previous year: the impact of the GYM intervention is persistent in the *regional* sample, with ITT impacts by 10%, 12% and 10% of a standard deviation in *Mathematics*, *Reading*

not differ in most tested characteristics. Balancing tests failed in only two covariates: whether the school segregated by gender; and, whether the school runs only morning shift or morning and afternoon shifts. The imbalanced variables could be driven by outliers as they represent very small distributions. For the robustness check we included the imbalanced school characteristics as controls, finding virtually the same results. Results are available upon request.

Comprehension, and *History, Economics, and Geography*, respectively. Consistent with previous findings, no impact is observed in the *metropolitan* sample.[21]

<u>*Mechanisms*</u>

The administrative data available do not allow us to fully uncover the mechanisms through which the intervention had an impact on students' test scores, however, there is some information available on the self-administered questionnaire that students assessed by MINEDU have to answer. In **Table 8** we provide evidence of the impact of the intervention on educational expectations (in the short and medium term), as well as on students' self-perception of their own abilities in math and reading comprehension (in the short-term). In the short-term, students in schools assigned to receive the 'GYM package' improved their expectation to attend university by 1.1 p.p. Consistent with previous findings, the ITT impact on expectations is virtually zero in the *metropolitan* sample, and approximately 1.8 p.p. in the *regional* sample. Results are very similar in the medium-term estimates, albeit marginally insignificant. Finally, our results do not suggest the intervention had an impact on students' beliefs about their own abilities in the short-term. It is possible that the intervention might have had an impact on this domain in the medium-term, once students realized they had improved their performance on both subjects—potentially thanks to an increased academic effort. Unfortunately, this hypothesis cannot be tested because the scales used to measure these self-perceptions were discontinued from the MINEDU questionnaire.

[21] By 2016, the research team knew the intervention had had an impact in the *regional* sample and not in the *metropolitan* sample. Because of this, we decided to randomly allocate half the schools from the *regional* sample to receive the 'GYM package' so that they could repeat the session at approximately the same month of the year. The instruments and protocols were identical, and the same monitoring procedure was put in place. The only difference was that, in this case, the letter to the head teacher and tutor teachers gave emphasis to the idea that the 'GYM session' should be repeated. Our analysis shows no additional impact from being assigned to repeat the 'GYM session'.

8. Conclusions and further agenda

Peru is a middle-income country with high educational inequalities and where less than 2 out of 10 students from the secondary-level achieve an academic performance consistent with their school grade. We evaluated the impact of '*¡Expande tu Mente!*', a growth-mindset intervention, on standardized test scores among secondary-level students from public, urban schools in Peru. The study randomized the introduction of an in-class, 90-minute session delivered by school teachers. Our analysis shows that even when compliance was not perfect, the benefits of the intervention were relevant and long-lasting in the Peruvian context, especially in the regional context. At a cost of less than $0.20 per pupil, this intervention holds the promise of a highly cost-effective, scalable and replicable educational tool. At the same time, the fact that we were able to detect gains in the regional sample but not in the metropolitan sample shows that the way the tool is introduced (design of the intervention) and the context in which this occurs (i.e., school and teacher characteristics) both matter to understand potential gains. While the results obtained in this study are encouraging, more still needs to be learned about specific ways through which Grow Your Mind interventions can be successful in different contexts (e.g., in metropolitan areas and rural areas in Peru; in other developing countries), for different students (e.g., article versus video versus animated story version), and how this type of psychosocial intervention can complement existing interventions. To the best of our knowledge, our study is the first to successfully test the impact of growth-mindset psychosocial stimulation tools in a developing country. Moreover, few have done so at the scale reported in this study.

References

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

Aronson, E. (1999). The power of self-persuasion. *American Psychologist,* 54, 875-884.

Aronson, J., C. Fried y C. Good (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38, 113-125.

Bandura, M. y C. Dweck (1985). The relationship of conceptions of intelligence and achievement goals to achievement-related cognition, affect and behavior. Unpublished manuscript, Harvard University.

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

Blackwell, L. A., K. H. Trzesniewski y C. S. Dweck (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. *Child Development*, 78, 246-263.

Bruhn, M. y D. McKenzie (2008). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4), 200-232.

Cohen, G. L., J. García, N. Apfel y A. Master (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, *313*, 1307-1310.

Cohen, G. L., J. García, V. Purdie-Vaugns, N. Apfel y P. Brzustoski (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, 324, 400-403.

Countryman, J. (1992). Writing to learn mathematics: Methods that work, Portsmouth: Heinemann.

Cunha, F. y J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, 43 (4), 738-782.

Dweck, C., y Leggett, E. (1988). A Social-Cognitive Approach to Motivation and Personality. *Psychological Review*, 95(2), 256-273.

Dweck, C., Chiu, C., y Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological inquiry*, 6 (4), 267-285.

Dweck, C. (2006). Mindset: The new psychology of success. Random House Incorporated.

Evans, D. y A. Popova (2015). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. World Bank Policy Research Working Paper 7203.

Ganimian, A. (2019). Growth mindset interventions at scale: experimental evidence from Argentina. Mimeo.

Glewwe, P. y K. Muralidharan (2015). Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. RISE Working paper.

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

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

Hulleman, C. S. y J. M. Harackiewicz (2009). Making education relevant: Increasing interest and performance in high school science classes. *Science*, *326*, 1410-1412.

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, 26, 629-640.

Miyake, A., N. D. Kost-SmothFinkelstein, S. J. Pollock, G. L. Cohen y A. Ito (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330, 1234-1237.

Mueller, C. y C. Dweck (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology*, 75 (1), 33-52.

Outes, I., Sánchez, A., Vakis, R. (2017). Cambiando la mentalidad de los estudiantes: evaluación de impacto de ¡Expande tu mente! sobre el rendimiento académico en 3 regiones del Perú. Documento de Investigación 83. Lima: GRADE.

Oyserman, D., D. Bybee y K. Terry (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, 91, 188-204.

Paunesku, D., G. M. Walton, C. L. Romero, E. N. Smith, D. S. Yeager y C. S. Dweck (2015). Mindset interventions are a scalable treatment for academic underachievement. *Psychological Science,* 26(6), 784-793.

Shnabel, N., V. Purdie-Vaughns, J. E. Cook, J. García y G. L. Cohen (2013). Demystifying values-affirmation interventions: writing about social belonging is a key to buffering against identity threat. Personality and Social Psychology Bulletin, 39(5), 663-676.

Sisk, V., Burgoyne, A., Sun, J., Butler, J., Macnamara, B. (2018) To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science,* 29(4), 549-571.

Walton, G. M. y G. L. Cohen (2011). A brief social-belonging intervention improves academic and health outcomes among minority students. *Science,* 331, 1447-1451.

Walton, G. M. y G. L. Cohen (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92, 82-96.

Wilson, T. D. y P. W. Linville (1982). Improving the academic performance of college freshmen: Attribution therapy revisited. *Journal of Personality and Social Psychology*, *42*, 367-376.

Wilson, T. D. y P. W. Linville (1985). Improving the performance of college freshmen with attributional techniques. *Journal of Personality and Social Psychology*, 49, 287-293.

Yeager, D., Hanselman, P., Walton, G., Murray, J., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C., Hinojosa, C., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Man Yang, S., Carvalho, C., Hahn, R., Gopalan, M., Mhatre, P., Ferguson, R., Duckworth, A., Dweck, C. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573.

Yeager, D. S., C. Romero, C. S. Hulleman, B. Schneider, C. Hinojosa, K. Flint y D. Greene (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., R. Johnson, B. Spitzer, K. Trzesniewski, J. Powers y C. S. Dweck (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. y G. Walton (2011). Social-Psychological Interventions in Education. *Review of Educational Research*, 81(2), 267-301.

Figure 1: description of the '¡Expande tu mente!' theory of change

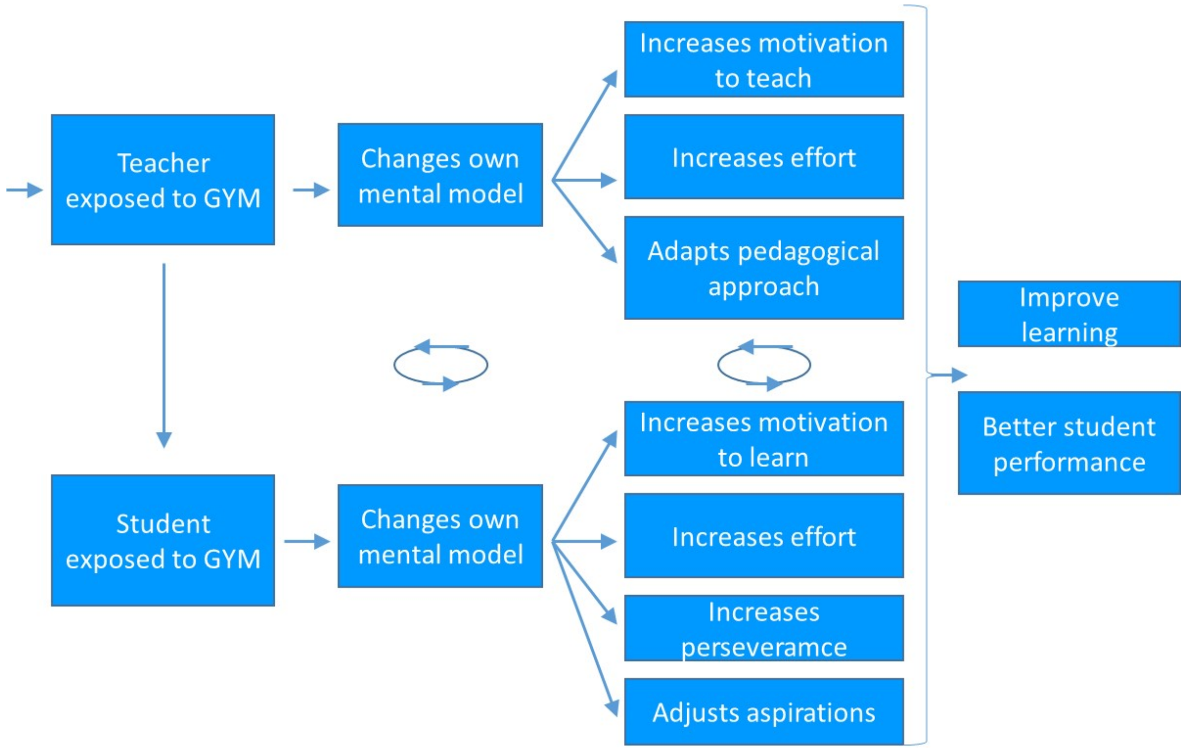
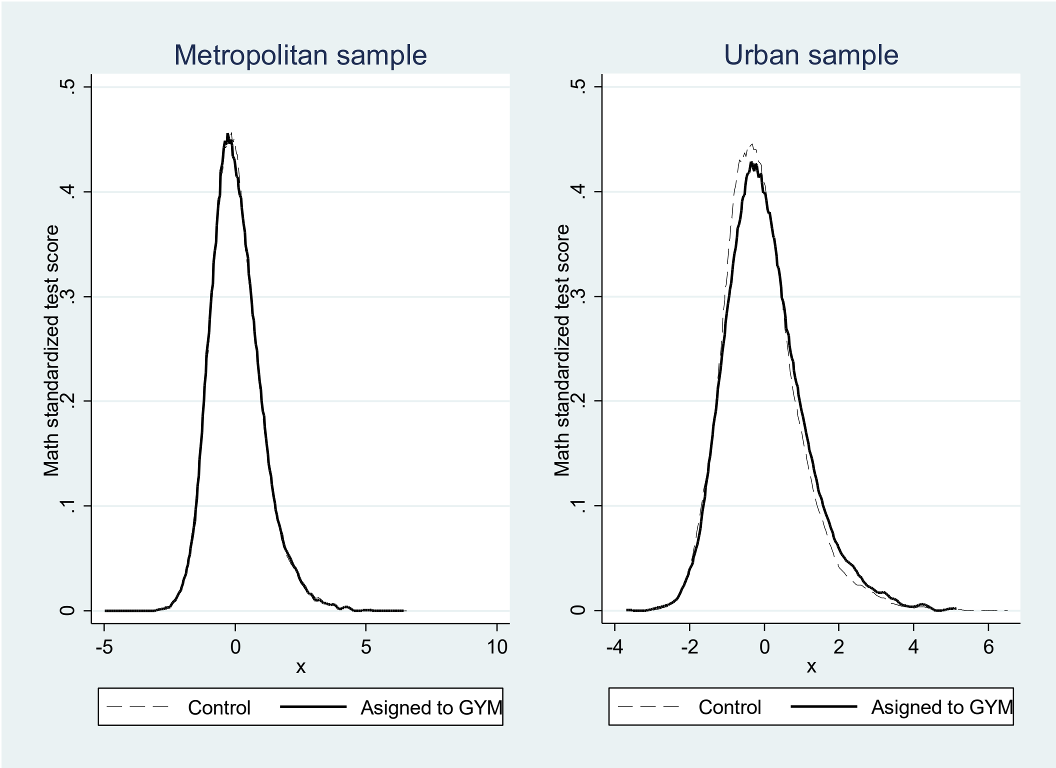


Figure 2: kernel densities of test scores in 2015 (short-term effects)

Part 2.a. Mathematics



Part 2.b. Reading comprehension

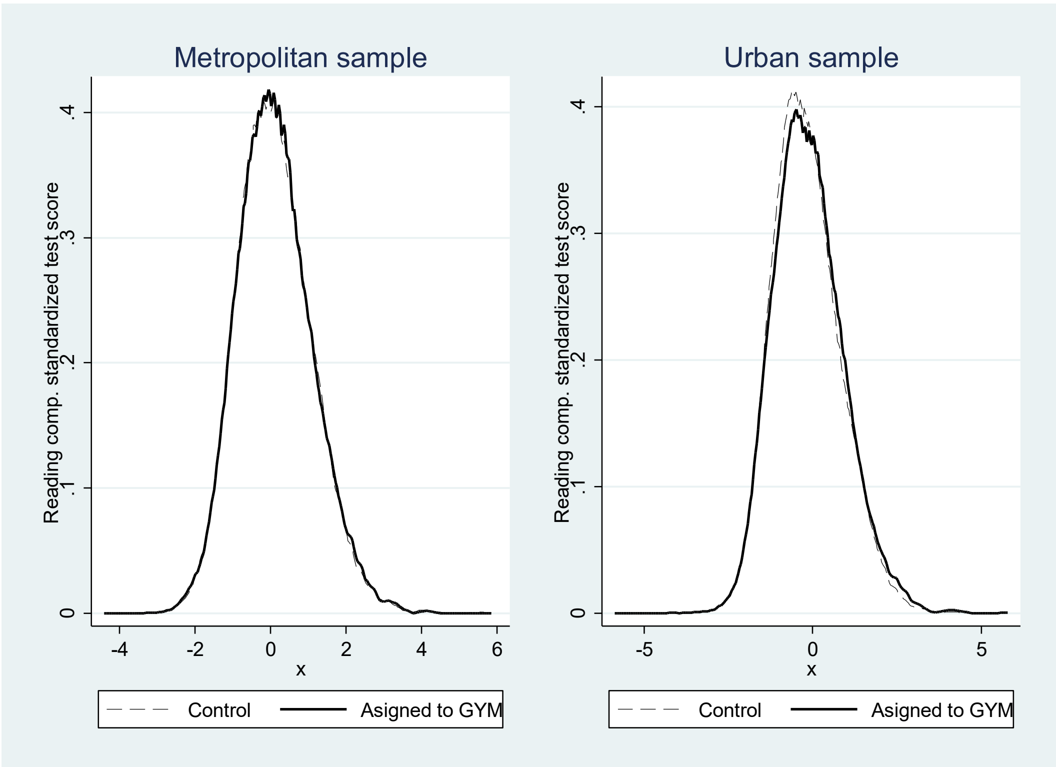
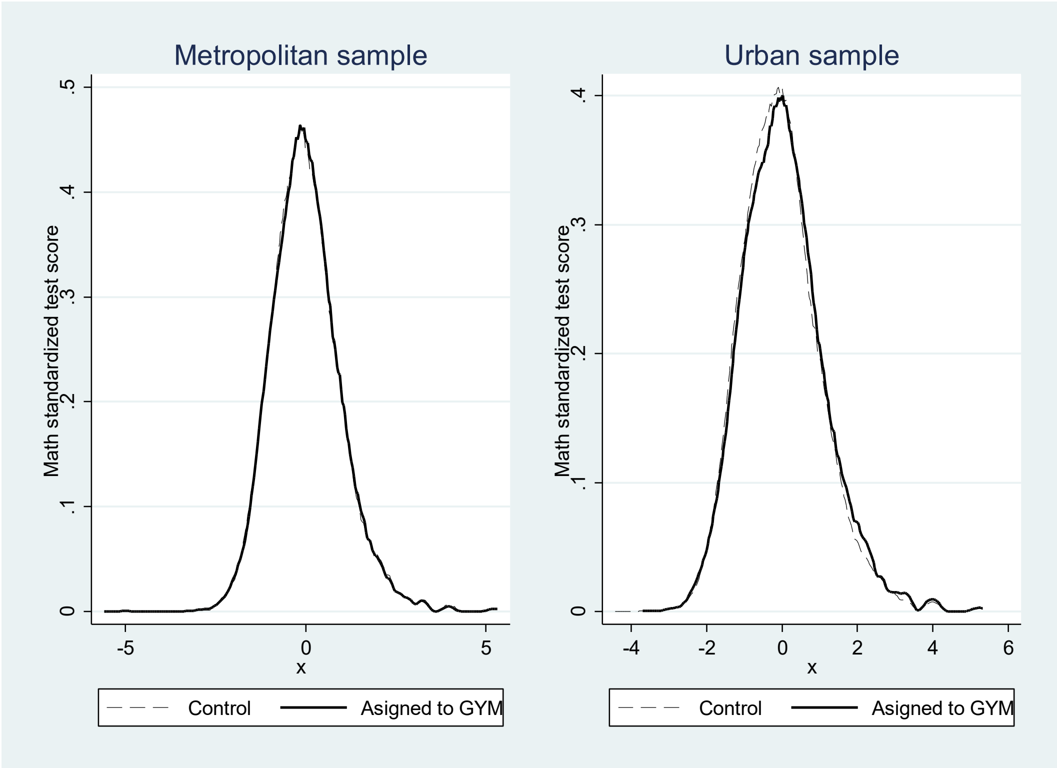
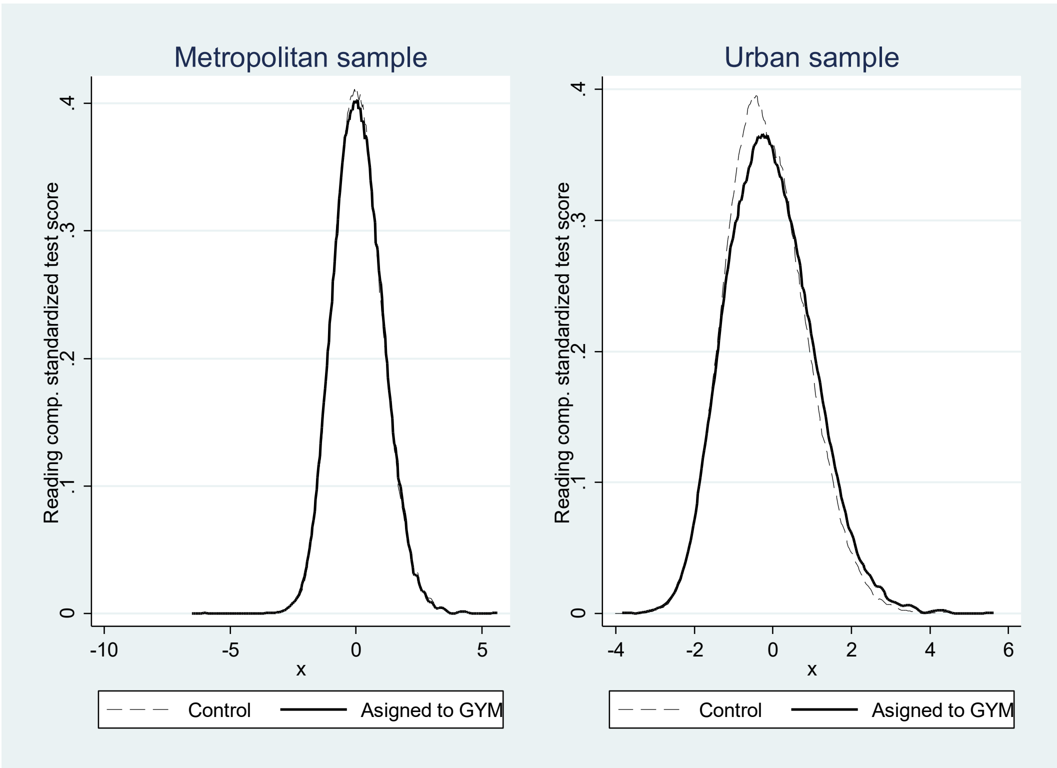


Figure 3: kernel densities of test scores in 2016 (medium-term effects)

Part 2.a. Mathematics



Part 2.b. Reading comprehension



Part 2.c. History, Geography, and Economics

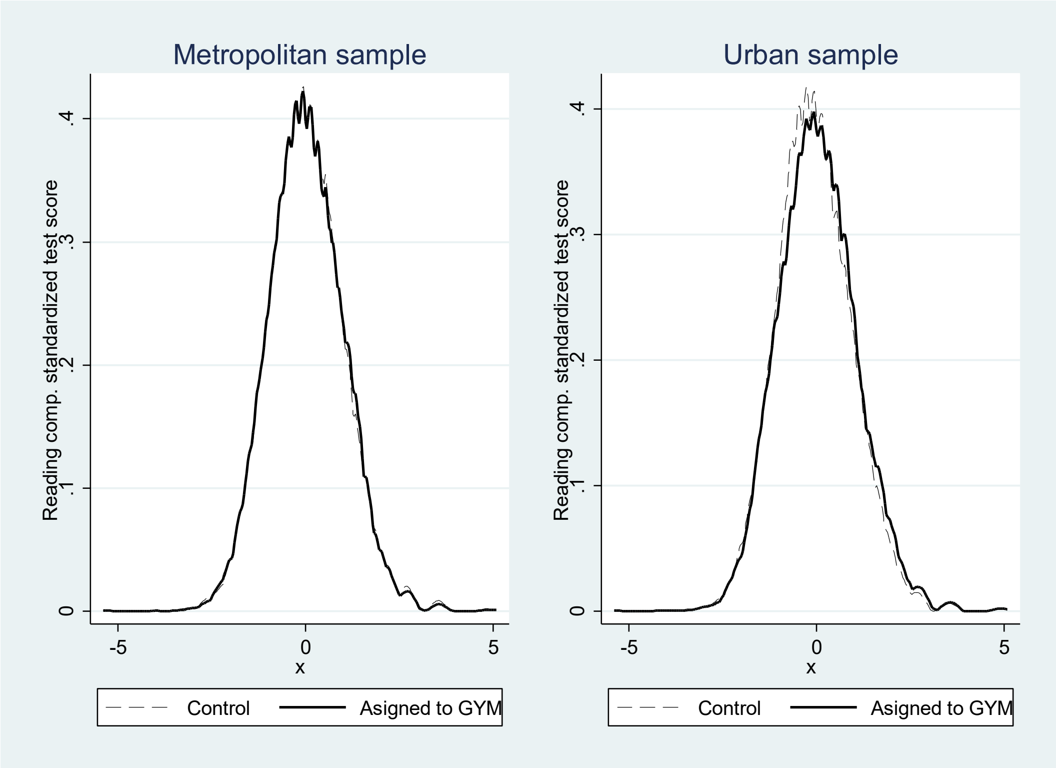


Table 1:
Experimental design: number of schools assigned to treatment by sub-sample and quartile

	Metropolitan sample					Regional sample					Total
	Quartile					Quartile					
	1	**2**	**3**	**4**	**Total**	**1**	**2**	**3**	**4**	**Total**	
Assignment:											
Control	3	35	56	101	**195**	70	91	23	21	**205**	400
Treatment	3	36	56	101	**196**	70	91	23	20	**204**	400
						0	0	0	0		
Total	**6**	**71**	**112**	**202**	**391**	**140**	**182**	**46**	**41**	**409**	**800**

Table 2:
GYM package and GYM sessions: delivery and compliance rates

	Full sample	Metropolitan sample	Regional sample
Panel A: Assigned to treatment			
Number of schools	**400**	**196**	**204**
Compliance among those schools assigned to treatment (in %)			
Administered at least 1 GYM session	*60*	*56*	*65*
Sent evidence of at least 1 GYM session	*55*	*50*	*59*
Sent evidence of all GYM sessions:			
7th grade	*43*	*36*	*49*
8th grade	*45*	*39*	*50*
Panel B: Received GYM package			
Assigned to treatment (in %)	**85**	**79**	**91**
Number of schools	***340***	***154***	***186***
Compliance among those schools that received GYM package (in %)			
Administered at least 1 GYM session	*71*	*71*	*71*
Sent evidence of at least 1 GYM session	*64*	*64*	*65*
Sent evidence of all GYM sessions:			
7th grade	*51*	*48*	*53*
8th grade	*53*	*50*	*54*

Table 3:
School characteristics

School characteristics	Total		
	Total	Metropolitan sample	Urban sample
Panel A: main characteristics			
By day shift (%)			
Morning	60%	38%	81%***
Afternoon	20%	30%	10%***
Morning & Afternoon	20%	32%	9%***
By sex (%)			
Males only	1%	1%	2%
Females only	2%	1%	3%
Mixed	96%	97%	96%
By bilingual status (%)			
No bilingual school	3%	0%	6%***
Panel B: School size (second grade of secondary)			
By number of classrooms (%)			
1 classroom	35%	22%	47%***
2 classrooms	24%	23%	26%
3 classrooms	14%	20%	9%***
4 classrooms	9%	12%	6%*
5 classrooms	7%	9%	6%*
6 or more classrooms	10%	14%	7%***
By number of students (%)			
Less than 40 students	46%	29%	62%***
Between 41 and 80 students	15%	33%	18%***
Between 81 and 120 students	12%	16%	8%***
Between 121 y 160 students	8%	11%	5%***
More than 160 students	9%	12%	7%***
Average number of students per classroom	23	24	20***
Panel C: Physical infrastructure			
Science laboratory (%)	69%	73%	64%***
Library (%)	75%	73%	77%
Access to computers (%)	97%	98%	95%*
Access to electricity (%)	98%	99%	96%***
Access to water (%)	91%	91%	91%
Access to sewage (%)	85%	91%	80%***
Panel D: Access to social programs in the district			
Cash conditional transfer program (JUNTOS) (%)	28%	0%	56%***
District located in VRAEM area (%)	4%	0%	7%***
Number of schools	*799*	*391*	*408*
Panel E: Performance in ECE (student level, results from control group)*			

Grade attainment in Mathematics			
Prior to beginning level	37.8%	35.0%	41.8%***
Beginning level	44.4%	46.0%	41.2%***
In process level	11.8%	12.6%	10.6%***
Satisfactory	6.1%	6.3%	5.7%
Number of students	*27,476*	*16,491*	*10,985*
Grade attainment in Reading comprehension			
Before beginning level	21.2%	16.9%	27.7%***
Beginning level	45.3%	45.3%	45.2%
In process level	23.0%	25.7%	18.9%***
Satisfactory	10.5%	12.1%	8.1%***
Number of students	*27,474*	*16,487*	*10,987*

Note: In the third column, we report results from t-tests to calculate differences between the metropolitan and regional samples (*** p<0.01, ** p<0.05, * p<0.1). In Panel E, grade attainments are calculated using the information at the student level and correspond to the control group.

Table 4:
Short-term impact of the GYM package in test scores

		Dep. variable: Mathematics test score	Dep. variable: Reading comprehension test score
		(I)	(II)
Panel A: ITT estimates			
Full sample	*Std. Coef.*	0.054*	0.040
	Std. Error	(0.030)	(0.028)
	n	[54,510]	[54,526]
Metropolitan sample	*Std. Coef.*	0.001	0.008
	Std. Error	(0.035)	(0.036)
	n	[32,810]	[32,826]
Regional sample	*Std. Coef.*	0.135***	0.088*
	Std. Error	(0.051)	(0.045)
	n	[21,700]	[21,700]
Panel B: LATE estimates			
Total	*Std. Coef.*	0.153*	0.112
	Std. Error	(0.086)	(0.081)
	n	[54,510]	[54,526]
Metropolitan sample	*Std. Coef.*	0.003	0.024
	Std. Error	(0.107)	(0.109)
	n	[32,810]	[32,826]
Regional sample	*Std. Coef.*	0.345**	0.225*
	Std. Error	(0.140)	(0.122)
	N	[21,700]	[21,700]

Note: Each coefficient comes from a different model specification (*** p<0.01, ** p<0.05, * p<0.1). All specifications control for strata dummies and have standard errors clustered at the school level. In Panel A (intent-to-treat estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school was assigned to receive a GYM package, 0 otherwise. In Panel B (LATE estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school complied with the GYM intervention, 0 otherwise. Compliance is instrumentalized by a school assignment to the GYM package (ivreg2 command in STATA is used).

Table 5:
Short-term impact of the GYM package in different segments of the test scores distribution

	n	(I) Dep. Variable: Prob(Before beginning=1)			(II) Dep. Variable: Prob(Beginning=1)			(III) Dep. Variable: Prob(In process=1)			(IV) Dep. Variable: Prob(Satisfactory=1)		
		Mean	ITT (a)	LATE (b)	Mean	ITT (a)	LATE (b)	Mean	ITT (a)	LATE (b)	Mean	ITT (a)	LATE (b)
Panel A: Dependent variable: Mathematics test score													
Full sample	54,510	38%	-0.020*	-0.056*	44%	0.000	0.001	12%	0.009*	0.027*	6%	0.010*	0.028*
			(0.010)	(0.030)		(0.006)	(0.016)		(0.005)	(0.014)		(0.006)	(0.016)
Metropolitan sample	32,810	35%	0.001	0.003	46%	-0.004	-0.012	13%	0.001	0.004	6%	0.001	0.004
			(0.013)	(0.038)		(0.007)	(0.020)		(0.006)	(0.018)		(0.007)	(0.020)
Regional sample	21,700	42%	-0.052***	-0.132***	42%	0.007	0.017	11%	0.022***	0.056**	6%	0.023**	0.059**
			(0.018)	(0.049)		(0.010)	(0.026)		(0.008)	(0.023)		(0.009)	(0.025)
Panel B: Dependent variable: Reading comprehension test score													
Full sample	54,526	21%	-0.010	-0.027	45%	-0.006	-0.016	23%	0.006	0.018	11%	0.009	0.025
			(0.008)	(0.022)		(0.007)	(0.020)		(0.006)	(0.017)		(0.007)	(0.020)
Metropolitan sample	32,826	17%	-0.000	-0.001	45%	-0.001	-0.004	26%	-0.002	-0.005	12%	0.004	0.011
			(0.008)	(0.026)		(0.009)	(0.028)		(0.007)	(0.023)		(0.009)	(0.030)
Regional sample	21,700	28%	-0.023	-0.061	45%	-0.012	-0.030	19%	0.019*	0.048*	8%	0.017*	0.043*
			(0.015)	(0.039)		(0.011)	(0.029)		(0.010)	(0.027)		(0.010)	(0.026)

Note: Each coefficient comes from a different model specification (*** p<0.01, ** p<0.05, * p<0.1). All specifications control for strata dummies and have standard errors clustered at the school level. The four dependent variables in blocks (I), (II), (III) and (IV) are binary variables that take the value of 1 if a student's test scores are classified as "Before beginning", "Beginning", "In process", and "Satisfactory", respectively (0 otherwise). In (a), the coefficient corresponds to an (independent) binary variable that takes the value of 1 if the school was assigned to receive a GYM package, 0 otherwise. In column (b), the coefficient is associated with a binary variable that takes the value of 1 if the school complied with the GYM intervention, 0 otherwise. Compliance is instrumentalized by a school assignment to the GYM package (ivreg2 command in STATA is used).

Table 6:
Short-term heterogeneous effects of the GYM package (ITT estimates): by gender and ethnicity

	(I)				(II)			
	n	Female (a)	Male (b)	F-test (p-value)	n	Non-Spanish (a)	Spanish (b)	F-test (p-value)
Panel A: Dependent variable: Mathematics test score								
Full sample	54,382	0.044	0.067**	(0.501)	52,713	0.040	0.056*	(0.421)
		(0.039)	(0.030)			(0.050)	(0.030)	
Metropolitan sample	32,726	-0.003	0.010	(0.764)	31,404	-0.033	0.001	(0.599)
		(0.049)	(0.035)			(0.068)	(0.036)	
Regional sample	21,656	0.117*	0.153***	(0.436)	21,309	0.038	0.144***	(0.134)
		(0.061)	(0.053)			(0.066)	(0.053)	
Panel B: Dependent variable: Reading comprehension test score								
Full sample	54,397	0.043	0.038	(0.861)	52731	-0.007	0.042	(0.303)
		(0.034)	(0.029)			(0.046)	(0.029)	
Metropolitan sample	32,742	0.009	0.008	(0.982)	31422	0.005	0.007	(0.966)
		(0.043)	(0.037)			(0.071)	(0.036)	
Regional sample	21,655	0.094*	0.084*	(0.803)	21309	-0.009	0.098**	(0.110)
		(0.054)	(0.047)			(0.067)	(0.047)	

Note: In each row, paired coefficients (Female/Male in block (I); Non-Spanish/Spanish in block (II)) come from the same model specification (*** p<0.01, ** p<0.05, * p<0.1). Estimates are obtained from intent-to-treat models. In block (I), the assignment variable (1 if the school was assigned to receive a GYM package, 0 otherwise), is interacted with a Male dummy, and the model controls for a Male dummy. In block (II), the assignment variable is interacted with a Spanish dummy, and the model controls for a Spanish dummy. All specifications control for strata dummies and have standard errors clustered at the school level. In each case, the F-test corresponds to the null hypothesis that the coefficient is the same for each sub-group.

Table 7:
Medium-term impact of the GYM package

		Dep. variable: Mathematics test score	Dep. variable: Reading comprehension test score	Dep. variable: History, geography, and economics test score
		(I)	(II)	(III)
	Panel A: ITT estimates			
Full sample	*Std. Coef.*	0.038	0.044	0.036
	Std. Error	(0.029)	(0.029)	(0.025)
	n	[54,933]	[54,980]	[53,678]
Metropolitan sample	*Std. Coef.*	0.001	-0.003	-0.005
	Std. Error	(0.033)	(0.035)	(0.029)
	n	[34,019]	[34,066]	[32,968]
Regional sample	*Std. Coef.*	0.099*	0.122**	0.103**
	Std. Error	(0.052)	(0.050)	(0.043)
	n	[20,914]	[20,914]	[20,710]
	Panel B: LATE estimates			
Total	*Std. Coef.*	0.111	0.128	0.104
	Std. Error	(0.083)	(0.086)	(0.072)
	n	[54,933]	[54,980]	[53,678]
Metropolitan sample	*Std. Coef.*	0.005	-0.008	-0.017
	Std. Error	(0.103)	(0.110)	(0.092)
	n	[34,019]	[34,066]	[32,968]
Regional sample	*Std. Coef.*	0.254*	0.312**	0.264**
	Std. Error	(0.138)	(0.136)	(0.116)
	n	[20,914]	[20,914]	[20,710]

Note: Each coefficient comes from a different model specification (*** p<0.01, ** p<0.05, * p<0.1). All specifications control for strata dummies and have standard errors clustered at the school level. In Panel A (intent-to-treat estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school was assigned to receive a GYM package the year before (0 otherwise). In Panel B (LATE estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school complied with the GYM intervention the year before (0 otherwise). Compliance is instrumentalized by a school assignment to the GYM package (ivreg2 command in STATA is used).

Table 8:
Impact of the GYM package on educational expectations and self-perceptions

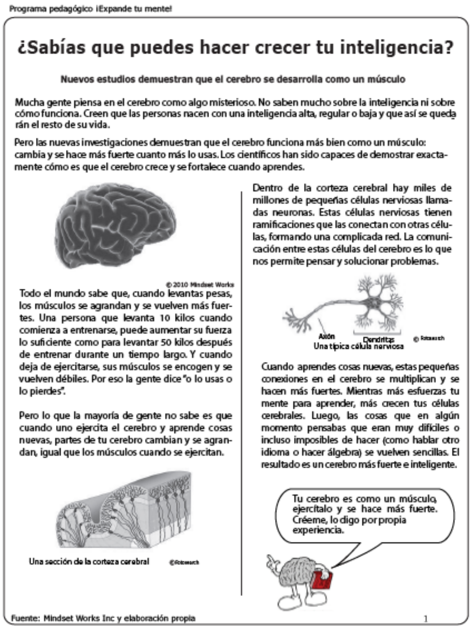
	Dep. variable: expectation for university education (short-term estimate)		Dep. variable: expectation for university education (medium-term estimate)		Dep. variable: Standardized student`s self-perception of own capacity in math (short-term estimate)	Dep. variable: Standardized student`s self-perception of own capacity in reading comprehension (short-term estimate)
	Mean	Coef.	Mean	Coef.	Coef.	Coef.
		(I)		(II)	(III)	(IV)
		Panel A: ITT estimates				
Full sample	80%	0.011*	78%	0.010	-0.002	-0.008
		(0.006)		(0.006)	(0.016)	(0.017)
		[51,896]		[53,539]	[48,890]	[48,914]
Metropolitan sample	81%	0.007	79%	0.006	-0.007	-0.022
		(0.007)		(0.007)	(0.020)	(0.022)
		[30,899]		[32,938]	[29,192]	[29,151]
Regional sample	80%	0.018*	78%	0.016	0.004	0.012
		(0.011)		(0.011)	(0.027)	(0.028)
		[20,997]		[20,601]	[19,698]	[19,763]
		Panel A: LATE estimates				
Full sample	80%	0.032*	78%	0.028	-0.007	-0.024
		(0.018)		(0.018)	(0.045)	(0.048)
		[51,896]		[53,539]	[48,890]	[48,914]

Metropolitan sample	81%	0.021	79%	0.019	-0.021	-0.067
		(0.023)		(0.023)	(0.061)	(0.065)
		[30,899]		[32,938]	[29,192]	[29,151]
Regional sample	80%	0.046	78%	0.042	0.011	0.030
		(0.029)		(0.029)	(0.068)	(0.072)
		[20,997]		[20,601]	[19,698]	[19,763]

Note: Each coefficient comes from a different model specification (*** p<0.01, ** p<0.05, * p<0.1). All specifications control for strata dummies and have standard errors clustered at the school level. In Panel A (intent-to-treat estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school was assigned to receive a GYM package, 0 otherwise. In Panel B (LATE estimates), the coefficient is associated with a binary variable that takes the value of 1 if the school complied with the GYM intervention, 0 otherwise. Compliance is instrumentalized by a school assignment to the GYM package (ivreg2 command in STATA is used). For short-term estimates, assignment/compliance and outcomes are measured the same calendar year (2015); for medium-term estimates, outcomes are measured the next calendar year (2016).

Annex 1

‘Grow Your Mind!’ article



Programa pedagógico ¡Expande tu mente!

¿Sabías que puedes hacer crecer tu inteligencia?

Nuevos estudios demuestran que el cerebro se desarrolla como un músculo

Mucha gente piensa en el cerebro como algo misterioso. No saben mucho sobre la inteligencia ni sobre cómo funciona. Creen que las personas nacen con una inteligencia alta, regular o baja y que así se quedarán el resto de su vida.

Pero las nuevas investigaciones demuestran que el cerebro funciona más bien como un músculo: cambia y se hace más fuerte cuanto más lo usas. Los científicos han sido capaces de demostrar exactamente cómo es que el cerebro crece y se fortalece cuando aprendes.

© 2010 Mindset Works

Todo el mundo sabe que, cuando levantas pesas, los músculos se agrandan y se vuelven más fuertes. Una persona que levanta 10 kilos cuando comienza a entrenarse, puede aumentar su fuerza lo suficiente como para levantar 50 kilos después de entrenar durante un tiempo largo. Y cuando deja de ejercitarse, sus músculos se encogen y se vuelven débiles. Por eso la gente dice "o lo usas o lo pierdes".

Pero lo que la mayoría de gente no sabe es que cuando uno ejercita el cerebro y aprende cosas nuevas, partes de tu cerebro cambian y se agrandan, igual que los músculos cuando se ejercitan.

Una sección de la corteza cerebral ©Fotosearch

Dentro de la corteza cerebral hay miles de millones de pequeñas células nerviosas llamadas neuronas. Estas células nerviosas tienen ramificaciones que las conectan con otras células, formando una complicada red. La comunicación entre estas células del cerebro es lo que nos permite pensar y solucionar problemas.

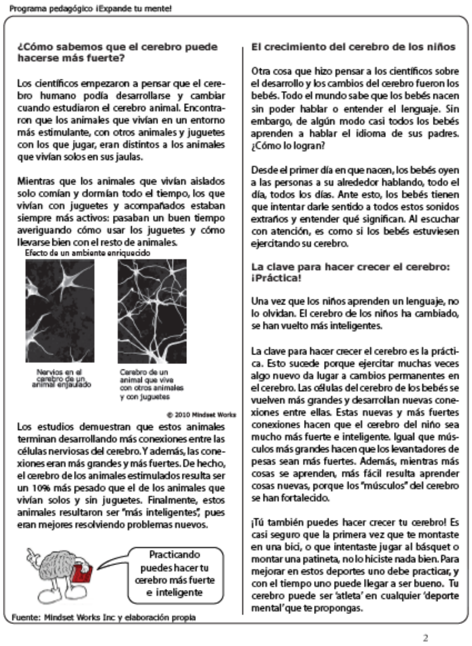
Axón Dendritas © Fotosearch

Una típica célula nerviosa

Cuando aprendes cosas nuevas, estas pequeñas conexiones en el cerebro se multiplican y se hacen más fuertes. Mientras más esfuerzas tu mente para aprender, más crecen tus células cerebrales. Luego, las cosas que en algún momento pensabas que eran muy difíciles o incluso imposibles de hacer (como hablar otro idioma o hacer álgebra) se vuelven sencillas. El resultado es un cerebro más fuerte e inteligente.

Tu cerebro es como un músculo, ejercítalo y se hace más fuerte. Créeme, lo digo por propia experiencia.

Fuente: Mindset Works Inc y elaboración propia 1



Programa pedagógico ¡Expande tu mente!

¿Cómo sabemos que el cerebro puede hacerse más fuerte?

Los científicos empezaron a pensar que el cerebro humano podía desarrollarse y cambiar cuando estudiaron el cerebro animal. Encontraron que los animales que vivían en un entorno más estimulante, con otros animales y juguetes con los que jugar, eran distintos a los animales que vivían solos en sus jaulas.

Mientras que los animales que vivían aislados solo comían y dormían todo el tiempo, los que vivían con juguetes y acompañados estaban siempre más activos: pasaban un buen tiempo averiguando cómo usar los juguetes y cómo llevarse bien con el resto de animales.

Efecto de un ambiente enriquecido

Nervios en el cerebro de un animal enjaulado

Cerebro de un animal que vive con otros animales y con juguetes

© 2010 Mindset Works

Los estudios demuestran que estos animales terminan desarrollando más conexiones entre las células nerviosas del cerebro. Y además, las conexiones eran más grandes y más fuertes. De hecho, el cerebro de los animales estimulados resulta ser un 10% más pesado que el de los animales que vivían solos y sin juguetes. Finalmente, estos animales resultaron ser "más inteligentes", pues eran mejores resolviendo problemas nuevos.

Practicando puedes hacer tu cerebro más fuerte e inteligente

El crecimiento del cerebro de los niños

Otra cosa que hizo pensar a los científicos sobre el desarrollo y los cambios del cerebro fueron los bebés. Todo el mundo sabe que los bebés nacen sin poder hablar o entender el lenguaje. Sin embargo, de algún modo casi todos los bebés aprenden a hablar el idioma de sus padres. ¿Cómo lo logran?

Desde el primer día en que nacen, los bebés oyen a las personas a su alrededor hablando, todo el día, todos los días. Ante esto, los bebés tienen que intentar darle sentido a todos estos sonidos extraños y entender qué significan. Al escuchar con atención, es como si los bebés estuviesen ejercitando su cerebro.

La clave para hacer crecer el cerebro: ¡Práctica!

Una vez que los niños aprenden un lenguaje, no lo olvidan. El cerebro de los niños ha cambiado, se han vuelto más inteligentes.

La clave para hacer crecer el cerebro es la práctica. Esto sucede porque ejercitar muchas veces algo nuevo da lugar a cambios permanentes en el cerebro. Las células del cerebro de los bebés se vuelven más grandes y desarrollan nuevas conexiones entre ellas. Estas nuevas y más fuertes conexiones hacen que el cerebro del niño sea mucho más fuerte e inteligente. Igual que músculos más grandes hacen que los levantadores de pesas sean más fuertes. Además, mientras más cosas se aprenden, más fácil resulta aprender cosas nuevas, porque los "músculos" del cerebro se han fortalecido.

¡Tú también puedes hacer crecer tu cerebro! Es casi seguro que la primera vez que te montaste en una bici, o que intentaste jugar al básquet o montar una patineta, no lo hiciste nada bien. Para mejorar en estos deportes uno debe practicar, y con el tiempo uno puede llegar a ser bueno. Tu cerebro puede ser 'atleta' en cualquier 'deporte mental' que te propongas.

Fuente: Mindset Works Inc y elaboración propia

2

‘Grow Your Mind!’ poster