MEL
MOROCCO INNOVATION AND EVALUATION LAB

A collaboration between

UM6P
University Mohammed VI Polytechnic

J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

HARVARD CENTER FOR INTERNATIONAL DEVELOPMENT

Developing Socioemotional Skills in Early Adolescence: Experimental Evidence from Morocco's Public Schools

Andreas de Barros
Alejandra Campos Quintero
Paul Glewwe
Nikhil Kumar
Laure Lépine

June 17, 2026

Developing Socioemotional Skills in Early Adolescence: Experimental Evidence from Morocco's Public Schools*

Andreas de Barros[†]
Alejandra Campos Quintero[‡]
Paul Glewwe[§]
Nikhil Kumar[¶]
Laure Lépine[‖]

June 17, 2026

We investigate a government-delivered socioemotional support intervention in Morocco's public lower-secondary schools. The intervention trained existing school social specialists to deliver small-group workshops to students in grades 7 and 8. In a cluster-randomized trial across 200 schools, all in the first year of a broader reform, 84 were assigned to the intervention and 116 to control. Assignment increased workshop participation by 43.2 percentage points and students' knowledge of their social specialist by 16.9 percentage points. Over one school year, however, this increase in exposure did not translate into detectable average effects on the pre-specified primary outcomes (interpersonal and intrapersonal skills) or on overall learning, dropout, creativity, well-being, or study habits. If anything, students' self-control declined by 0.097 standard deviations. Yet, we observe improvements in two sets of secondary outcomes: grade repetition fell by 1.7 percentage points, and student achievement increased in select subjects and subdomains: physics, chemistry, and

*This document presents the phase-two disposition of a Registered Report conditionally accepted by the Journal of Development Economics pre-results review process. We thank the Ministry of National Education, Preschools and Sports in Morocco for supporting both this study and a separate impact evaluation of the Pioneer School Program in public lower-secondary schools. We gratefully acknowledge generous funding from the Ministry. We are grateful to Naila Ebeid, Alice Escande, Naomi Duhau-Marmon, and Maxime Sirera at the Behavioural Insights Team (BIT), who designed the socioemotional support intervention with the Ministry and conducted the diagnostic analysis and qualitative research that informed it. We are grateful to the Morocco Innovation and Evaluation Lab (MEL) and J-PAL staff—Sara Ameziane-Hassani, Youssef Assarssah, Arthur Bluche, Ayoub Chaabi, Florencia Devoto, Najiba El Amrani Mida, and Zakaria Mansouri—for their excellent field coordination, research assistance, and project leadership. We thank Quentin Daviot and Thomas Escande of EVAL-LAB for their excellent support in the data collection process. Data protection was secured in compliance with the relevant local legal provisions and regulations concerning Human Subjects research at the University of California, Irvine, Harvard University, and the University of Minnesota. The order of authors is alphabetical; co-authorship is shared equally among the authors.

[†]University of California, Irvine. E-mail: adb@uci.edu.
[‡]Columbia University. E-mail: aac2271@tc.columbia.edu.
[§]University of Minnesota. E-mail: pglewwe@umn.edu.
[¶]Harvard University. E-mail: nikhil_kumar1@g.harvard.edu.
[‖]University Mohammed VI Polytechnic, E-mail: llepine@povertyactionlab.org.

biology, written expression in Arabic and French, and oral expression in Arabic. All of these results hold under a pre-specified hierarchy of hypothesis tests that controls the false discovery rate. Thus, although existing school staff can expand students' exposure to support without additional staff or pay, a low-intensity universal protocol did not measurably improve the targeted socioemotional skills; instead, positive effects fell on academic and progression outcomes.

Keywords: Education; socioemotional skills; human capital; Morocco.

JEL classification: I20, I21, I28, O15, O22.

Study pre-registration: This study was pre-registered as a Registered Report and conditionally accepted by the *Journal of Development Economics* through its pre-results review process.

1 Introduction

School systems increasingly aim to develop socioemotional skills that shape learning, persistence, relationships, and later-life outcomes (Deming, 2017). Early adolescence brings new academic demands, changing peer relationships, and heightened risk of disengagement from school. Yet policies for producing socioemotional skills among children in this age group at scale remain poorly understood. Teachers who raise test scores are not necessarily the same teachers who improve socioemotional outcomes (Jackson, 2018), and much of the evidence on targeted socioemotional programming comes from high-income settings or from interventions delivered by researchers and non-governmental organizations. Less is known about whether governments can use existing public-school personnel to deliver socioemotional support inside ordinary schools.

This paper studies a government-led socioemotional support intervention in Morocco's public lower-secondary schools. The intervention was embedded in the first year of the Pioneer School Program, a broader reform of those schools. It trained and equipped school social specialists to deliver four two-hour workshops per year in small classes for students in grades 7 and 8. The workshops focused on self-awareness, emotion recognition and regulation, belonging, cooperation, empathy, active listening, and conflict resolution. The intervention also aimed to increase students' familiarity with their school's social specialist and to build trust between students and the specialist.

We evaluate the intervention through a cluster-randomized trial across 200 public lower-secondary schools, all participating in the Pioneer School Program. Of these, 84 were randomly assigned to receive the socioemotional support intervention, while 116 served as a control group. The design therefore identifies the marginal effect of assignment to the socioemotional component within this reform setting. Our primary outcomes are students' interpersonal and intrapersonal skills. We also examine downstream outcomes, including academic skills and dropout, as well as additional outcomes related to creativity, well-being, and study habits.

The intervention changed exposure to socioemotional support. Students in treatment schools were 43.2 percentage points more likely to report participating in socioemotional workshops, relative to the control mean of 21.7 percent. They were also 16.9 percentage points more likely to know their school social specialist, relative to a control mean of 43.5 percent. Exposure to other components of the Pioneer School Program was balanced across treatment and control schools. These results indicate that the intervention created a substantial contrast in students' interaction with the social specialists and the workshops. At the same time, roughly one in three treatment students did not report participating in

a workshop. We view this incomplete take-up as a feature of studying the intent-to-treat effects of a reform delivered at scale by the government.

Despite this increase in exposure, we find no detectable average improvements in the pre-specified primary outcomes over the first year. Specifically, the estimated effect on interpersonal skills is 0.029 standard deviations (s.d.), and that on intrapersonal skills is 0.012 s.d., both of which are small and statistically indistinguishable from zero. If anything, one subdomain of intrapersonal skills declined by 0.097 s.d. (students' self-control). We also find no detectable effects on overall academic skills or dropout rates. The estimated effect on test scores is 0.034 s.d., and the estimated effect on dropout by the end of the school year is approximately zero. Additional outcomes, including creativity, school climate and well-being, and study habits, similarly show no detectable effects.

Yet, improvements appear in two sets of secondary outcomes: grade repetition fell by 1.7 percentage points, and student achievement rose in select subjects and subdomains, including physics, chemistry, and biology, written expression in Arabic and French, and oral expression in Arabic. All of these estimates emerge from a pre-specified hierarchy of hypothesis tests that controls the false discovery rate, even as overall test scores and dropout show no detectable change.

These findings contribute to three literatures. First, we add to a small literature on school-based socioemotional interventions in low- and lower-middle-income countries. Meta-analyses and reviews document impacts of socioemotional programs in high-income settings (Wilson et al., 2025; Cipriano et al., 2023), and related studies in middle-income contexts evaluate interventions such as growth-mindset or socioemotional curricula (Alan et al., 2019; Ganimian, 2020; Santos et al., 2022). Evidence from lower-income settings remains thinner and often comes from researcher- or NGO-delivered programs in Central America, India, and Zambia (Dinarte-Diaz et al., 2024; Edmonds et al., 2023; Dhar et al., 2022; Ashraf et al., 2020). To our knowledge, this is the first large-scale, experimental evaluation of a government-led socioemotional skill intervention implemented in public schools in a low- or lower-middle-income country.

Second, the paper contributes to research on public-sector implementation of education reforms. Many successful education interventions in developing countries have relied on NGOs, private-public partnerships, or other non-governmental implementation structures (Banerjee et al., 2017; Eble et al., 2021; Romero et al., 2020). Attempts to improve public schools using only state personnel and state administrative systems have often struggled (de Barros et al., 2024). The intervention studied here was not a stand-alone pilot delivered outside the public education system; it was implemented through regular public schools and social specialists.

Third, the results speak to the design of socioemotional supports during adolescence. Existing work in low- and lower-middle-income countries often focuses on early childhood, primary school, or late adolescence (e.g., Dillon et al., 2017; Barrera-Osorio et al., 2024; Dam et al., 2025; Beaman et al., 2021; Krishnan and Krutikova, 2013). We add to this work by studying socioemotional support delivered at the start of lower-secondary school, a developmental window in which disengagement and dropout risk rise, and by characterizing how a broad, low-intensity design implemented through existing public-school staff performs when a government takes it to scale.

The rest of the paper proceeds as follows. Section 2 describes the Moroccan lower-secondary school context and the socioemotional support intervention. Section 3 presents the study design and analytical strategy. Section 4 reports the results, Section 5 presents heterogeneity and robustness analyses, and Section 6 concludes.

2 Context and Intervention

2.1 Morocco's Public Lower-Secondary Schools

Public provision of education, as governed by the Ministry of National Education, Preschools and Sports, is the most common type of primary and secondary education in Morocco. Even though the share of private enrollment has increased over the recent years, as of 2019, public schools still served 83 percent of primary school students and 89 percent of lower secondary school students in the country (The World Bank, 2025). Enrollment rates are high, with net enrollment rates of 99.6 percent at the primary level and 90.6 percent at the lower secondary level in 2019 (The World Bank, 2025).

From 2018 to 2023, the lower secondary completion rate increased from 64.5 to 74.2 percent (from World Bank World Development Indicators). However, these high enrollment and completion rates mask very low student performance. In the 2021 PIRLS reading assessment, Morocco's fourth-graders ranked second to last (out of 57 countries), and more than half of the students (59 percent) did not reach the minimum proficiency benchmark (Mullis et al., 2023). In the 2023 TIMSS math assessment, Morocco's fourth graders ranked third to last (out of 58 countries), and more than half (54 percent) did not attain the minimum proficiency benchmark (von Davier et al., 2024). At the secondary level, in the 2023 TIMSS math assessment, Morocco's eighth graders ranked second to last out of 42 countries, and 64 percent of grade 8 students were below the lowest international benchmark in mathematics (von Davier et al., 2024).

We conducted this study within the context of a broader reform effort to address low learning levels in Morocco's public lower secondary schools. The reform is locally known as the "Pioneer School Program (PSP)". The socioemotional skill intervention is one intervention component of that reform. The program was launched in 232 of Morocco's public lower secondary schools in September of 2024.[1] The Ministry considers the program to be its flagship intervention in lower secondary schools, and it has started to scale the program to another 535 schools in the 2025-26 school year, covering approximately 32 percent of the students in the country. The remaining schools are expected to be reached by 2028.

2.2 The Socioemotional Support Intervention

The program's diagnostic phase drew on the socio-ecological model (Kilanowski, 2017) to map multiple levels of factors that may contribute to dropping out of school in Morocco, spanning individual, relational, community, and societal factors. The socioemotional support intervention itself is grounded in the foundational life-skills framework of the World Health Organization (World Health Organization, 1993, 2003) and the CASEL framework for social and emotional learning (Durlak et al., 2015). Its hypothesized link to dropout is best understood through the participation-identification model of Finn (1989), which casts school withdrawal as a cumulative process driven by declining behavioral and emotional engagement, providing a direct rationale for targeting students' sense of belonging, self-efficacy, and relationship with a trusted adult.

In practice, the intervention consists of training and equipping social specialists to give four two-hour workshops per year in small classes for students in the first and second year of lower secondary education (grades 7 and 8).[2] These workshops aim to promote awareness of the role of the social specialist, build a relationship of trust with the students, and develop intrapersonal and interpersonal skills with the aim of improving individual well-being, adaptive capacity, and school climate. The workshops are also designed to pay particular attention to at-risk students in need of individualized support and to refer them for individual follow-up.

The staff member in charge of implementing the socioemotional support intervention is the school's social specialist. Twelve social specialists across Morocco were selected to receive

[1] The 232 schools volunteered to participate in the first year of the reform. The launch in lower secondary schools comes on the heels of launching the PSP program in primary schools one year prior. We evaluate the overall effect of (the bundle of interventions provided by) the reform in both types of schools in other research.

[2] The social specialist is a civil servant who has been trained after a competitive examination open to graduates in psychology, philosophy, or social science. Their work consists of ensuring student well-being at school and identifying students in need of support or attention. A social specialist is assigned to a school and works on-site a total of 24 hours per week.

training and, thereafter, train other social specialists in the schools we randomly assigned to receive the intervention. Social specialists received training slides and detailed session guides, and students received booklets for each workshop, along with a notebook to record their feelings and emotions.

Each workshop focuses on a specific topic related to intra- or interpersonal skills. The first workshop focuses on self-awareness, covering self-assessment (strengths and weaknesses) and a growth mindset. The second workshop focuses on effective listening and empathic communication, and managing interpersonal difficulties and conflict. The third workshop focuses on identifying one's own emotions and regulating stress. The fourth workshop focuses on cooperation, collaboration, and students' sense of belonging.

2.3 Theory of Change

In terms of *need*, the intervention aimed to address two related challenges common to many low- and middle-income countries. First, initial qualitative insights into Morocco's lower secondary schools revealed an environment largely unconducive to developing socioemotional skills, marked by a high prevalence of bullying and intimidation, conflicted relationships between students and teachers, feelings of insecurity, and general distress and discomfort among students. Second, while personnel exists (in this case, social specialists), existing services that could promote socioemotional skills appear ineffective, with limited support for students, support relying solely on voluntary participation, social specialists being assigned to administrative tasks (or only to the most serious, disciplinary cases), a lack of clarity about responsibilities, and an absence of tools and training. Taken together, the study context appears to reveal a serious need for the intervention and a high likelihood that it would generate a meaningful contrast with the status quo.

The *inputs* offered by the intervention are training sessions for social specialists working in the schools and guidebooks on how to implement the intervention (described in the previous section), following a government order for the specialists to do so. Other inputs for schools selected for the ongoing reform are held constant; the intervention is implemented with the existing resources, with no additional staff or pay allocated to intervention schools.

The expected *outputs* are that social specialists hold four two-hour workshops per year in small classes for students in the first and second years of lower secondary education and that students participate in these workshops. The expected *intermediate outcomes*, mechanisms, and secondary outcomes are that students' well-being increases, their study habits improve, and students think more creatively. The intervention may also trigger changes in students' participation in extra-curricular activities, but to the extent that these

activities do not foster academic or socio-emotional skills, this effect may have detrimental consequences on downstream variables.

The expected *outcomes* are that students improve their interpersonal (or social) skills and their intrapersonal skills. The former relates to positive impacts on emotion perception and pro-sociality. The latter relates to positive impacts on students' perceived control (incl. growth mindset, locus of control, and perceived self-efficacy) and students' self-regulation and discipline (incl. self-control and grit).[3]

Lastly, the expected *impact* is that students improve their academic skills (in Arabic, French, math, and science) and become less likely to drop out of school.

3 Study Design

3.1 Sampling of Schools

Our sample consists of 200 public lower-secondary schools. We constructed the sample in two steps, using administrative data on the universe of government lower secondary schools, which includes the 232 lower-secondary PSP schools in Morocco. First, using the "post-double-selection" (PDS) methodology (Belloni et al., 2014), 30 variables were identified that were predictive of either test scores or participation in the PSP reform. Second, using these 30 selected variables, Mahalanobis nearest-neighbor matching (without replacement) was used to identify 100 pairs of PSP schools that were very similar to one another (in terms of their Mahalanobis distance).[4]

Table 1 presents an analysis of the sample's representativeness, both with respect to the population of lower secondary schools in Morocco and to other PSP schools in the country, along with balance checks. Columns (1) to (3) compare the 200 (public) lower secondary schools in the study's sample of schools with the 2,344 remaining (public) lower secondary schools in Morocco.[5] Columns (4) to (6) present the representativeness within the reform program by comparing the 200 program (PSP) lower secondary schools in the study sample with the 32 other PSP lower secondary schools in Morocco. Overall, schools in the study

[3]Here, we focus on expected outcomes that arose from our discussions with the Ministry prior to pre-registering the study. They are broad goals that also align with the wider "Pioneer School" reform process and its evaluation, as presented in our companion research (de Barros et al., 2026). We acknowledge that the intervention developers may have instead focused on a narrower set of more proximal outcomes.

[4]We initially considered assigning 100 PSP schools to the socioemotional support intervention, in a pairwise, stratified trial. However, as specified during the phase-1 review and registration of the study design, we discarded this plan. The following subsection describes our random assignment of 84 PSP schools to the intervention, as pre-specified.

[5]At the stage of finding matched control schools for each of the PSP schools, we had received data on 2,544 public lower secondary schools.

sample have slightly more teachers, are less likely to be rural, have more students, and serve students with a slightly higher average primary and middle-school school leaving exam scores.

3.2 Random Assignment of Schools

Of the 200 PSP schools, we randomly assigned 84 to receive the socioemotional support intervention (and the remaining 116 not to receive that intervention). Specifically, to conduct a stratified randomization strategy, we started by calculating a principal component based on the average primary school passing grade, number of teachers, number of students, and fraction of female teachers in the 200 schools. We then used this principal component to sort the schools into an alternating pattern of triplets and pairs (84 groups in total). Finally, within each group, we randomly assigned one school to receive the socioemotional support intervention.

Columns (7) to (9) of Table 1 present the balance among the two groups of schools included in this study. School characteristics are balanced between treatment and control groups, with a small imbalance in the percentage of students who are female and the average age of students. The joint test of balance is marginal (column (9); $p = 0.101$). Reassuringly, the student-level balance checks presented just below show no systematic differences between the two groups.[6]

3.3 Sampling of Students

The study's unit of analysis is a student. In each of the study's lower secondary schools, surveyors were given a priority list of randomly selected students for each grade. We constructed these lists before the baseline data collection began, using enrollment records. The lists allowed for random replacements if students were absent on the day of the baseline assessment.

Students were sampled to take assessments in only one subject (Arabic, French, mathematics, or science). For the written Arabic and French assessments, as well as the one-on-one oral Arabic and French assessments and interviews capturing socioemotional skills, we subsampled up to 12 students per school (6 students from grade 7, and 6 students

[6] All joint F-tests in Tables 1, 2, and A1 report randomization-inference p-values, following Kerwin et al. (2024).

from grade 8).[7] For the written math and science assessments, we subsampled up to 18 students per school (all attending grade 7).[8]

Specifically, the students in each grade and track (APIC or ASCG) were randomly split into four groups, one for each subject.[9] Within each group of students sampled for a subject, students were randomly ordered into a ranked list of students. This randomly ordered list of students ensured that each grade and subject had the required number of students sampled for the assessments and the interviews, as well as that the number of students on the list from each track was proportional to the number of enrolled students in each track for that subject. Thus, a randomized priority list of students was generated, which had students with "high" priority who enumerators would try to assess/interview first before moving down the list to the students with "low" priority.

The study's effective sample consists of 11,132 students (6,483 in the control group and 4,649 in the treatment group). These are all the students who took the baseline assessment and who we attempted to track for the endline assessment. Among these students, 2,138 were tested in Arabic, 2,148 in French, 3,428 in math, and 3,418 in science.

Table 2 examines balance at baseline in the sampled students, in terms of learning outcomes, socioemotional skills, and student characteristics, between the control schools and the treatment schools. As expected, given the random assignment of the socioemotional support intervention and the random sampling described above, the students in the control schools are generally similar to the students in the treatment schools, and joint F-tests confirm no systematic differences across the two groups.[10] Individually, two baseline characteristics differ at the 10-percent level: treatment students score 0.083 s.d. higher on the baseline Arabic assessment and are slightly younger. Our intent-to-treat specification controls for baseline student characteristics and outcomes (Section 3.6), and neither difference would survive the multiple-hypothesis-testing correction we apply to our treatment effects, so these small imbalances do not affect our conclusions. Attrition between baseline and endline was low, at about 5 percent overall, with a slight imbalance of roughly one percentage point between the treatment and control groups; Section 5.2 shows that our estimates are robust to this differential attrition.

[7] We also sampled and tested up to 6 students in grade 9, but since the socioemotional skill intervention focuses on the lower grades only, we exclude them from our data. We also specified this decision during the phase-1 review of our analysis plan.

[8] For these two subjects, we did not include students from grades 8 or 9 as the broader PSP reform did not yet include math and science interventions for these grades.

[9] APIC refers to "Année Secondaire Collégial Originel Parcours International" and ASCG refers to "Année Secondaire Collégial Général". The two are the academic tracks into which lower-secondary students are streamed, with APIC denoting the internationally oriented pathway and ASCG the general pathway.

[10] Table A1 examines the balance at baseline among the non-attriting sample of 10,526 students and confirms that non-attriting students in the two randomized groups look similar at baseline.

3.4 Outcome Measures

Our measurement framework aligns closely with the Theory of Change described above.[11] The pre-specified primary outcomes were two families of socioemotional skills. The interpersonal skills index combines pro-sociality and emotion perception. Pro-sociality is measured using the pro-social subscale of the Strengths and Difficulties Questionnaire (Goodman, 1997). Emotion perception is measured using the Perceiving AI-Generated Emotions scale (PAGE), which asks students to recognize emotions in facial expressions (Weidmann and Xu, 2024).

The intrapersonal skills index combines perceived control and self-regulation/discipline. Perceived control includes growth mindset, locus of control, and self-efficacy. Self-regulation and discipline include work discipline and grit, as well as self-control (Duckworth and Quinn, 2009; Huillery et al., 2025). These measures capture students' beliefs about their own agency, their ability to persist in the face of difficulty, and their capacity to manage attention and behavior.

We also pre-specified downstream outcomes. Academic skills are measured through written assessments in Arabic, French, mathematics, and science. In Arabic and French, students also completed one-on-one oral assessments. The main learning outcome stacks standardized scores across subjects. School-progression outcomes are measured using administrative records and include dropout or exclusion by the end of the school year, non-re-enrollment at the beginning of the following school year, repetition, and exclusion.

Additional outcomes capture potential mechanisms and broader developmental domains. These include creativity, school climate and well-being, and study habits. Creativity is measured using tasks based on the Torrance Tests of Creative Thinking (Torrance, 1968). School climate and well-being combine measures of belonging, bullying, and perceived stress (Cohen et al., 1983). Study habits are measured by whether students report spending more than 30 minutes per day on homework after school.

Finally, we measure take-up and implementation using student reports at endline. The key exposure outcomes are whether students participated in socioemotional workshops and whether they know the school social specialist. We also measure exposure to the other components of the Pioneer School Program (PSP), including targeted remediation, explicit teaching, and tutoring.

[11] As in this description of the Theory of Change, and our pre-registration of broad outcome families, we acknowledge that the intervention developers may have instead chosen a measurement framework with a narrower set of proximal outcome measures that match the workshop content more tightly. Bailey et al. (2020) caution that "overalignment" between intervention content and outcome measures is a pervasive concern in education evaluations; for related evidence, see de Barros and Lubozha (2026).

Item-level responses are aggregated using item response theory models chosen to match the structure of each instrument (Jacob and Rothstein, 2016). Binary items are scored using two-parameter logistic models, and ordered categorical items using graded response models. When multiple latent scores are combined into a family index, we use inverse-covariance-weighted averages. Endline scores are standardized relative to the endline control-group distribution. Appendix B provides details on measurement and scoring.[12]

Appendix Table B1 reports the average conditional reliability of each instrument at endline. Precision is high for the academic assessments (0.78–0.90) and for most socioemotional components, with emotion perception the main exception (0.51); because every instrument enters the analysis as an outcome rather than a right-hand-side variable, this measurement error inflates standard errors but does not bias the treatment-effect estimates.

3.5 Program Exposure

Panel A of Table 3 shows that the intervention substantially increased students' exposure to socioemotional support. In control schools, 21.7 percent of students reported participating in socioemotional workshops. Assignment to the treatment group increased this share by 43.2 percentage points. The intervention also increased students' awareness of their school social specialist by 16.9 percentage points, relative to a control mean of 43.5 percent. Both estimates are statistically significant at the one-percent level.[13]

In turn, Panel B of Table 3 shows that treatment and control schools had similar exposure to the other components of the Pioneer School Program (PSP). We find no meaningful differences in students' reported receipt of targeted remediation, explicit teaching, or tutoring. This lack of effect supports interpreting the estimates below as reflecting the effect of the socioemotional support component rather than differential exposure to the broader reform package.[14]

3.6 Empirical Strategy

Our identification strategy rests on the study's random assignment of schools to the two experimental groups within each randomization stratum. We estimate the intent-to-treat

[12]For related papers on the measurement of socioemotional skills in lower-income countries, see Laajaj and Macours (2021) and Danon et al. (2024).

[13]Panel A of tables A2 and A3 shows that the intervention led to a similar increase in exposure to socioemotional support among students at risk of dropout and among female students, respectively.

[14]Panel B of tables A2 and A3 shows that the intervention did not lead to a differential exposure to components of the broader PSP reform package among students at risk of dropout and among female students, respectively.

(ITT) effect of the socioemotional support component on the outcomes of interest by comparing lower secondary schools randomly assigned to receive the socioemotional support intervention with those randomly assigned not to receive the intervention. For all outcomes, we use the following empirical specification:

$$Y_{igsr}^{t=1} = \lambda_r + \beta_1 T_{sr} + \boldsymbol{\delta}' \boldsymbol{X}_{igsr}^{t=0} + \epsilon_{igsr} \tag{1}$$

Here, $Y_{igsr}^{t=1}$ is the endline outcome for student i in grade g in school s and randomization stratum r. The fixed-effect term λ_r absorbs randomization-stratum effects; for the survey-outcomes, we additionally absorb subject fixed effects, and for dropout outcomes, we absorb grade-by-stratum effects only. T_{sr} is the treatment dummy indicating a school's random assignment to the socioemotional support intervention, and ϵ_{igsr} is the residual. The coefficient of interest is β_1, which captures the intent-to-treat effect of assignment to the socioemotional support intervention among schools in the Pioneer School Program (PSP). To increase precision, all specifications include $\boldsymbol{X}_{igsr}^{t=0}$ as covariates. Measured at baseline ($t = 0$), $\boldsymbol{X}_{igsr}^{t=0}$ is a vector of baseline controls selected by a post-double selection (PDS) Lasso procedure on student and school characteristics, partialing out the stratum fixed effects and, when available, $Y_{igsr}^{t=0}$ (a student's outcome of interest at baseline).[15] All inference clusters at the school level.

We also examine effects among two pre-specified subgroups: students flagged as being at risk of dropout and female students. As pre-specified, we interact the treatment indicator with the subgroup indicator and report the overall effect for each subgroup.

We adjust for multiple hypothesis testing and report sharpened false discovery rate (FDR) adjusted q-values computed using the Anderson (2008) procedure, following the tiered outcome-family approach in Vivalt et al. (2024). The main family of outcomes (K0) consists of interpersonal and intrapersonal skills. The second family of outcomes (K1) consists of academic skills and dropout. The third family (K2) consists of creativity, school climate and well-being, and study habits. Component- and subgroup-level tests are adjusted within the relevant pre-specified testing families. Appendix C provides additional details on the pre-specified multiple hypothesis testing procedure.

[15] We thank Jacobus Cilliers for pointing us to Cilliers et al. (2024) and for recommending this approach. Following Cilliers et al. (2024), we use the default, "plug-in" penalty parameter of Stata's *pdslasso* command (not cross-validation).

4 Results

4.1 Socioemotional Skills

Table 4 reports effects on the two primary socioemotional outcomes. In panel A, we find that assignment to the intervention led to a small and statistically insignificant change in the interpersonal skills index by +0.029 s.d. When we examine the individual components of this index, the estimate for pro-sociality is -0.007 s.d. and that for emotion perception is $+0.058$ s.d. The emotion perception estimate is statistically significant at the 5-percent significance level before adjustment for multiple hypothesis testing, but does not remain statistically significant after adjustment; pro-sociality is not statistically significant at conventional levels.

In panel B, we find that assignment to the intervention did not lead to a detectable change in the intrapersonal skills index. When we focus on the individual components of this index, perceived control increased by 0.045 s.d. and self-regulation and discipline decreased by 0.027 s.d., neither of which is statistically significant.

These results indicate that, despite increases in reported exposure to workshops and social specialists, the intervention did not generate detectable average improvements in students' socioemotional skills over the first year. One subdomain is an exception: among the self-regulation and discipline measures, students' self-control declined by 0.097 s.d. (Appendix Table A4), an estimate that is statistically significant and survives adjustment for multiple hypothesis testing ($q = 0.003$). The remaining underlying socioemotional-skill subdomains follow the broader pattern: the estimates are small in magnitude and do not remain statistically significant after adjustment for multiple hypothesis testing.

4.2 Learning

Table 5 reports effects on academic skills and student dropout. In panel A, we find that assignment to the intervention led to a small and statistically insignificant increase in test scores by 0.034 s.d. When we focus on individual subjects, scores increased by 0.023 s.d. in Arabic, 0.070 s.d. in French, and 0.067 s.d. in science, while math scores decreased by 0.033 s.d.; none of these subject-level estimates is statistically significant after adjustment for multiple hypothesis testing.

Within science, the estimated increase of 0.123 s.d. in physics and chemistry is statistically significant and remains so after adjustment ($q = 0.016$), while the 0.033 s.d. increase in life science is not. Decomposing the science assessment further (Appendix Table A5), the gains

appear in physics (0.148 s.d.), chemistry (0.135 s.d.), and biology (0.095 s.d.), each of which survives adjustment for multiple hypothesis testing.

Among the language subdomains (Appendix Table A5), the positive estimates concentrate in written production. Assignment raised written-production scores by 0.124 s.d. in Arabic and 0.132 s.d. in French, each significant at the 1-percent level before adjustment and similar in magnitude across the two language assessments. The Arabic written-production estimate remains statistically significant after adjustment for multiple hypothesis testing at the 5-percent level ($q = 0.044$), while the French written-production estimate remains significant at the 10-percent level ($q = 0.095$). Arabic oral expression shows a comparable gain (0.118 s.d.) that likewise remains significant after adjustment at the 5-percent level ($q = 0.044$), while the French oral estimate is small and statistically insignificant (0.047 s.d.), so we read the oral result more cautiously. These estimates fall in the components that require students to produce language (writing and, in Arabic, speech) rather than in receptive reading comprehension, a pattern consistent with workshops centered on communication and expression but also with greater student engagement on open-ended productive tasks.

4.3 Dropout and Repetition

In panel B of Table 5, we find no evidence that the intervention reduced dropout by the end of the school year. The estimated effect on our main dropout measure is an increase of 0.1 percentage point in the share of students who did not continue the following year, whether voluntarily or due to exclusion, relative to a control mean of 2.6 percent. The estimates for dropout alone, non-re-enrollment in the following school year, and exclusion are also small and statistically insignificant after adjustment for multiple hypothesis testing. Assignment to the intervention reduced grade repetition by 1.7 percentage points, relative to a control mean of 9.6 percent, and this estimate remains statistically significant at the 5-percent significance level after adjusting for multiple hypothesis testing.

4.4 Additional Outcomes

Table 6 examines additional outcomes. We find that assignment to the intervention did not lead to detectable improvements in creativity, school climate and well-being, or study habits. Creativity decreased by 0.079 s.d., while school climate and well-being increased by 0.014 s.d., and the share of students spending more than 30 minutes per day on after-school homework increased by 0.1 percentage point. The creativity estimate is statistically significant at the 10-percent significance level before adjustment for multiple hypothesis

testing, but does not remain statistically significant after adjustment; the other estimates are not statistically significant at conventional levels. These estimates do not support the hypothesis that the intervention led to measurable changes in well-being, creativity, or out-of-school study habits.

5 Heterogeneity and Robustness

5.1 Subgroup Effects

Table 7 reports pre-specified subgroup analyses for students flagged as being at risk of dropout at the beginning of the study period and for female students. In panel A, we find that assignment to the intervention increased intrapersonal skills among at-risk students by 0.144 s.d. This estimate is statistically significant at the 10-percent significance level before adjustment for multiple hypothesis testing but does not remain statistically significant after adjustment ($q = 0.281$). The corresponding estimate for interpersonal skills is 0.004 s.d. and is not statistically significant. We find no evidence of effects on at-risk students' learning, dropout, creativity, school climate and well-being, or study habits.

In panel B, we find no evidence of effects for female students. Assignment to the intervention increased interpersonal skills by 0.043 s.d., decreased intrapersonal skills by 0.014 s.d., and increased learning by 0.064 s.d. The estimated effect on the global dropout measure is essentially zero. The learning estimate is statistically significant at the 10-percent level before adjustment for multiple hypothesis testing, but does not remain statistically significant after adjustment; the other estimates are small and statistically insignificant, and the estimates on additional outcomes are similarly small and insignificant.[16]

[16] In addition to the above pre-registered investigation of program impacts among at-risk and female students, we conducted an exploratory analysis using causal forests (Wager and Athey, 2018) to assess whether treatment effects varied with a broader set of baseline student and school characteristics, evaluating the implied prioritization rule with the rank-weighted average treatment effect (RATE) of Yadlowsky et al. (2025). The only outcome with a RATE significantly different from zero was intrapersonal skills, and the estimate was negative (RATE $= -0.095$, s.e. 0.029): the students the forest ranks as most likely to benefit had lower-than-average treatment effects. This agrees in substance with the pre-specified subgroup analysis even though the point estimates differ in sign: there, the at-risk intrapersonal estimate is positive but does not survive correction (0.144 s.d., $q = 0.281$), and the forest likewise fails to isolate a higher-responding group, so neither establishes a subgroup with above-average effects. We therefore do not interpret this analysis as identifying a beneficiary subgroup, in either direction.

5.2 Robustness and Sensitivity

As the sample of initially enrolled students varies by the month in which it is constructed, we assess the robustness of dropout-related findings to alternative sample definitions. Table A7 shows that the estimated effects on dropout are stable across alternative definitions of the enrollment sample.

We also examine whether social desirability bias affects the estimates of effects on socioemotional skills, following Dhar et al. (2022). Table A8 shows no evidence that the effects on socioemotional skills differ meaningfully between students with above- and below-median baseline social-desirability scores. The difference in effects is 0.022 s.d. for interpersonal skills and -0.056 s.d. for intrapersonal skills, and neither estimate is statistically significant. We take this as evidence that, more broadly, our results are unlikely to be affected by social-desirability bias in students' survey responses.

Finally, we assess whether the slight differential attrition between treatment and control schools could account for our findings. Appendix Table A9 reports treatment effects on the two socioemotional indices and the overall test-score index under three approaches to attrition: Lee (2009) bounds, the tightened bounds of Behaghel et al. (2015) based on survey tracking effort, and inverse-probability weighting on baseline covariates. Across all three approaches, the estimates are close to those in our main analysis, which we take as evidence that, more broadly, the results we present are not affected by differential attrition.

6 Conclusion

This paper investigated a government-delivered socioemotional support intervention in Morocco's public lower-secondary schools. The intervention trained and equipped existing school social specialists to deliver small-group workshops to students in grades 7 and 8. In a cluster-randomized trial across 200 schools, assignment to the intervention increased students' reported participation in workshops by 43.2 percentage points and their awareness of the school social specialist by 16.9 percentage points.

Despite this increase in reported exposure, we find no detectable average effects on the pre-specified primary outcomes (interpersonal and intrapersonal skills), nor on overall learning, dropout, creativity, well-being, or study habits. If anything, students' self-control declined by 0.097 s.d. Yet, improvements appear in two sets of secondary outcomes: grade repetition fell by 1.7 percentage points ($q = 0.032$), and achievement rose in select science and language subdomains, namely physics, chemistry, and biology, written expression in Arabic and French, and oral expression in Arabic. All of these estimates emerge from a pre-specified hierarchy of hypothesis tests that controls the false discovery rate.

We read the reduction in grade repetition as consistent with the program strengthening the social specialist's role in supporting students who might otherwise repeat a grade, the kind of individualized attention the intervention was designed to encourage. Because the intervention did not target subject content, we treat the gains in physics, chemistry, and biology as unexplained estimates that call for additional research. The language gains are more interpretable: written-production scores rose by 0.124 s.d. in Arabic and 0.132 s.d. in French, and Arabic oral expression by 0.118 s.d., concentrating in the productive components of language rather than in receptive reading comprehension, a pattern consistent with the workshops' emphasis on communication and expression; we interpret them as evidence that the intervention may have strengthened students' productive language skills. We have no comparable account for the decline in self-control, which we leave as a puzzle for future work.

We conclude with three qualifications. First, every study school participates in the broader Pioneer School Program, so our design identifies the marginal effect of the socioemotional component within a reforming system. It does not capture the effect of socioemotional programming relative to an unreformed status quo (nor does it speak to the overall effectiveness of the reform). If other components of the reform already addressed some of the needs the workshops target, the workshops' marginal contribution here may understate what socioemotional support would achieve in schools outside the reform; if the reform instead created conditions that helped the workshops operate, it may overstate it.

Second, the four workshops per year reflect the dose the Ministry chose to deploy at scale; meta-analyses of universal socioemotional programs typically find detectable average effects, mainly among programs that run a higher number of sessions of explicit skill practice (Cipriano et al., 2023). We also have limited insight into how faithfully this dose was delivered, since we observe students' self-reported participation but not the quality of the sessions that took place; a null finding on the targeted socioemotional skills may therefore reflect either small effects of well-run workshops at this dose or shortfalls in their delivery, and our data cannot separate the two. Future work could examine the intervention's dose-response relationship, including a higher-dose version of the workshops; qualitative research could assess how the sessions are delivered in practice.

Third, we measured socioemotional skills broadly, within a pre-registered outcome framework that also considers the Ministry's broader reform efforts (de Barros et al., 2026), rather than tracking a narrow set of proximal outcome measures tailored to the workshop content. Breadth is deliberate: it tests whether skill gains generalize to outcomes of general interest, guards against mechanical effects in instruments built around the intervention, and keeps our measures comparable to the broader literature. Its cost is

that some instruments sit at a distance from what the workshops most directly train.[17] Timing adds to this distance, since we observe students at the end of the same school year, when socioemotional skills may reflect initial acquisition rather than change that may become visible only at longer-term follow-ups. We therefore read the findings as most informative about the broad skill domains our framework was designed to capture, and as less informative about the more proximal skills the workshops trained most directly.

Taken together, the study shows that a government can use its existing school staff to change students' exposure to socioemotional support under routine public-school conditions, raising workshop participation and familiarity with the social specialist without additional staff or pay. At the dose and in the form deployed here, this contact did not detectably improve the socioemotional skills the workshops targeted, though it coincided with improvements in two sets of pre-specified downstream outcomes, grade repetition and achievement in select science and language subdomains, that survive correction for multiple hypothesis testing. Reaching students with socioemotional support through regular schools and existing staff thus appears feasible, but turning that contact into gains on the targeted socioemotional skills may require a more intensive, and potentially more targeted, design than the one tested here.

[17] For example, the emotion-perception measure asks students to read emotions in others rather than to identify their own, which the workshops treat as the basis for regulation; self-efficacy stands in for the broader self-esteem the sessions cultivate; and several targeted skills, including active listening, empathy, and conflict resolution, are not captured at all, while the relationship between students and the social specialist enters only as whether a student knows the specialist.

Table 1: *Sample of schools, representativeness, and balance tests*

	Representativeness (overall)			Representativeness (PSP)			Balance Checks		
	Non-Study	Study	Difference	Other PSP	Study PSP	Difference	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of teachers	29.61	32.38	2.77***	27.12	32.38	5.25**	32.88	31.69	-1.10
	[15.20]	[14.35]	(1.06)	[12.53]	[14.35]	(2.41)	[14.25]	[14.56]	(1.08)
Rural (%)	49.27	37.00	-12.27***	65.62	37.00	-28.62***	36.21	38.10	1.76
	[50.01]	[48.40]	(3.57)	[48.26]	[48.40]	(9.10)	[48.27]	[48.85]	(6.69)
Total Enrolment	774.30	913.21	138.90***	834.62	913.21	78.58	941.09	874.70	-60.21
	[457.18]	[466.13]	(34.22)	[457.39]	[466.13]	(86.48)	[487.00]	[435.60]	(41.64)
Female students (%)	47.30	47.09	-0.215	48.04	47.09	-0.948	47.46	46.58	-0.903**
	[4.90]	[2.90]	(0.229)	[3.53]	[2.90]	(0.650)	[2.99]	[2.72]	(0.409)
Age	14.60	13.94	-0.665***	14.58	13.94	-0.642***	13.96	13.91	-0.053*
	[0.365]	[0.218]	(0.017)	[0.260]	[0.218]	(0.048)	[0.203]	[0.235]	(0.028)
Primary school passing exam score	6.85	6.94	0.093***	6.79	6.94	0.150***	6.94	6.94	0.001
	[0.285]	[0.280]	(0.021)	[0.229]	[0.280]	(0.045)	[0.256]	[0.311]	(0.037)
Lower secondary school passing exam score	8.95	9.14	0.189*	8.75	9.14	0.386*	9.07	9.22	0.143
	[1.46]	[1.36]	(0.101)	[1.09]	[1.36]	(0.214)	[1.31]	[1.43]	(0.174)
Number of schools	2344	200	2544	32	200	232	116	84	200
Joint F-test (p-value)									0.101

Notes. This table reports on the study's sample of schools. Study refers to the 200 lower secondary schools included in the study's effective sample. Non-study refers to all other public lower secondary schools in Morocco. Study PSP refers to the 200 Pioneer Schools included in the study. Other PSP refers to the remaining 32 (lower secondary) Pioneer Schools in the country. Control refers to the 116 schools not assigned to receive the socioemotional support intervention, and Treatment refers to 84 schools assigned to receive this intervention. Difference reports on the regression-adjusted difference. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are heteroskedasticity-robust, since the unit of observation is the school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: *Balance checks*

	Number of observations		Balancing check	
	Control	Treatment	Control mean	Difference
	(1)	(2)	(3)	(4)
Panel A: Socioemotional skills				
Interpersonal skills	2509	1777	-0.368	-0.008
			[1.22]	(0.047)
Emotion perception	2509	1777	-0.288	-0.007
			[0.952]	(0.036)
Intrapersonal skills	2509	1777	1.90	0.001
			[0.796]	(0.027)
Perceived control	2509	1777	1.29	0.002
			[0.675]	(0.022)
Self-regulation and discipline	2509	1777	2.07	0.000
			[1.00]	(0.035)
Joint F-test (p-value)				0.997
Panel B: Academic skills and creativity				
Learning (stacked test scores)	6483	4649	-0.583	0.037
			[1.01]	(0.036)
Arabic	1256	882	-0.586	0.083*
			[1.06]	(0.048)
French	1253	895	-0.499	0.025
			[1.01]	(0.064)
Math	2002	1426	-0.355	0.043
			[1.08]	(0.045)
Science	1972	1446	-0.867	0.027
			[0.808]	(0.036)
Physics and Chemistry	1972	1446	-0.454	-0.006
			[0.786]	(0.032)
Life Science	1972	1446	-1.27	0.057
			[0.954]	(0.041)
Creativity	2509	1777	-0.320	0.029
			[1.04]	(0.054)
Joint F-test (p-value)				0.826
Panel C: Student characteristics				
Age	6483	4649	12.64	-0.039*
			[1.06]	(0.023)
Primary school passing exam score	6483	4649	7.10	0.032
			[1.17]	(0.037)
% Female	6483	4649	48.80	0.209
			[49.99]	(0.839)
% Attrited	6483	4649	5.92	-0.965*
			[23.61]	(0.515)
Joint F-test (p-value)				0.446

Notes. This table describes the study's sample of 11,132 students and presents balance checks. "Control" refers to 116 control schools. "Treatment" refers to the 84 treatment schools. "Difference" reports on the regression-adjusted difference between treatment schools and control schools, controlling for grade and for randomization-group fixed effects, in column (4). Reversed outcomes were flipped so higher scores represent desirable outcomes. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: *Program exposure*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Program exposure		
Student participated in socioemotional workshops	0.217	0.432***
	[0.412]	(0.023)
Knows social specialist	0.435	0.169***
	[0.496]	(0.032)
Panel B: Exposure to other interventions		
Student received TaRL	0.956	-0.002
	[0.205]	(0.009)
Student received explicit teaching	0.885	0.007
	[0.319]	(0.013)
Student participated in extra-curricular activities	0.666	0.005
	[0.472]	(0.019)
Student participated in tutoring program	0.503	-0.012
	[0.500]	(0.018)
At-risk student participated in tutoring program	0.674	0.034
	[0.470]	(0.034)

Notes. This table describes student exposure to the program in the 200 study schools among the non-attriting analysis sample; outcome availability varies by exposure measure. "Control" refers to the 116 Pioneer schools that did not receive the socioemotional workshops. "ITT Effect" reports on the regression-adjusted difference between treated schools and control schools, controlling for randomization-group fixed effects, in column (2). Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: *Intent-to-treat effects on socioemotional skills*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Interpersonal skills		
Overall (index)	-0.000	0.029
	[1.00]	(0.030)
		{1.000}
Pro-sociality	-0.000	-0.007
	[1.00]	(0.031)
		{1.000}
Emotion perception	-0.000	0.058**
	[1.00]	(0.029)
		{0.218}
Panel B: Intrapersonal skills		
Overall (index)	-0.000	0.012
	[1.00]	(0.029)
		{1.000}
Perceived control	-0.000	0.045
	[1.00]	(0.030)
		{0.762}
Self-regulation and discipline	-0.000	-0.027
	[1.00]	(0.027)
		{0.762}

Notes. This table reports on the program's intent-to-treat (ITT) effects following equation (1) on students' socioemotional skills among non-attriting students across the 200 government lower secondary schools who sat the written exams in Arabic or French. The estimation sample varies by outcome because of outcome-specific missingness. Column (1) reports the mean for the control group, and column (2) reports the ITT effect. "Prosociality" was not assessed at baseline, hence we do not include the baseline control when estimating the effect on this variable. Standard deviations for the control group means are shown in brackets; standard errors for the treatment effect are shown in parentheses and are clustered at the school level; *q*-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT) over the pre-specified outcome-family hierarchy, and following Vivalt et al. (2024). Main family measures are highlighted in bold font. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table 5: *Intent-to-treat effects on student learning and dropout*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Learning		
Overall (stacked test scores)	0.000	0.034
	[1.00]	(0.029)
		{1.000}
Arabic	0.000	0.023
	[1.00]	(0.037)
		{1.000}
French	-0.001	0.070*
	[1.00]	(0.039)
		{1.000}
Math	0.000	-0.033
	[1.00]	(0.042)
		{1.000}
Science	0.001	0.067
	[1.00]	(0.046)
		{1.000}
Physics & Chemistry	0.002	0.123***
	[1.00]	(0.042)
		{0.016}
Life Science	0.001	0.033
	[1.00]	(0.048)
		{0.394}
Panel B: Dropout and Repetition		
Dropout by end of school year (global: either excluded or dropout)	0.026	0.001
	[0.158]	(0.002)
		{1.000}
Dropout by end of the school year	0.023	0.001
	[0.150]	(0.002)
		{1.000}
Repeated at end of school year	0.096	-0.017***
	[0.295]	(0.006)
		{0.032}
Not re-enrolled in next school year	0.026	0.006*
	[0.159]	(0.004)
		{0.404}
Excluded by end of school year	0.003	0.000
	[0.053]	(0.001)
		{1.000}

Notes. This table reports on the program's intent-to-treat (ITT) effects following equation (1). Panel A reports the effect of the program on student learning among non-attriting students across the 200 government lower secondary schools. Panel B shows the effect of the program on dropout and repetition among all students present in the study schools. The estimation sample varies by outcome. Column (1) reports the mean of the control group, and column (2) reports the ITT effect. Standard deviations for the control group means are shown in brackets; standard errors for the treatment effect are shown in parentheses and are clustered at the school level; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT) over the pre-specified outcome-family hierarchy, and following Vivalt et al. (2024). Main family measures are highlighted in bold font. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table 6: *Intent-to-treat effects on other outcomes*

	Control mean	ITT Effect
	(1)	(2)
Creativity	0.000	-0.079*
	[1.00]	(0.045)
		{1.000}
School climate and well-being	0.000	0.014
	[1.00]	(0.035)
		{1.000}
Study habits (spent more than 30 min/day doing homework after school)	0.954	0.001
	[0.210]	(0.006)
		{1.000}

Notes. This table describes the potential mechanisms and other outcomes of the program for students in the 200 study schools among the non-attriting analysis sample. The estimation sample varies by outcome. Column (1) reports the mean of the control group and column (2) reports the ITT effect. Reversed outcomes were flipped so higher scores represent desirable outcomes. Standard deviations for the control group means are shown in brackets; standard errors for the treatment effect are shown in parentheses and are clustered at the school level; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT) over the pre-specified outcome-family hierarchy, and following Vivalt et al. (2024). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table 7: *Pre-registered analysis of subgroup effects*

	Control mean	ITT Effect
	(1)	(2)
Panel A: At-risk students		
Socioemotional skills		
Interpersonal skills	-0.260	0.004
	[1.10]	(0.089)
		{1.000}
Intrapersonal skills	-0.400	0.144*
	[0.866]	(0.075)
		{0.281}
Learning and dropout		
Learning (stacked test scores)	-0.609	0.015
	[0.762]	(0.048)
		{1.000}
Dropout by end of school year (global: either excluded or dropout)	0.085	-0.002
	[0.279]	(0.008)
		{1.000}
Additional outcomes		
Creativity	-0.196	-0.050
	[1.06]	(0.087)
		{1.000}
School climate and well-being	-0.194	0.002
	[1.00]	(0.083)
		{1.000}
Study habits (spent more than 30 min/day doing homework after school)	0.884	0.039
	[0.320]	(0.024)
		{1.000}
Panel B: Female students		
Socioemotional skills		
Interpersonal skills	0.102	0.043
	[0.971]	(0.041)
		{1.000}
Intrapersonal skills	0.065	-0.014
	[1.04]	(0.039)
		{1.000}
Learning and dropout		
Learning (stacked test scores)	0.169	0.064*
	[0.996]	(0.034)
		{0.868}
Dropout by end of school year (global: either excluded or dropout)	0.016	0.000
	[0.124]	(0.002)
		{1.000}
Additional outcomes		
Creativity	0.051	-0.054
	[1.00]	(0.051)
		{1.000}
School climate and well-being	0.148	-0.018
	[0.971]	(0.044)
		{1.000}
Study habits (spent more than 30 min/day doing homework after school)	0.963	-0.001
	[0.189]	(0.008)
		{1.000}

Notes. This table reports on the program's intent-to-treat (ITT) effects following equation (1) for pre-specified subgroup outcomes across the 200 government lower secondary schools. Panel A reports the effects for at-risk students, and Panel B reports the effects for female students. The estimation sample varies by outcome and subgroup. Rows are grouped by the corresponding main-table outcome families: socioemotional skills in Table 4, learning and dropout in Table 5, and additional outcomes and potential mechanisms in Table 6. Column (1) reports the mean of the control group and column (2) reports the ITT effect. Standard deviations for the control group means are shown in brackets; standard errors for the treatment effect are shown in parentheses and are clustered at the school level; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT) over the corresponding pre-specified subgroup and outcome-family hierarchy, and following Vivalt et al. (2024). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

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, 1121–1162.

Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association 103, 1481–1495. doi:10.1198/016214508000000841.

Ashraf, N., Bau, N., Low, C., McGinn, K., 2020. Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. The Quarterly Journal of Economics 135, 1095–1151. doi:10.1093/qje/qjz039.

Bailey, D.H., Duncan, G.J., Cunha, F., Foorman, B.R., Yeager, D.S., 2020. Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. Psychological Science in the Public Interest 21, 55–97. doi:10.1177/1529100620915848.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. Journal of Economic Perspectives 31, 73–102. doi:10.1257/jep.31.4.73.

Barrera-Osorio, F., de Barros, A., Filmer, D., 2024. Longterm Impacts of Primary School Scholarships: Evidence from Cambodia. Journal of Policy Analysis and Management 43, 10–38. doi:10.1002/pam.22533.

de Barros, A., Campos Quintero, A., El Amrani Mida, N., Glewwe, P., Kumar, N., Lépine, L., 2026. Beyond Basics: Whole-School Reform and Early Adolescent Development. Working Paper 260. Berkeley, CA. URL: https://doi.org/10.26085/C3RP5W.

de Barros, A., Fajardo-Gonzalez, J., Glewwe, P., Sankar, A., 2024. The Limitations of Activity-Based Instruction to Improve the Productivity of Schooling. The Economic Journal 134, 959–984. doi:10.1093/ej/uead099.

de Barros, A., Lubozha, T., 2026. Targeting Foundational Skills at Scale: Skill Specificity and Transfer. Working Paper 12542. Munich. URL: https://www.ifo.de/sites/default/files/docbase/docs/cesifo1_wp12542.pdf.

Beaman, L., Herskowitz, S., Keleher, N., Magruder, J., 2021. Stay in the Game: A Randomized Controlled Trial of a Sports and Life Skills Program for Vulnerable Youth

in Liberia. Economic Development and Cultural Change 70, 129–158. doi:10.1086/711651.

Behaghel, L., Crépon, B., Gurgand, M., Le Barbanchon, T., 2015. Please call again: Correcting nonresponse bias in treatment effect models. Review of Economics and Statistics 97, 1070–1080.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. Journal of Economic Perspectives 28, 29–50. doi:10.1257/jep.28.2.29.

Cilliers, J., Elashmawy, N., McKenzie, D., 2024. Using Post-Double Selection Lasso in Field Experiments. Working Paper 10931. The World Bank. Washington, D.C. URL: https://openknowledge.worldbank.org/entities/publication/0cde089d-33ba-4f51-8c03-b25b5114d41a.

Cipriano, C., Strambler, M.J., Naples, L.H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A.K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J.C., Durlak, J., 2023. The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. Child Development 94, 1181–1204. doi:https://doi.org/10.1111/cdev.13968.

Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. Journal of Health and Social Behavior 24, 385–396. doi:10.2307/2136404.

Dam, A., Gray-Lobe, G., Kremer, M., de Laat, J., Morsink, K., 2025. Learning to Work Towards Goals: A Sequential Evaluation of the Effect of Goal-Setting Course on Academic and Soft Skills. Working Paper 25-1344. Annenberg Institute at Brown University. Providence, RI. URL: https://edworkingpapers.com/sites/default/files/ai25-1344.pdf.

Danon, A., Das, J., de Barros, A., Filmer, D., 2024. Cognitive and Socioemotional Skills in Low-Income Countries: Measurement and Associations with Schooling and Earnings. Journal of Development Economics 168, 103132. doi:10.1016/j.jdeveco.2023.103132.

von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., Yin, L., 2024. TIMSS 2023 International Results in Mathematics and Science. Boston College, TIMSS & PIRLS International Study Center. doi:10.6017/lse.tpisc.timss.rs6460.

Deming, D., 2017. The Growing Importance of Social Skills in the Labor Market. The Quarterly Journal of Economics 132, 1593–1640. doi:10.1093/qje/qjx022.

Dhar, D., Jain, T., Jayachandran, S., 2022. Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. American Economic Review 112, 899–927. doi:10.1257/aer.20201112.

Dillon, M.R., Kannan, H., Dean, J.T., Spelke, E.S., Duflo, E., 2017. Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. Science 357, 47–55. doi:10.1126/science.aal4724.

Dinarte-Diaz, L., Egana-delSol, P., Martínez Alvear, C., Rojas Alvarado, C., 2024. When emotion regulation matters: The efficacy of socio-emotional learning to address school-based violence in Central America. Working Paper IDB-WP-1585. IDB Working Paper Series. doi:10.18235/0012854.

Duckworth, A.L., Quinn, P.D., 2009. Development and Validation of the Short Grit Scale (GritS). Journal of Personality Assessment 91, 166–174. doi:10.1080/00223890802634290.

Durlak, J.A., Domitrovich, C.E., Weissberg, R.P., Gullotta, T.P. (Eds.), 2015. Handbook of Social and Emotional Learning: Research and Practice. Guilford Press, New York.

Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Hsieh, P.T.J., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., Elbourne, D., 2021. How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. Journal of Development Economics 148, 102539. doi:10.1016/j.jdeveco.2020.102539.

Edmonds, E., Feigenberg, B., Leight, J., 2023. Advancing the Agency of Adolescent Girls. Review of Economics and Statistics 105, 852–866. doi:10.1162/rest_a_01074.

Finn, J.D., 1989. Withdrawing from school. Review of Educational Research 59, 117–142. doi:10.3102/00346543059002117.

Ganimian, A.J., 2020. Growth-Mindset Interventions at Scale: Experimental Evidence From Argentina. Educational Evaluation and Policy Analysis 42, 417–438. doi:10.3102/0162373720938041.

Goodman, R., 1997. The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry 38, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x.

Huillery, E., Bouguen, A., Charpentier, A., Algan, Y., Chevallier, C., 2025. The Role of Mindset in Education : A Large-Scale Field Experiment in Disadvantaged Schools. The Economic Journal , ueaf015doi:10.1093/ej/ueaf015.

Jackson, C.K., 2018. What Do Test Scores Miss? The Importance of Teacher Effects on NonTest Score Outcomes. Journal of Political Economy 126, 2072–2107. doi:10.1086/699018.

Jacob, B., Rothstein, J., 2016. The Measurement of Student Ability in Modern Assessment Systems. Journal of Economic Perspectives 30, 85–108. doi:10.1257/jep.30.3.85.

Kerwin, J., Rostom, N., Sterck, O., 2024. Striking the right balance: Why standard balance tests over-reject the null, and how to fix it .

Kilanowski, J.F., 2017. Breadth of the socio-ecological model. Journal of Agromedicine 22, 295–297.

Krishnan, P., Krutikova, S., 2013. Non-cognitive skill formation in poor neighbourhoods of urban India. Labour Economics 24, 68–85. doi:10.1016/j.labeco.2013.06.004.

Laajaj, R., Macours, K., 2021. Measuring Skills in Developing Countries. Journal of Human Resources 56, 1254–1295. doi:10.3368/jhr.56.4.1018-9805R1.

Lee, D.S., 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. The Review of Economic Studies 76, 1071–1102. doi:10.1111/j.1467-937X.2009.00536.x.

Mullis, I.V.S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K.A., Wry, E., 2023. PIRLS 2021 International Results in Reading. Boston College, TIMSS & PIRLS International Study Center. doi:10.6017/lse.tpisc.tr2103.kb5342.

Romero, M., Sandefur, J., Sandholtz, W.A., 2020. Outsourcing Education: Experimental Evidence from Liberia. American Economic Review 110, 364–400. doi:10.1257/aer.20181478.

Santos, I., Petroska-Beska, V., Carneiro, P., Eskreis-Winkler, L., Boudet, A.M.M., Berniell, I., Krekel, C., Arias, O., Duckworth, A.L., 2022. Can Grit Be Taught? Lessons from a Nationwide Field Experiment with Middle-School Students. Working Paper 15588. IZA Institute of Labor Economics. Bonn, Germany. URL: https://www.econstor.eu/handle/10419/265809.

The World Bank, 2025. Education Statistics (EdStats). URL: https://datatopics.worldbank.org/education/.

Torrance, E.P., 1968. Torrance Tests of Creative Thinking. Personnel, Princeton, NJ.

Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. doi:10.3386/w32719.

Wager, S., Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association 113, 1228–1242. doi:10.1080/01621459.2017.1319839.

Weidmann, B., Xu, Y., 2024. PAGE: A Modern Measure of Emotion Perception for Teamwork and Management Research. doi:10.48550/ARXIV.2410.03704.

Wilson, S.J., Freeman, B., Hedberg, E.C., 2025. Empirical Benchmarks for Effect Size Interpretation and Study Planning with Social and Behavioral Outcomes. Journal of Research on Educational Effectiveness , 1–26doi:10.1080/19345747.2024.2427767.

World Health Organization, 1993. Life Skills Education for Children and Adolescents in Schools. Technical Report WHO/MNH/PSF/93.7A.Rev.2. World Health Organization, Division of Mental Health. Geneva.

World Health Organization, 2003. Skills for Health: Skills-Based Health Education Including Life Skills. WHO Information Series on School Health 9. World Health Organization. Geneva.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., Wager, S., 2025. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects. Journal of the American Statistical Association 120, 38–51. doi:10.1080/01621459.2024.2393466.

Online Appendices

A Additional Tables and Figures

Table A1: *Balance checks (non-attrited sample)*

	Number of observations		Balancing check	
	Control	Treatment	Control mean	Difference
	(1)	(2)	(3)	(4)
Panel A: Socioemotional skills				
Interpersonal skills	2331	1681	-0.353	-0.012
			[1.21]	(0.046)
Emotion perception	2331	1681	-0.276	-0.010
			[0.946]	(0.036)
Intrapersonal skills	2331	1681	1.91	-0.002
			[0.793]	(0.026)
Perceived control	2331	1681	1.29	0.002
			[0.675]	(0.022)
Self-regulation and discipline	2331	1681	2.08	-0.006
			[0.996]	(0.035)
Joint F-test (p-value)				0.991
Panel B: Academic skills and creativity				
Learning (stacked test scores)	6099	4427	-0.576	0.040
			[1.01]	(0.036)
Arabic	1160	840	-0.585	0.087*
			[1.05]	(0.049)
French	1171	841	-0.488	0.037
			[1.01]	(0.065)
Math	1904	1362	-0.349	0.047
			[1.08]	(0.046)
Science	1864	1384	-0.857	0.033
			[0.801]	(0.035)
Physics and Chemistry	1864	1384	-0.448	0.002
			[0.782]	(0.032)
Life Science	1864	1384	-1.26	0.060
			[0.947]	(0.041)
Creativity	2331	1681	-0.328	0.040
			[1.03]	(0.055)
Joint F-test (p-value)				0.903
Panel C: Student characteristics				
Age	6099	4427	12.61	-0.033
			[1.02]	(0.022)
Primary school passing exam score	6099	4427	7.09	0.042
			[1.12]	(0.036)
% Female	6099	4427	49.37	0.204
			[50.00]	(0.861)
Joint F-test (p-value)				0.710

Notes. This table describes the study's non-attriting sample of 10,526 students and presents balance checks. "Control" refers to 116 control schools. "Treatment" refers to the 84 treatment schools. "Difference" reports on the regression-adjusted difference between treatment schools and control schools, controlling for grade and for randomization-group fixed effects, in column (4). Reversed outcomes were flipped so higher scores represent desirable outcomes. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level. Main family measures are highlighted in bold font. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: *Program exposure to at-risk students*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Program exposure		
Student participated in socioemotional workshops	0.231	0.407***
	[0.422]	(0.041)
Knows social specialist	0.431	0.135***
	[0.496]	(0.045)
Panel B: Exposure to other interventions		
Student received TaRL	0.913	0.004
	[0.281]	(0.016)
Student received explicit teaching	0.859	0.014
	[0.348]	(0.033)
Students participation in extra-curricular activities	0.575	0.063
	[0.495]	(0.051)
Student participated in tutoring program	0.684	0.042
	[0.465]	(0.047)

Notes. This table describes the program's exposure to students at the risk of dropping out in the 200 study schools among the non-attriting sample of students. Column (1) reports the control group mean among at-risk students, and column (2) reports the ITT effect for at-risk students, estimated in the full sample as the sum of the treatment coefficient and the treatment-by-at-risk interaction. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: *Program exposure to female students*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Program exposure		
Student participated in socioemotional workshops	0.217	0.443***
	[0.413]	(0.027)
Knows social specialist	0.466	0.190***
	[0.499]	(0.037)
Panel B: Exposure to other interventions		
Student received TaRL	0.962	0.001
	[0.192]	(0.011)
Student received explicit teaching	0.877	0.019
	[0.328]	(0.017)
Students participation in extra-curricular activities	0.661	0.008
	[0.474]	(0.025)
Student participated in tutoring program	0.469	-0.018
	[0.499]	(0.023)
At-risk student participated in tutoring program	0.678	-0.061
	[0.470]	(0.081)

Notes. This table describes the program's exposure to female students in the 200 study schools among the non-attriting sample of students. Column (1) reports the control group mean among female students, and column (2) reports the ITT effect for female students, estimated in the full sample as the sum of the treatment coefficient and the treatment-by-female interaction. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: *Effect on subdomains of socioemotional skills and creativity*

	Control mean	ITT Effect
	(1)	(2)
Interpersonal skills	-0.000	0.029
	[1.00]	(0.030)
		{1.000}
Pro-sociality	-0.000	-0.007
	[1.00]	(0.031)
		{1.000}
Emotion perception	-0.000	0.058**
	[1.00]	(0.029)
		{0.218}
Intrapersonal skills	-0.000	0.012
	[1.00]	(0.029)
		{1.000}
Perceived control	-0.000	0.045
	[1.00]	(0.030)
		{0.762}
Growth mindset	0.000	0.042
	[1.00]	(0.030)
		{0.480}
Locus of control	-0.000	0.035
	[1.00]	(0.034)
		{0.480}
Self-efficacy	-0.000	0.057**
	[1.00]	(0.028)
		{0.435}
Self-regulation and discipline	-0.000	-0.027
	[1.00]	(0.027)
		{0.762}
Work discipline and grit	-0.001	-0.007
	[1.00]	(0.026)
		{0.947}
Self-control	0.000	-0.097***
	[1.00]	(0.027)
		{0.003}
Creativity	0.000	-0.079*
	[1.00]	(0.045)
		{1.000}

Notes. This table reports on the program's intent-to-treat (ITT) effects following equation (1) on students' socioemotional skills and creativity among the 4,012 non-attriting students across the 200 government lower secondary schools who sat the written exams in Arabic or French. Column (1) reports the mean for the control group, and column (2) reports the ITT effect. "Prosociality" was not assessed at baseline, hence we do not include the baseline control when estimating the effect on this variable. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level; *q*-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT), and following Vivalt et al. (2024). Main family measures are highlighted in bold font. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table A5: *Intent-to-treat effects on subdomains of learning*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Arabic		
Written	0.000	0.029
	[1.00]	(0.034)
		{0.766}
Reading and Comprehension	0.000	-0.009
	[1.00]	(0.034)
		{1.000}
Written Production	0.000	0.124***
	[1.00]	(0.043)
		{0.044}
Oral	-0.000	0.118***
	[1.00]	(0.044)
		{0.044}
Panel B: French		
Written	-0.001	0.081*
	[1.00]	(0.042)
		{0.394}
Reading and Comprehension	-0.001	0.049
	[1.00]	(0.039)
		{0.567}
Written Production	-0.001	0.132***
	[1.00]	(0.049)
		{0.095}
Oral	-0.001	0.047
	[1.00]	(0.044)
		{0.567}
Panel C: Math		
Below-level	-0.001	-0.058*
	[1.00]	(0.032)
		{0.275}
At-level	-0.001	-0.058*
	[1.00]	(0.032)
		{0.275}
Applying or Reasoning	-0.001	-0.074**
	[1.00]	(0.033)
		{0.253}
Knowing	0.001	0.029
	[0.999]	(0.034)
		{0.576}
Algebra	0.001	-0.048
	[0.999]	(0.040)
		{0.510}
Data and Chance	0.000	-0.110**
	[1.00]	(0.044)
		{0.253}
Geometry	-0.000	-0.041
	[1.00]	(0.047)
		{0.576}
Number	-0.001	-0.039
	[1.00]	(0.043)
		{0.576}
Panel D: Science		
Applying or Reasoning	0.001	0.081*
	[1.00]	(0.044)
		{0.138}
Knowing	0.002	0.084*
	[0.999]	(0.045)
		{0.138}
Biology	0.001	0.095**
	[0.999]	(0.042)
		{0.085}
Physics	0.002	0.148***
	[1.00]	(0.042)
		{0.006}
Chemistry	0.002	0.135***
	[0.999]	(0.041)
		{0.007}

Notes. This table reports on the program's intent-to-treat (ITT) effects of the program on student learning among the 10,526 non-attriting students across the 200 government lower secondary schools following equation (1). Column (1) reports the mean of the control group, and column (2) reports the ITT effect. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT), and following Vivalt et al. (2024). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table A6: *Subject-wise subgroup effects on learning*

	Control mean	ITT Effect
	(1)	(2)
Panel A: At-risk students		
Arabic	-0.752	0.015
	[0.838]	(0.078)
		{1.000}
French	-0.651	-0.005
	[0.655]	(0.084)
		{1.000}
Math	-0.550	-0.016
	[0.789]	(0.082)
		{1.000}
Science	-0.553	0.037
	[0.732]	(0.079)
		{1.000}
Panel B: Female students		
Arabic	0.233	0.018
	[0.957]	(0.046)
		{1.000}
French	0.161	0.110**
	[0.970]	(0.047)
		{0.495}
Math	0.115	0.003
	[0.991]	(0.052)
		{1.000}
Science	0.191	0.091
	[1.04]	(0.060)
		{1.000}

Notes. This table reports on the program's intent-to-treat (ITT) effects of the program on student learning among the non-attriting students across the 200 government lower secondary schools following equation (1). Panel A reports the effects for the subgroup of students who are risk of dropout and Panel B reports the effects for female students. Column (1) reports the mean of the control group, and column (2) reports the ITT effect. Standard deviations are shown in brackets; Standard errors are shown in parentheses and are clustered at the school level; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT), and following Vivalt et al. (2024). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

Table A7: *Robustness of intent-to-treat effects on student dropout to alternative sample definitions*

	Control mean	ITT Effect
	(1)	(2)
Panel A: Students enrolled at the start of July		
Dropout by end of school year (global: either excluded or dropout)	0.035	0.002
	[0.183]	(0.003)
Dropout by end of the school year	0.030	0.002
	[0.171]	(0.002)
Repeated at end of school year	0.105	-0.017***
	[0.307]	(0.006)
Not re-enrolled in next school year	0.032	0.008**
	[0.176]	(0.004)
Excluded by end of school year	0.004	-0.000
	[0.067]	(0.001)
Panel B: Students enrolled at the start of August		
Dropout by end of school year (global: either excluded or dropout)	0.027	0.001
	[0.161]	(0.002)
Dropout by end of the school year	0.024	0.001
	[0.152]	(0.002)
Repeated at end of school year	0.097	-0.017***
	[0.296]	(0.006)
Not re-enrolled in next school year	0.027	0.007*
	[0.161]	(0.004)
Excluded by end of school year	0.003	-0.000
	[0.053]	(0.001)
Panel C: Students enrolled at the start of October		
Dropout by end of school year (global: either excluded or dropout)	0.025	0.001
	[0.156]	(0.002)
Dropout by end of the school year	0.023	0.002
	[0.148]	(0.002)
Repeated at end of school year	0.094	-0.019***
	[0.291]	(0.006)
Not re-enrolled in next school year	0.027	0.007*
	[0.161]	(0.004)
Excluded by end of school year	0.003	-0.000
	[0.050]	(0.001)

Notes. This table reports on the program's intent-to-treat (ITT) effects following equation (1). Column (1) reports the mean for the control group and column (2) reports the ITT effect. Panel A reports the effect of the program on students who were reported as enrolled at the start of July (when the previous academic year ends). Panel B reports the effect of the program on students who were reported as enrolled at the start of August (when the new academic year starts). Panel C reports the effect of the program on students who were reported as enrolled at the start of October (one month into the new academic year). Standard deviations are shown in brackets. Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: *Effect on socioemotional skills: Robustness to social desirability bias*

		Social desirability	
	Difference	Below median	Above median
	(1)	(2)	(3)
Interpersonal skills	0.022	-0.007	0.015
	(0.069)	(0.039)	(0.054)
		{1.000}	{1.000}
Pro-sociality	-0.028	-0.004	-0.032
	(0.061)	(0.040)	(0.049)
		{1.000}	{1.000}
Emotion perception	0.097	0.007	0.104**
	(0.067)	(0.039)	(0.052)
		{1.000}	{0.217}
Intrapersonal skills	-0.056	0.029	-0.027
	(0.062)	(0.036)	(0.051)
		{1.000}	{1.000}
Perceived control	-0.022	0.042	0.021
	(0.060)	(0.036)	(0.050)
		{1.000}	{1.000}
Self-regulation and discipline	-0.084	0.011	-0.073
	(0.062)	(0.035)	(0.049)
		{1.000}	{1.000}

Notes. This table reports on the heterogeneity in the program's intent-to-treat (ITT) effects following equation (1) on students' socioemotional skills among the 4,012 non-attriting students across the 200 government lower secondary schools who have non-missing social desirability scores at baseline and who sat the written exams in Arabic or French. Column (1) reports the difference in the effect on students who exhibit high social desirability bias and students who exhibit low social desirability bias, and columns (2) and (3) respectively report the effects separately for the two subsamples of students. "Prosociality" was not assessed at baseline, hence we do not include the baseline control when estimating the effect on this variable. Standard errors are shown in parentheses and are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: *Sensitivity to differential attrition*

	Interpersonal skills	Intrapersonal skills	Learning (stacked test scores)
	(1)	(2)	(3)
Panel A: Lee (2009) bounds			
Lower bound			
Treatment	0.020	-0.010	0.030
	(0.030)	(0.030)	(0.030)
	{0.764}	{1.000}	{1.000}
Upper bound			
Treatment	0.020	0.020	0.030
	(0.030)	(0.030)	(0.030)
	{0.777}	{1.000}	{1.000}
Panel B: Behaghel et al. (2015) bounds: Attrition based on number of days tracked			
Lower bound			
Treatment	0.020	0.010	0.030
	(0.030)	(0.030)	(0.030)
	{0.786}	{1.000}	{1.000}
Upper bound			
Treatment	0.020	0.010	0.040
	(0.030)	(0.030)	(0.030)
	{0.786}	{1.000}	{1.000}
Panel C: IPW - Attrition based on baseline covariates			
Treatment	0.030	0.000	0.030
	(0.030)	(0.030)	(0.030)
	{0.407}	{1.000}	{1.000}

Notes. Estimated treatment effects. The dependent variables in columns (1), (2), and (3) are the family indices from Tables 4 and 5. Treatment captures effects for schools which received the treatment. Panel A estimates bounds for the ITT estimates following Lee (2009). Panel B estimates tightened bounds for the ITT estimates following Behaghel et al. (2015). Panel C estimates an IPW estimate for the ITT effect using baseline covariates as predictors of attrition. In panel C, the probability of attrition is estimated through a probit regression of an attrition indicator on treatment status and baseline covariates. "Number of days tracked" information includes the number of days visits to a school before a student was surveyed. All regressions follow the specifications in tables 4 and 5 except for the weights. Standard errors are clustered at the school level and are shown in parentheses; q-values are shown in curly brackets, computed using the Anderson (2008) correction for multiple hypothesis testing (MHT), and following Vivalt et al. (2024). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

References

Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association 103, 1481–1495. doi:10.1198/016214508000000841.

Behaghel, L., Crépon, B., Gurgand, M., Le Barbanchon, T., 2015. Please call again: Correcting nonresponse bias in treatment effect models. Review of Economics and Statistics 97, 1070–1080.

Lee, D.S., 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. The Review of Economic Studies 76, 1071–1102. doi:10.1111/j.1467-937X.2009.00536.x.

Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. doi:10.3386/w32719.

B Measurement

IRT models are used to estimate the probability of answering an item correctly as a function of a latent parameter representing the student's ability and of parameters relating to the item (Jacob and Rothstein, 2016). To aggregate responses to test and interview questions ("items"), we estimate different types of item response theory (IRT) models, each selected to align with the structure and characteristics of a given instrument.

This appendix briefly describes each of the models used to aggregate our items, along with the methods used to further aggregate IRT scores into indices. Appendix Table B1 reports the resulting psychometric properties of each assessment and survey instrument, including the number of items, mean item discrimination and difficulty, and the endline average conditional reliability. To estimate students' ability scores, each of our IRT methods uses expected a posteriori (EAP) scoring. Soland et al. (2024) show that, for EAP scoring, failure to account for between-group variation in ability distributions may lead to biased estimates of treatment effects in educational interventions. To avoid such bias, we follow the guidance in Soland et al. (2024) and estimate scores using a two-group IRT model. Thus, the treatment and control groups are allowed to have separate ability distributions (i.e., different means and variances) while assuming common item parameters.

B.1 Two-Parameter Logistic (2PL) Model

The two-parameter logistic (2PL) model estimates the probability of answering an item correctly based on the examinee's ability (θ) and each item's difficulty (b_i). This type of model is only applicable to instruments with strictly binary items (including test questions rated as correct vs. incorrect), and, therefore, we apply it to the language, math, and science tests, along with the perceived emotions instrument (PAGE).

The two-parameter logistic (2PL) model is given by:

$$P(Y_{ij} = 1 \mid \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \tag{2}$$

Where θ_j is the latent ability of student j, a_i is the discrimination parameter, and b_i is the difficulty of item i.

B.2 Graded Response Model (GRM)

For Likert-type items with ordered categorical responses (e.g., 1 = Never to 5 = Always), we use the Graded Response Model (GRM), appropriate for instruments such as *SDQ (pro-sociality)*.

The probability of responding in category k or higher is given by:

$$P(Y_{ij} \geq k \mid \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_{ik})]}$$

where b_{ik} represents the threshold between category $k-1$ and k. This model captures both item sensitivity (via a_i) and the varying difficulty of response thresholds.

We also pre-registered the Partial Credit Model (PCM) out of concern that the Graded Response Model could not accommodate items with varying or unequally spaced ordinal categories. On implementation, the GRM handles these directly (and estimates an item-discrimination parameter the PCM omits), so we scored all indices that involve ordinal response categories with the GRM.

B.3 Poisson Principal Component Analysis (PCA)

To construct the Creativity Index, we use a Poisson principal component analysis (Poisson PCA). Unlike standard PCA, which assumes continuous variables, Poisson PCA models the discrete, non-negative nature of count data more accurately. This approach aligns with the creativity measure, which captures the number of responses students produce.

B.4 Aggregation of Latent Scores

Latent scores ($\hat{\theta}_{j1}, \hat{\theta}_{j2}, \ldots, \hat{\theta}_{jn}$) are estimated using the empirical Bayes method, which computes scores based on students' observed item responses and the estimated item parameters. To account for potential bias that can be introduced by shrinkage, we follow Soland et al. (2024) and estimate latent traits using a two-group IRT model. This model allows the treatment and control groups to have separate ability distributions while holding item parameters fixed.

To compare scores across baseline and endline assessments, we use a standard IRT linking procedure. Specifically, we estimate item parameters from the control group at endline and then hold those parameters fixed when scoring the baseline data. This approach ensures scores are placed on a common scale and changes over time are interpretable.

After estimating individual-level latent scores $(\hat{\theta}_{j1}, \hat{\theta}_{j2}, \ldots, \hat{\theta}_{jn})$ from each IRT model, we construct composite indices taking the inverse covariance-weighted average of component scores. Specifically, we compute the inverse covariance matrix of the component scores, normalize these weights to sum to one, and use them to compute a weighted average:

$$\text{Index}_j = \sum_{i=1}^{n} w_i \hat{\theta}_{ji}, \quad \text{where } w_i \propto \frac{1}{\text{Var}(\hat{\theta}_i)}$$

Table B1: *Psychometric properties*

	Number of items			Classical test theory (CTT)				Item response theory (IRT)		
	Effective	Excluded	Anchors	% correct	% NA	Cronbach's alpha	Mean item-test corr.	Mean discrimination	Mean difficulty	Avg. conditional reliability (endline)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Assessments										
Arabic	75	1	52	45.31	0.056			1.16	0.191	0.897
French	56	1	36	25.12	4.56			1.22	1.01	0.889
Math	32	0	11	27.93	7.96	0.787	0.369	0.888	1.18	0.801
Science	34	0	10	35.74	2.79	0.765	0.339	0.744	1.02	0.781
Physics and Chemistry	19	0	5	37.18	3.37	0.631	0.362	0.740	1.35	0.667
Life Science	15	0	5	41.14	2.04	0.639	0.408	0.848	0.550	0.652
Panel B: Other measures										
Interpersonal skills										
Pro-sociality	5	5						1.34		0.599
Emotion perception	16	1						0.529		0.505
Intrapersonal skills										
Perceived control	12	0						0.840		0.723
Growth mindset	3	0						1.91		0.720
Locus of control	4	0						1.09		0.434
Self-efficacy	5	0						1.18		0.627
Self-regulation and discipline	25	0						0.886		0.851
Work discipline and grit	17	0						0.998		0.821
Self-control	8	0						0.959		0.677
Creativity	7	0								0.791
Well-being										
Well-being index	18	0						0.995		0.737
Feeling of belonging in school	5	0						1.15		0.607
Bullying (reversed)	9	0						1.54		0.676
Perceived stress (reversed)	4	0						0.886		0.458

Notes. Sample and unit of observation: 10,514 assessed students across the 200 schools in the study. This table reports on the measurement properties of the student assessment instruments and survey instruments included in the study's endline. "Anchors" refers to items administered at both baseline and endline, used to link the two waves onto a common scale. "NA" refers to non-response. Mean discrimination and mean difficulty refer to the mean discrimination and difficulty parameters from the IRT model used for each instrument (a two-parameter logistic model for the assessments and a graded-response model for the Likert survey scales). Average conditional reliability represents the mean of individual-level reliabilities across the endline sample, based on each respondent's estimated ability and associated predicted standard error.

References

Jacob, B., Rothstein, J., 2016. The Measurement of Student Ability in Modern Assessment Systems. Journal of Economic Perspectives 30, 85–108. doi:10.1257/jep.30.3.85.

Soland, J., Kuhfeld, M., Edwards, K., 2024. How survey scoring decisions can influence your studys results: A trip through the IRT looking glass. Psychological Methods 29, 1003–1024. doi:10.1037/met0000506.

C Prespecified Adjustments for Multiple Hypothesis Testing

Accounting for a pre-registered hierarchy of research hypotheses, we adjust for multiple hypothesis testing by computing the sharpened false discovery rate adjusted q-values. Following Vivalt et al. (2024), we place our hypotheses into tiers (denoted K0, K1, and K2), which correspond to our prioritization of tests.

K0: Family of main outcomes: Intra- and interpersonal skills. We compute q-values for these two variables (sets of skills), which we refer to as our two K0 items. These are our highest priority outcomes.

K1: Family of final outcomes: Academic skills and dropout. These are our second-highest priority outcomes. We compute the q-values for these two variables (outcomes), which we refer to as our two K1 items, in combination with our two K0 items.

K2: Family of additional outcomes: Creativity, well-being, and study habits. These are our third-highest priority outcomes. We compute the q-values for these three variables (outcomes), which we refer to as our three K2 items, in combination with our four K0 and K1 items.

In addition to these three tiers of family-level tests, we also explore effects at the underlying component- and sub-component levels. For example, the "family" of intrapersonal skills consists of two "components": an index of perceived control, and an index of self-discipline. In turn, the perceived control component consists of three sub-components: a measure of growth mindset, a measure of locus of control, and a measure of perceived self-efficacy. To provide another example, the "family" of academic skills includes four subject-wise components (Arabic, French, mathematics, science), which can be further subdivided into sub-components (e.g., two indices of math items capturing at-grade-level-content vs. below-grade-level content, respectively). The q-values for component-level estimates are computed using the family-level items up to the given family level (K0, K1, or K2), in addition to the component-level items in the outcome's same family. For example, the q-values for effects on Arabic skills account for the four family-level comparisons at the K0 and K1 levels, as well as the four component-level comparisons (Arabic, French, math, and science).

Within each family tier, and at each level (family, component, or sub-component), we explore treatment effects among two subgroups: students identified as being at risk of dropping out, and girls. Across these two subgroups, we prioritize the tests for the subgroup of at-risk students. The q-values for these comparisons are computed using all

the items for that family, and up to the given level, in addition to the tests of heterogeneous effects. For example, tests for effects on Arabic skills among students at risk of dropping out account for the four comparisons at the K0 and K1 levels, the four component-level comparisons (Arabic, French, math, and science), as well as four additional subgroup tests (one per subject). Tests for effects on Arabic skills among girls account for the four comparisons at the K0 and K1 levels, the four component-level comparisons (Arabic, French, math, and science), the four subgroup tests for students at risk of dropping out, as well as four additional subgroup tests (one per subject).

Our analyses of intervention take-up and implementation fidelity are largely based on descriptive statistics and do not account for these analyses in our adjustments for multiple hypothesis testing.

References

Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. doi:10.3386/w32719.