Title page with author details

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||

October 16, 2026

*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.

Revised manuscript without author details (marked)
Click here to access/download;Revised manuscript without author details (marked);Manuscript No Author Details

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

October 16, 2026

Abstract

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 of which were participating in the first year of a broader reform, 84 were assigned to the intervention and 116 to a control group. Assignment to the intervention 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. Grade repetition fell by 1.7 percentage points, and achievement rose in specific science and language subdomains. All of these estimates survive correction for multiple hypothesis

testing with the pre-specified school-level clustering; yet with the larger standard errors from clustering by randomization stratum, only the self-control decline and the physics and chemistry subdomains reach the 10-percent threshold, so we treat the improvements as suggestive. 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.

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

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

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 to strengthen 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 or non-governmental organizations. Much less is known about whether governments can use existing personnel to deliver socioemotional support inside public 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 implemented in those schools. It trained and equipped existing school social specialists to deliver four workshops per year, each lasting about an hour, 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 of which were 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 intervention 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.

More positively, improvements appear in two sets of secondary outcomes: grade repetition fell by 1.7 percentage points, and student achievement rose in specific subjects and subdomains, including physics, chemistry, and biology, written expression in Arabic and French, and oral expression in Arabic. With standard errors clustered at the school level, as pre-specified, these estimates survive correction for multiple hypothesis testing. When standard errors are instead clustered at the level of the 84 randomization strata, as de Chaisemartin and Ramirez-Cuellar (2024) recommend for experiments with small strata, the coefficients are essentially unchanged but the standard errors are larger, so that none of these estimates survives at the 5-percent level and only the physics and chemistry subdomain estimates reach the 10-percent threshold. We therefore treat these improvements as suggestive. The decline in self-control survives correction under both levels of clustering.

Section 6 examines why the added contact did not translate into the targeted skills. Beyond workshop participation, the intervention increased individual meetings with the social specialist, by 6.1 percentage points. Take-up varied across schools, but effects on the targeted skills are not significantly larger in the treatment schools where it was high. The null estimates are precise at the intent-to-treat level: they rule out the average effects of universal socioemotional programs in a recent meta-analysis (Cipriano et al., 2023). By design, the dose (four workshops over the year) was low by the standard of those programs, and individual follow-up rose no more for at-risk than for not-at-risk students.

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 sparse and often comes from researcher- or NGO-delivered programs (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 by existing 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, Section 6 examines why the intervention did not improve the targeted skills, and Section 7 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 in recent years, as of 2019, public schools still served 83 percent of primary school students and 89 percent of lower secondary school students in Morocco (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 (World Bank, 2025).

From 2018 to 2023, the lower secondary completion rate increased from 64.5 to 74.2 percent (World Bank, 2026). 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).

2.2 The Pioneer School Program

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 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 scaled the program to another 554 schools in the 2025-26 school year, covering approximately 32 percent of Morocco's lower secondary students in public schools. The remaining schools are expected to be reached by 2028.

The reform bundles several components, and the socioemotional support intervention we evaluate is one of them. All 200 schools in our study received the reform's other components, including targeted remediation, explicit teaching, tutoring, and extracurricular activities. The socioemotional component differs only in that its rollout was partial: the Ministry decided that only 84 program schools would receive it in the first year of implementation, and we randomly assigned which of the 200 study schools those would be, leaving 116 as a control group (see subsection 3.2). Our design therefore identifies the marginal effect of adding socioemotional support to schools already implementing the rest of the reform, not the effect of the reform as a whole. Panel B of Table 3 confirms that students in the two groups report similar exposure to the other components; Appendix

[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 the PSP program's launch 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 Ibrahim et al. (2026) and de Barros et al. (2026).

Tables A2 and A3 report the same checks for students at risk of dropout and for female students. Subsection 2.3 just below describes the socioemotional support intervention.

2.3 The Socioemotional Support Intervention

The intervention's diagnostic phase drew on the socio-ecological model (Bronfenbrenner, 1977) 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 consisted of training and equipping social specialists to give four 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 aimed 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 were 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. Without the intervention, this person's role varies from school to school: specialists look after students' well-being and refer cases to the school director (principal), but many are assigned administrative or disciplinary tasks, and their responsibilities, tools, and training are poorly defined (subsection 2.4). Specialists may also run workshops and individual interviews on their own initiative.

The training followed a cascade. The Behavioural Insights Team (BIT), a social purpose organisation that designed the intervention with the Ministry, trained twelve social specialists who served as regional focal points, four of whom worked in study schools; these focal-point specialists then trained 80 other social specialists, so that 84 specialists in

[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. Each social specialist is assigned to a single lower-secondary school.

all were trained to deliver the workshops. The training of the specialists lasted five days. The first day covered facilitation and classroom management: how to arrange the room, agree on ground rules with students, keep a session interactive, and respond to a student in distress. Each of the four remaining days prepared one workshop: participants read the facilitator guide for each activity, one of them ran the activity as if with students, and the group discussed the simulation. Social specialists received training slides and detailed session guides, and students received booklets for each of the four workshops, along with a notebook to record their feelings and emotions.

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

The facilitator guide set a common structure for every workshop: an introduction and an icebreaker, two main parts, and a closing discussion in which students shared what they had learned. The activities alternated individual reflection, pair work, small-group work, and whole-class discussion. The training asked specialists to act as facilitators, not as teachers giving a lesson: to seat students in a circle where possible, to encourage every student to speak without forcing anyone, and to rely on discussion and role-play instead of lecturing or testing. In the first workshop, for example, students listed their own strengths in their booklets and discussed them in pairs, then repeated a short drawing task to see how practice improves performance; in the second, students took turns recounting an experience to a partner, who paraphrased it back.

Each session was held with half a class, about 15 students and never more than 20, and lasted about an hour. Sessions took place during regular school hours without replacing any class, and each combined the content set out in the facilitator guide with time for open discussion. The first workshops took place in mid-November 2024.

The workshops were also meant to be the entry point to individual support. The training asked specialists to watch during sessions for students who withdrew from their peers, could not name any of their strengths, voiced negative beliefs about themselves, or behaved in unusual ways, and to check in with such students privately at the end of the session. Individual follow-up by the specialist was meant to complement the work of each school's monitoring committee (the *cellule de veille*, a group of school staff that identifies students at risk of dropping out and follows them through individual action plans). Specialists raised the students they identified at the committee's meetings but were not asked to record their names.

We observe the delivery of the intervention almost entirely through students' reports. The process-monitoring visits specified in the Stage 1 plan were not conducted (Appendix B). The endline interview records whether a student attended a workshop, and two short monitoring surveys that we fielded during the school year record how many workshops students had attended. Subsection 6.1 reports what these data show.

2.4 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 exist (in this case, social specialists), existing services that could promote socioemotional skills appear ineffective, with limited support for students, support relying solely on students' voluntary participation, social specialists being assigned to administrative tasks (or only to the most serious cases, typically disciplinary), a lack of clarity about responsibilities, and an absence of tools and training. Taken together, the study context reveals 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 subsection), following a government order for the specialists to implement it. 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 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 desired downstream outcomes.

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 (including

growth mindset, locus of control, and perceived self-efficacy) and students' self-regulation and discipline (including 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. Improvements in socioemotional skills are also goals in their own right, alongside this impact.

3 Study Design

As a Stage 2 Registered Report, our study closely follows its pre-registered design and analysis plan. Appendix B lists the deviations from, and corrections to, that plan, most of which are corrections to the Stage 1 document or additions that extend the registered analysis.

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 public 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 from the administrative data 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 representativeness of this sample of 200 PSP schools, both with respect to the population of all lower secondary schools in Morocco and to other PSP schools in that 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

[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 Program" 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 stage-1 review and registration of the study design, we adjusted our plan because the Ministry decided at the beginning of the study that only 84 PSP schools would receive the support intervention in the first year of implementation. The following subsection describes our random assignment of 84 PSP schools to the intervention, which was done as pre-specified.

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

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, the 200 PSP schools in the study sample have slightly more teachers, are less likely to be rural, have more students, and serve students with slightly higher average primary and middle-school school leaving exam scores, compared to the 32 non-study PSP schools.

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 lower secondary 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 strata in total). Finally, within each stratum, we randomly assigned one school to receive the socioemotional support intervention.

Columns (7) to (9) of Table 1 present the balance among these two groups of schools included in this study. School characteristics are balanced between treatment and control schools, 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 we subsampled

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

up to 12 students per school (6 students from grade 7, and 6 students from grade 8).[7] For the written math and science assessments, we subsampled up to 18 students per school (all attending grade 7).[8] Students in the Arabic and French subsample also sat a one-on-one interview session, which combined the oral language assessment with the instruments used to measure socioemotional skills, creativity, school climate and well-being, and study habits. Students in the mathematics and science subsample sat written assessments only.

Specifically, students were randomly sampled for each subject within each school, grade, and track. Mathematics and science were administered only to students in the international (APIC) track, because the general (ASCG) track studies these subjects in Arabic rather than in the language of our assessments (French); Arabic and French were administered in both tracks, with the number of students drawn from each track proportional to its share of enrollment.[9] Within each subject group, a random priority list of students was generated, and enumerators worked from the top of the list ("high" priority) toward the bottom ("low" priority), drawing on the lower-ranked students as replacements for any who were absent, so that each grade-by-subject group reached its target size.

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, at baseline, 2,138 were tested in Arabic, 2,148 in French, 3,428 in math, and 3,418 in science. The analysis samples therefore differ by outcome. Outcomes measured in the one-on-one interview session are estimated on the Arabic and French subsample, 4,286 students at baseline and 4,012 at endline. Learning outcomes draw on all four subjects, or 10,526 non-attriting students. School-progression outcomes are constructed from Ministry enrollment records and cover all 128,055 students enrolled in grades 7 and 8 of the study schools.

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 very similar to the students in the treatment

[7] We also sampled and tested up to 6 students in grade 9, but since the socioemotional skill intervention focuses on grades 7 and 8 only, we exclude them from our data. We also specified this decision during the stage-1 review of our pre-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] Lower-secondary students are streamed into a general track (ASCG) and an international track (APIC), which differ mainly in language of instruction. The international track, which covers the large majority (roughly 95 percent) of students in these schools, studies mathematics and science in French; the general track studies them in Arabic.

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 (subsection 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. Panel C of Table 2 reports the balance test in the row labeled "% Attrited": attrition was 1.0 percentage points lower in the treatment group than in the control group, against a control-group rate of 5.9 percent, and the difference is significant at the 10-percent level. Subsection 5.2 shows that our estimates are robust to this differential attrition.

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. Our main learning outcome stacks standardized scores across all four subjects. School-progression outcomes are measured using administrative records and include a summary measure capturing dropout

[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.

[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).

or exclusion by the end of the school year; we also report separately on students' non-re-enrollment at the beginning of the following school year, repetition, and exclusion.[12] These records cover every student in the study schools, not only the subsample of students who participated in the assessments. The school records sample was not foreseen in the Stage 1 plan, which registered dropout only for students who took the baseline assessment; repetition and exclusion were likewise not registered as separate outcomes. Appendix B records both departures and explains why we use the full records.

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. Further measures record whether students named the specialist correctly, met the specialist individually during the school year, and received help from the specialist with psychological or academic difficulties. We also measure exposure to the other components of the Pioneer School Program (PSP), including targeted remediation, explicit teaching, and tutoring.

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 C provides details on measurement and scoring.[13]

Appendix Table C1 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, which is plausibly random, inflates standard errors but does not bias the treatment-effect estimates for any given outcome.

[12] Here, we maintain the Ministry's standard language for referring to students' end-of-year enrollment status. "Exclusion" refers to students who failed and are no longer permitted to repeat, typically because they have exhausted their allowed number of repetitions.

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

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.[14] Subsection 6.1 discusses the other measures of contact with the specialist in Panel A.

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.[15]

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 (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 that 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. 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

[14] Panel A of Appendix 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.

[15] Panel B of Appendix 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.

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).[16] As pre-specified, we cluster standard errors at the school level (Abadie et al., 2023); to check robustness, subsection 5.2 reports inference with standard errors clustered at the level of the randomization strata (as recommended by de Chaisemartin and Ramirez-Cuellar, 2024).

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 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. (2026). 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) comprises creativity, students' perception of school climate and well-being, and study habits. Component- and subgroup-level tests are adjusted for multiple hypothesis testing within the relevant pre-specified testing families. Appendix D provides additional details on the pre-specified multiple hypothesis testing procedure.

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 loses statistical significance 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,

[16] 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).

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 ($q = 0.003$), including when standard errors are clustered by randomization stratum ($q = 0.044$; subsection 5.2). 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 multiple hypothesis testing 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, biology at the 10-percent level ($q = 0.085$), and physics and chemistry at the 1-percent level. With standard errors clustered by randomization stratum, these science estimates are less precise, and only the physics and chemistry subdomains reach the 10-percent threshold after adjustment (subsection 5.2).

Among the language subdomains (Appendix Table A5), the positive estimates concentrate on written production. Assignment to the intervention 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 (0.047 s.d.) and statistically insignificant, so we interpret the oral result more cautiously. None of these language estimates retains significance after adjustment when standard errors are clustered by randomization stratum (subsection 5.2). These estimates are for the skills that require students to produce language (writing and, in Arabic, speech) rather than for 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 ($q = 0.032$). It does not retain significance after adjustment when standard errors are clustered by randomization stratum ($q = 0.296$; subsection 5.2).

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 loses statistical significance 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 loses statistical significance 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 loses statistical significance after adjustment; the other estimates are small and statistically insignificant at conventional levels, and the estimates on additional outcomes are similarly small and insignificant.[17] Appendix Table A6 reports the subgroup effects on learning separately by subject.

5.2 Robustness Checks

For dropout and repetition outcomes, the sample of initially enrolled students can vary by the month in which enrollment is constructed; therefore, we assess the robustness of findings to alternative sample definitions. Table A7 shows that the estimated effects on dropout and repetition 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

[17] In addition to the above pre-registered investigation of the intervention's 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 subgroup that was affected by the intervention, in either direction.

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.

We then 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. Inverse-probability weighting leaves the estimates essentially unchanged, and the tightened bounds are narrow and statistically insignificant. The Lee bounds are the widest of the three, as they are worst-case bounds that use no auxiliary information on which students attrited. Because attrition was lower in the treatment group, their upper bound trims the lowest-scoring treated students; for interpersonal skills, that bound is statistically significant even after adjustment for multiple hypothesis testing. We take this as evidence that, more broadly, the results we present are not affected by differential attrition.

Finally, we examine how our inference depends on the level at which standard errors are clustered. Our main specification clusters them at the school level, as pre-specified. Treatment was assigned within 84 randomization strata of two or three schools, however, and de Chaisemartin and Ramirez-Cuellar (2024) show that in experiments with such small strata, standard errors clustered at the level of the randomized unit can be too small; they recommend clustering at the level of the strata. Appendix Tables A10 and A11 report our estimates for socioemotional skills, learning, dropout, and repetition under both levels of clustering, with q-values recomputed over the same pre-specified families. The coefficients barely change (by at most 0.015 s.d., because the lasso's penalty loadings also depend on the clustering), and the estimates for the primary outcomes remain statistically insignificant. The decline in self-control survives adjustment for multiple hypothesis testing under both levels of clustering ($q = 0.003$ and $q = 0.044$). The downstream improvements are estimated less precisely. Under clustering by randomization stratum, each remains statistically significant before adjustment (biology at the 10-percent level, the others at the 5-percent level), but after adjustment only the physics and chemistry subdomains reach the 10-percent threshold ($q = 0.100$). The q-values rise to 0.135 for physics and chemistry combined, 0.389 for biology, 0.296 for grade repetition, 0.316 for written production and oral expression in Arabic, and 0.946 for written production in French. We therefore treat these downstream improvements as suggestive.

6 Interpreting the Null Results

The intervention changed students' exposure to socioemotional support but did not detectably change the skills it targeted. This section follows the theory of change (subsection 2.4) from delivery to outcomes. Most of the analyses in this section were not specified in the Stage 1 plan, and Appendix B records them as additions. The null on the targeted skills is precise and holds even in the half of treatment schools where take-up was highest. The evidence points to the design of the intervention (a low dose, and individual follow-up that rose no more for at-risk than for not-at-risk students) rather than to weak delivery.

6.1 Delivery and Contact

Assignment to the intervention raised workshop participation by 43.2 percentage points and students' knowledge of the social specialist by 16.9 percentage points (subsection 3.5). The second measure, whether students know the specialist's name, does not rest on self-report alone: students who said they did were asked to give it, and enumerators recorded whether it was correct. Of these students, 87.5 percent gave the correct name, and assignment raised the share of all students who did so by 16.8 percentage points, from a control mean of 37.2 percent (Table 3, Panel A).

The intervention also increased individual contact with the specialist, which the workshops were designed to generate for students in need of follow-up. Assignment to the intervention raised the share of students reporting an individual meeting with the specialist during the school year by 6.1 percentage points, from a control mean of 24.8 percent, and the share reporting that the specialist had helped them with psychological difficulties by 2.4 percentage points, from 5.1 percent; these estimates are statistically significant at the 1-percent and 5-percent levels, respectively. The share reporting help from the specialist with academic difficulties rose by 1.5 percentage points, from 3.6 percent, an estimate significant at the 10-percent level. We collected these measures, like the other exposure measures, as measures of take-up and not as registered outcomes; they enter no family of the multiple-testing correction, and Appendix B records which of them the plan specified.

Two short monitoring surveys that we fielded in the study schools during the year asked students whether they had attended a workshop and, if so, how many.[18] In February 2025, two to three months after the first workshops, 62.3 percent of students in grades 7 and 8

[18] Each survey interviewed six randomly sampled students per grade in each study school it reached: 197 schools from 28 January to 15 February 2025, and 194 schools from 4 April to 22 May 2025. The figures in the text are for students in grades 7 and 8.

of treatment schools reported having attended at least one, against 7.1 percent in control schools. By April and May, the shares were 70.0 and 8.1 percent, and 63 percent of the treatment-school participants who gave a number reported having attended three or more workshops. In treatment schools, these rates agree with the endline interview, in which about two-thirds of students report participating in a workshop, and schools with higher take-up at endline also had higher take-up in the surveys (correlations of 0.57 and 0.50 across schools). In control schools, the endline interview gives a higher rate, 21.7 percent, than in the monitoring surveys. There, endline reports come disproportionately from students who had met the specialist individually, 41 percent of whom report participating in a workshop against 15 percent of the others, and from students who attended tutoring sessions, which suggests that some students counted other contact with the specialist, or other support sessions, as workshops.[19] Take-up nevertheless varied across schools. At endline, the share of surveyed students reporting workshop participation ranges from 15.0 to 100 percent across the 84 treatment schools, with a mean of 65.6 percent, and in 10 of them fewer than one student in three reports participating in a workshop.

If the null on the targeted skills reflected weak delivery alone, effects should be larger where delivery was stronger. Appendix Table A12 examines this descriptively. It splits the treatment schools at the median of realized take-up and estimates the intent-to-treat effect separately for each half, comparing each treatment school with the control schools in its randomization stratum. The two halves differ sharply in delivery: workshop participation was 63.2 percentage points higher in schools with high take-up and 23.4 percentage points higher in schools with low take-up, and individual meetings with the specialist were 11.7 percentage points higher in the former and no higher in the latter. The estimates for the two socioemotional indices, however, are small in both halves and do not differ significantly between them. The learning index was 0.10 s.d. higher in high-take-up schools and essentially no higher in low-take-up schools, a difference that is statistically significant ($p = 0.009$). Realized take-up is itself an outcome of assignment, and schools that delivered more workshops may differ from the others in other ways, so this comparison does not estimate a dose-response relationship. The null on the targeted skills is therefore not confined to schools where few students attended, which argues against weak delivery as its only explanation. Because take-up was not randomly assigned, we do not interpret the learning pattern as an effect of the workshops.

[19]The endline question asked about participation in "psychosocial support workshops" during the school year, and the Arabic word for "support" is the one the next question uses for the mathematics and French tutoring sessions that all study schools offer. The workshops do not appear to have spread to control schools: in only 6 of the 116 control schools did a majority of surveyed students report attending workshops in two or more of the three surveys.

6.2 Precision

At the intent-to-treat level, the null estimates are precise. The 95-percent confidence intervals are $[-0.029, 0.087]$ for interpersonal skills and $[-0.045, 0.070]$ for intrapersonal skills, and the ex post minimum detectable effect, at 80-percent power and a 5-percent significance level, is 0.083 s.d. for both indices. These intervals exclude the mean effects of universal school-based socioemotional programs in the meta-analysis of Cipriano et al. (2023), which are 0.22 s.d. (Hedges' g) for socioemotional skills and 0.19 s.d. across all outcome domains. The interval for the stacked learning index, $[-0.023, 0.091]$, likewise excludes their mean effect on academic achievement, 0.11 s.d. With standard errors clustered by randomization stratum, the intervals widen by about 30 percent, but their upper bounds for the two socioemotional indices, 0.105 and 0.085 s.d., remain below those averages.

Assignment to the intervention raised workshop participation by 43.2 percentage points, so if assignment affected skills only through participation, the effect on students induced to participate is 2.3 times the intent-to-treat effect. On that scale, the upper bounds of the confidence intervals are 0.20 s.d. for interpersonal skills and 0.16 s.d. for intrapersonal skills, against a meta-analytic mean of 0.22 s.d. for socioemotional skills.[20]

By design, the intervention's dose was low by the standard of those programs: four workshops of about an hour each over the school year, whereas the universal programs for early adolescents (grades 6 to 8) in Cipriano et al. (2023) averaged 19 sessions of about 51 minutes each. 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.

6.3 Reaching At-Risk Students

The intervention was meant to reach students at risk of dropping out in two ways: through the workshops, which included all other students, and through the specialist's referral of students in need to individual follow-up (subsection 2.3). The workshops reached at-risk students about as well as other students. Among at-risk students, assignment raised workshop participation by 40.7 percentage points and knowledge of the specialist by 13.5

[20]Compliance was two-sided: at endline, 21.7 percent of control students also report participating in socioemotional workshops, against 7 to 8 percent in the monitoring surveys (subsection 6.1). The scaled estimate therefore refers to compliers, the students who participated because their school was assigned to the intervention, not to all participants. If part of the endline control rate reflects other activities, the endline first stage understates the contrast in participation, and the scaled bounds are conservative: scaling instead by the difference of about 62 percentage points in the April and May monitoring survey gives upper bounds of 0.14 and 0.11 s.d.

percentage points (Appendix Table A2), against 43.2 and 16.9 percentage points in the full sample (Table 3). The share of at-risk students who could name the specialist correctly rose by only 6.7 percentage points, an estimate that is not statistically significant, against 16.8 percentage points in the full sample. Individual follow-up rose no more for them either: among at-risk students, individual meetings with the specialist rose by 3.1 percentage points, an estimate that is not statistically significant, against 6.1 percentage points in the full sample, although at-risk students in control schools already met the specialist more often (33.1 percent, against 24.8 percent in the full sample). The intervention thus did not differentially engage the students its targeted component was designed for, and none of its estimated effects on at-risk students survives correction for multiple hypothesis testing (Table 7). A higher dose of the same workshops need not change this: reaching at-risk students preferentially may require changing how they are identified and referred.

6.4 Measurement and the Downstream Estimates

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. We therefore interpret the findings as most informative about the broad skill domains our framework was designed to capture.

Grade repetition and some subdomains of achievement improved while the targeted skills did not. These estimates are less precise when standard errors are clustered by randomization stratum (subsection 5.2), so we treat them as suggestive. If they reflect real effects, they need not run through the skills we measured, since grade repetition and performance on productive language tasks also depend on effort, engagement, and follow-through. The reduction in grade repetition is consistent with the intervention 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; the increase in individual meetings with the specialist (subsection 6.1) points the same way, although this increase was small and statistically insignificant for students classified as being at risk (subsection 6.3), and our data cannot link those meetings to the students who would otherwise have repeated a grade. The language gains concentrate on the productive components of language rather than on receptive reading comprehension, a pattern consistent with the workshops' emphasis on communication and expression.

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 decline in self-control, which survives correction under both levels of clustering, is harder to account for. One possibility is that the workshops raised the standard against which students judge themselves, depressing self-assessment even where behavior did not change; Lafortune et al. (2024) suggest the same explanation for an entrepreneurship program in Rwanda that also reduced self-reported targeted skills.

7 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. These results suggest that governments can use existing school staff to deliver socioemotional support at scale, but that support at this dose is unlikely to improve students' socioemotional skills.

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. More positively, improvements appear in two sets of secondary outcomes: grade repetition fell by 1.7 percentage points, and achievement rose in specific science and language subdomains, namely physics, chemistry, and biology, written expression in Arabic and French, and oral expression in Arabic. All of these estimates survive correction for multiple hypothesis testing with the pre-specified school-level clustering; yet with the larger standard errors from clustering by randomization stratum, only the self-control decline and the physics and chemistry subdomains reach the 10-percent threshold, so we treat the improvements as suggestive.

Section 6 examined why the added contact did not translate into the targeted skills. Take-up varied across schools, but the targeted skills did not improve even where it was high. By design, the intervention's dose was low: four workshops over the year, far fewer sessions than the typical program in the meta-analysis of Cipriano et al. (2023). Our estimates rule out intent-to-treat effects of the size that meta-analysis reports. Individual follow-up rose no more for at-risk than for not-at-risk students.

One further qualification concerns what our design identifies. 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 that 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.

Taken together, this 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, and the improvements we observe in grade repetition and in specific science and language subdomains remain suggestive. 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.

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 study PSP schools not assigned to receive the socioemotional support intervention, and Treatment refers to 84 study PSP 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 at baseline*

	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 of variables measured at baseline. "Control" refers to the 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-stratum 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: Exposure to the socioemotional intervention		
Student participated in socioemotional workshops	0.217	0.432***
	[0.412]	(0.023)
Knows social specialist	0.435	0.169***
	[0.496]	(0.032)
Correctly names social specialist	0.372	0.168***
	[0.483]	(0.033)
Met individually with social specialist	0.248	0.061***
	[0.432]	(0.019)
Received psychological support from social specialist	0.051	0.024**
	[0.221]	(0.010)
Received academic support from social specialist	0.036	0.015*
	[0.186]	(0.008)
Panel B: Exposure to other reform components		
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. Panel A reports students' exposure to the socioemotional intervention, which was assigned to 84 of the 200 study schools; Panel B reports exposure to the reform components that all 200 schools received. The sample is the non-attriting students, and availability varies by measure: 4,012 for questions administered with the Arabic and French exams, and 10,526 for the TaRL question. "Correctly names social specialist" equals one for students who named their school's specialist correctly and zero for all others, including those who say they do not know the specialist's name. Column (1) reports the control-group mean, where control refers to the 116 Pioneer schools that did not receive the socioemotional workshops. Column (2) reports the intent-to-treat effect, the regression-adjusted difference between treated and control schools, controlling for randomization-stratum fixed effects. 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 intervention'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 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. (2026). 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}
Excluded by end of school year	0.003	0.000
	[0.053]	(0.001)
		{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}

Notes. This table reports on the intervention's intent-to-treat (ITT) effects following equation (1). Panel A reports the effect of the intervention on student learning among the 10,526 non-attriting students across the 200 government lower secondary schools. Panel B shows the effect of the intervention on dropout and repetition among all 128,055 students in grades 7 and 8 in the study schools' administrative enrollment records. 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. (2026). 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 intervention among the 4,012 non-attriting students in the 200 study schools who sat the written exams in Arabic or French. 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. (2026). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, 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 intervention'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, each estimated in the full sample as the sum of the treatment coefficient and the relevant treatment-by-subgroup interaction. The estimation sample varies by outcome: 4,012 for socioemotional and mechanism outcomes (519 at risk, 13%; 2,028 girls, 51%) and 10,526 for learning (1,401 at risk, 13%; 5,197 girls, 49%). 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. (2026). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, before adjustment for MHT.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M., 2023. When Should You Adjust Standard Errors for Clustering? The Quarterly Journal of Economics 138, 1–35. doi:10.1093/qje/qjac038.

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 WPS-260. Center for Effective Global Action. 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.

Bronfenbrenner, U., 1977. Toward an Experimental Ecology of Human Development. American Psychologist 32, 513–531. URL: http://psycnet.apa.org/journals/amp/32/7/513/.

de Chaisemartin, C., Ramirez-Cuellar, J., 2024. At What Level Should One Cluster Standard Errors in Paired and Small-Strata Experiments? American Economic Journal: Applied Economics 16, 193–212. doi:10.1257/app.20210252.

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.

Ibrahim, H., de Barros, A., Deschênes, S., Glewwe, P., 2026. The Best Buy? From Global Evidence to Government Practice. Working Paper 12866. CESifo. Munich. URL: https://doi.org/10.65864/afr3uhpaci.

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 .

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.

Lafortune, J., Pugatch, T., Tessada, J., Ubfal, D., 2024. Can gamified online training make high school students more entrepreneurial? Experimental evidence from Rwanda. Economics of Education Review 101, 102559. doi:10.1016/j.econedurev.2024.102559.

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.

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., 2026. The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. The Quarterly Journal of Economics 141, 2675–2734. doi:10.1093/qje/qjag042.

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

World Bank, 2026. World Development Indicators. Data. URL: https://data.worldbank.org/products/wdi.

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.