How Far Can a Short Course Go? Evidence from Remedial Summer Camps in the Dominican Republic

Carolina Concha-Arriagada, Guthrie Gray-Lobe, and Michael Kremer

June 2026

[UPDATED FREQUENTLY: LATEST VERSION HERE]

School systems in low- and middle-income countries routinely respond to low achievement by holding students back a grade—a remedy that consumes a full year of schooling and seldom raises learning. We evaluate a cheaper alternative within the same system: a short, government-run remedial summer camp in the Dominican Republic. Randomizing invitations among overage students in grades 3–5, we find that the average attendee gains about 0.10 SD in test scores—roughly a fifth of a year of learning, produced in a tenth of the time—and that, per dollar, the camp delivers about 50% more learning than a year of schooling, the resource grade retention consumes. Camp effectiveness, however, is heterogeneous: a camp one standard deviation above the mean is nearly twice as effective as the average. For systems where grade retention remains the default response to overage, these results make the case for short remedial camps as a far cheaper way to recover lost learning.

Keywords: Education, Summer Camp, Foundational Learning, Grade Repetition, Literacy, Numeracy, Migrant Students

JEL Codes: I21, I25

Concha-Arriagada: Columbia University. Gray-Lobe: University of Chicago. Kremer: University of Chicago.
This project received IRB approval from the University of Chicago (IRB24-1522); and was preregistered in the AEA RCT registry with ID 0015257 (https://www.socialscienceregistry.org/trials/15257) (Concha-Arriagada, Gray-Lobe, and Kremer 2025). We are grateful to Aurora Salvador for outstanding research assistance. We thank the Dominican Republic's Ministry of Education (MINERD), in particular Ancell Scheker, Vice Minister of Pedagogical and Technical Affairs; Madiery Vásquez, our main counterpart at the ministry throughout the project; and, in the department in charge of the program, Norma Mena, Junior García, and Aneudy Minier. We also thank SIGERD for data support. All errors are our own.

1. Introduction

School systems in low- and middle-income countries routinely respond to low achievement by holding students back a grade—a remedy that consumes a full year of schooling for learning gains that may never materialize. The costs are large and certain: a second year of public spending on material already covered, a year of the student's own time, and a delayed entry into the labor market. The benefits, by contrast, are far from established—the evidence on whether repetition improves learning is thin and contested (Manacorda 2012; Jacob and Lefgren 2004; Koppensteiner 2014). Remediation outside the regular school year is a natural alternative, but the rigorous evidence comes largely from NGO-led pilots and rich-country tutoring. Whether a government can deliver it effectively at national scale, especially in Latin America and the Caribbean, remains an open question.

This paper studies *Aprender es Divertido*, a remedial summer camp designed and run by the Dominican Republic's Ministry of Education—a setting where more than a quarter of fifteen-year-olds report having repeated a grade yet only 8 percent reach minimum mathematics proficiency on PISA. The program offers low-performing students in grades 3–5 three to four weeks of full-day instruction in mathematics and Spanish during the school vacation, in classes of roughly 25 taught by two regular public-school teachers. Attendance is free and students receive meals on site. The intervention is short by international standards—105 to 143 instructional hours, depending on cohort—and runs on existing public-school teachers and facilities rather than NGO-trained tutors or dedicated infrastructure.

We evaluate the program through a field experiment conducted across two consecutive cohorts, in 2024 and 2025. Eligible students were randomly assigned to receive an invitation to attend the camp, with approximately 60 percent assigned to treatment, in a sample of over 16,000 students across 478 schools and 49 districts. Attendance was voluntary, so invited students could decline and some non-invited students attended.

We report two estimands. The first is the intent-to-treat effect of being invited, which compares endline test scores across the randomly assigned invitation. Because take-up was incomplete, the intent-to-treat effect understates the return to attending, so we also instrument days of camp attended with the random invitation. Under a monotonicity assumption on treatment intensity, this recovers an average causal response: the per-day return to attendance. All specifications were pre-registered in our pre-analysis plan (AEA RCT Registry 0015257).

We find that the camp raised foundational learning. Fifteen days of attendance increased aggregate test scores by about 0.10 standard deviations, with similar gains in mathematics and Spanish, and the intent-to-treat estimates are positive and significant in both cohorts. These gains were broadly uniform across the eligible population: we find no significant differences across the baseline test-score distribution, by gender, or between Dominican and Haitian-origin students. The main exception is that students three or more years overage gained significantly less in Spanish, a pattern present in both intent-to-treat and per-day estimates that partly reflects their lower attendance. The two cohorts differed in their subject profiles—2024 gains were concentrated in Spanish, with a near-zero point estimate in mathematics, whereas 2025 showed comparable gains in both. We cannot reject equality of the cohort effects, but this reflects the wide confidence intervals in the smaller, lower-compliance 2024 cohort rather than genuine similarity.

Between 2024 and 2025 the ministry made several changes at once—tightening enforcement, adding transportation, shortening travel distances, grouping classrooms by ability, and extending the camp by a week—and compliance improved sharply. In 2024, weak enforcement let nearly one in five non-invited students attend and the invitation raised full-program attendance by only 7 percentage points; in 2025, contamination fell below 2 percent and the invitation's effect on full attendance rose to 40 percentage points. Because the changes were introduced together, we cannot say which of them drove the improvement, or how much each contributed. The cross-cohort pattern shows that both take-up among the invited and exclusion of the non-invited are malleable to within-system implementation choices.

We then ask whether selection into attendance drives these results. Comparing OLS estimates of the per-day return with the instrumented estimates, the two coincide closely

(0.11 versus 0.10 SD on aggregate scores), and Wu–Hausman tests fail to reject exogeneity of attendance. Under linear and homogeneous per-day returns, this convergence points to a limited role for selection on unobservables—of interest for the broader literature on voluntary remedial programs, where observational estimates are routinely discounted on exactly this concern.

Effectiveness nonetheless varied widely across camps, and the variation came from how productively students learned rather than how much they attended. Following Walters (2015), we estimate the cross-camp standard deviation of the per-day return for the 2025 cohort and reject equal camp effects ($p < 0.001$). The dispersion is about 0.07 standard deviations—roughly 80 percent of the mean per-day return—so a camp one standard deviation above the mean is nearly twice as effective as the average. Observable camp, teacher, and student characteristics explain essentially none of this variation.

Finally, the camp is highly cost-effective relative to a year of schooling. Benchmarking against the cross-grade learning gradient among uninvited students—each grade corresponds to about 0.47 standard deviations—a per-attendee gain of 0.11 SD is roughly a quarter of a school year, achieved in a tenth of the days. At $234 per invited student against about $2,700 for a pupil-year, the camp delivers around 0.025 standard deviations of learning per $100 spent, against 0.017 for a year of schooling—about 45 percent more learning per dollar. Consistent with this, we find suggestive evidence that the camp raised grade progression in the 2024 cohort, though the estimate is imprecise.

This paper makes four contributions. First, we contribute to the literature on remedial instruction and its delivery at scale. Such instruction produces some of the largest known learning gains, whether through community tutors (Banerjee et al. 2016) or one-on-one and small-group tutoring (Nickow, Oreopoulos, and Quan 2024), but the effects come largely from NGOs and volunteers and hinge on delivery: teacher-led, intent-to-treat effects run an order of magnitude below high-fidelity volunteer delivery (Angrist and Meager 2023), even as governments can still raise learning at scale (Muralidharan, Singh, and Ganimian 2024; Duflo, Kiessel, and Lucas 2020). We evaluate a program designed and run by a national government with its own teachers; its intent-to-treat effect of 0.06

standard deviations sits at this teacher-delivered benchmark, and compliance, we show, responds to within-system implementation choices rather than being fixed.

We also contribute to the literature on remediation delivered outside the school year. The rigorous evidence here is mostly from the United States and is mixed, ranging from modest, subject-specific gains (Matsudaira 2008; Mariano and Martorell 2013; Kim and Quinn 2013) to null intent-to-treat effects (Borman and Dowling 2006). We provide the first rigorous evidence on short, vacation-period remediation run by a government in a low- or middle-income country, where the learning deficits and delivery constraints differ sharply.

We further contribute to the literature on grade retention and its alternatives. Retention — the status-quo response to low achievement — is expensive, and its learning benefits are contested: grade failure raises dropout in Uruguay (Manacorda 2012) and averages near zero across studies (Goos, Pipa, and Peixoto 2021), with some exceptions (Gomes-Neto and Hanushek 1994; Jacob and Lefgren 2004). Whatever its effect on learning, retention costs a full year, so a short program that recovers part of that progress for a fraction of the cost is an attractive alternative. We benchmark the camp's effect and cost against an additional year of schooling, and find suggestive evidence that it raises on-time grade progression — the margin retention targets.

Finally, we contribute to the literature on heterogeneity and selection in program effects. Walters (2015) finds that observable inputs explain only part of the cross-site variance in Head Start effects, while voluntary take-up raises the familiar concern of selection on unobservables. We speak to both: adapting his decomposition, we find that camp effectiveness varies in ways observed inputs do not explain, and the close agreement of OLS and instrumental-variables estimates points to limited selection. The camp's 0.10 standard-deviation effect sits at the median of trials in low- and middle-income countries (Evans and Yuan 2022; Kraft 2020).

The paper proceeds as follows. Section 2 describes the Dominican education context, the camp, and the design changes between the 2024 and 2025 cohorts. Section 3 presents the data and checks baseline balance. Section 4 lays out the estimation framework. Section 5 reports effects on attendance, intent-to-treat and average-causal-response estimates, and heterogeneity by baseline score, grade, demographics, and overage status, and closes with the OLS–ACR comparison that bears on selection. Section 7 concludes.

2. Context and Experiment Design

The Dominican Republic is an upper-middle-income country with a GDP per capita of approximately $10,900. Despite sustained investment in primary education, Foundational learning outcomes in the Dominican Republic are among the lowest in Latin America: over 80% of sixth-grade students fail to reach minimum reading proficiency, and more than 95% do not meet minimum standards in mathematics (LLECE 2021). This is consistent with World Bank learning poverty estimates, which place 81% of late-primary-age Dominican children below minimum reading proficiency after adjusting for out-of-school children (World Bank 2019).

2.1. Educational Context in the Dominican Republic

The Dominican Republic's primary education system faces persistent challenges related to low academic achievement, grade repetition, and overage enrollment. In PISA 2022, the country ranked among the lowest-performing participants globally, with only 8 percent of 15-year-old students reaching minimum proficiency in mathematics — far below the OECD average of 69 percent. Grade repetition is widespread: 26 percent of Dominican students reported having repeated at least once. National standardized assessments at the primary level confirm these deficits — 61 percent of 3rd-grade students scored at the elementary level in Spanish and 59 percent in math, meaning they had not acquired foundational competencies.

These learning gaps are closely linked to overage enrollment and academic failure. According to MINERD administrative records for the 2023–2024 school year, almost 30 percent of students in grades 3–5 are at least one year behind their age-appropriate grade level, with a failure rate of 7.2 percent and a dropout rate of 1.3 percent in the same grades.

Failure rates also rise sharply with overage status: 5 percent among on-age students, 11 percent among students one year overage, and 14 percent among students two or more years overage. We do not interpret this gradient as a causal effect of retention — it reflects both selection (lower-performing students are more likely to be retained in the first place) and any direct consequences of being held back. Causal evidence on the latter, summarized in our discussion of grade retention in the introduction, suggests that retention itself contributes to subsequent academic and behavioral difficulties (Manacorda 2012; Jacob and Lefgren 2004; Koppensteiner 2014).

The regions selected by MINERD to participate in the summer camp program have greater academic need and socioeconomic disadvantage. Participating regions have slightly higher shares of overage students in grades 3 to 5 — approximately 31 percent — and higher failure rates approaching 9 percent. These regions also have a larger share of rural schools, reaching 34 percent in the 2024 participating regions compared to 24 percent nationally. On the 2024 national diagnostic assessment (*Pruebas Diagnósticas*), students in participating regions score 5 to 8 points below the national mean in both subjects, with a larger share at the elementary proficiency level. Participating regions are also more socioeconomically disadvantaged: in the 2024 regions, 31 percent of students were in the lowest socioeconomic quintile, compared to 18 percent nationally. Moreover, participating regions exhibit higher failure rates within every overage category, suggesting that the program targets not only regions with more overage students but also regions with worse academic outcomes conditional on overage status.

Finally, participating regions serve a disproportionate share of Haitian migrant students. Haitian students constitute approximately 8 percent of the national student population in grades 3–5, but between 10 and 12 percent in participating regions. These students may face additional barriers to learning, including language differences, as Spanish is the language of instruction in Dominican public schools. We examine whether program effects differ for this population in Section 5.

2.2. Program Design and Implementation

In July 2024 and July 2025, we conducted a randomized controlled trial in collaboration with the Dominican Republic's Ministry of Education (MINERD) to evaluate a government-designed remedial summer camp targeting low-performing primary school students in grades 3–5. This subsection describes the implementation of the program in each year, followed by the curriculum and pedagogical approach.

2.2.1. 2024 Implementation

MINERD implemented the first edition of the summer camp from July 8 to July 26, 2024, across 29 districts (out of 122), with one camp operating in each district. The program provided remedial instruction in mathematics and Spanish, focused on strengthening foundational skills aligned with the national curriculum. The camp ran for 15 instructional days, with a daily schedule from 8:30 a.m. to 3:30 p.m. that alternated between mathematics and Spanish instruction blocks.

Each camp organized students into classrooms by grade: two classrooms for 3rd grade, one for 4th grade, and one for 5th grade. Class sizes were approximately 25 students, with two teachers assigned to each classroom. Students received three daily meals at the camp.

Instructional methods combined interactive games and activities and smaller-group instruction for targeted support. Students facing particular difficulties were temporarily separated for more focused reinforcement during Spanish and, in some cases, mathematics sessions. Daily activities were adjusted based on observed learning levels, though no formal differentiation by performance level was implemented (for details of camp staff's vision of on-site implementation see Appendix B).

2.2.2. 2025 Implementation

The second edition of the camp incorporated several design changes informed by the experience of the first year. The program expanded to 49 districts, with 46 districts operating one camp each and 3 districts operating two camps, for a total of 52 camp sites. The camp ran for 19 instructional days, with a daily schedule from 8:00 a.m. to 3:30 p.m. that alternated between mathematics and Spanish instruction blocks.

The most significant pedagogical change was the introduction of ability-based grouping. Rather than organizing classrooms by grade, students were grouped into three performance levels — Exploradores, Creadores, and Innovadores — based on results from an initial diagnostic assessment. During the first week, students who had not previously completed the baseline diagnostic took it at the camp, and these results were used to finalize classroom assignments. Class sizes remained at approximately 25 students with two teachers per classroom.

The 2025 curriculum also broadened in scope, integrating literacy and numeracy instruction with socio-emotional learning, identity development, technology use, arts, and values education. Activities were organized weekly around thematic axes and adapted to students' performance levels, with continued emphasis on interactive, participatory, and game-based learning.

Two additional operational changes aimed at improving attendance. Transportation was provided to reduce access barriers, addressing the long travel distances that contributed to low take-up in 2024. Students continued to receive three daily meals, with the addition of an afternoon snack designed to enhance motivation and enjoyment (for details of camp staff's vision of on-site implementation see Appendix B).

2.2.3. Math and Spanish Curriculum and Pedagogical Approach

The camp curriculum focused on Spanish and mathematics, with content aligned to the national primary curriculum and designed to reinforce foundational literacy and numeracy skills. In Spanish, instruction covered reading comprehension across different text genres, identification of text structure and purpose, and written production at increasing levels of complexity. In mathematics, instruction progressed from numeration and place value through basic operations (addition, subtraction, multiplication, and division) to problem-solving using arithmetic.

While the core content domains remained consistent across both years, the structure and delivery of the curriculum evolved substantially between 2024 and 2025.

In 2024, the camp ran for three weeks (15 instructional days). All classrooms followed the same instructional guide, with content organized in a weekly progression. In Spanish, students worked with instructive texts in the first week, narrative texts in the second, and informational texts in the third. In mathematics, instruction moved from number sequences and place value to addition and subtraction, and then to multiplication and division. Topics in both subjects were defined broadly, and teachers were expected to adjust depth and pacing depending on student skill levels. There was no formal differentiation in content by performance level or grade, and students were grouped by the grade they attended in the last academic year.

In 2025, the camp was extended to four weeks (19 instructional days) and introduced explicit differentiation by student ability. Students were grouped into three performance levels (Exploradores, Creadores, and Innovadores). The same core content domains were maintained, but objectives were tailored to each level. In mathematics, differentiation operated through varying number ranges and levels of operational complexity — for example, the lowest level worked within 1–20 while higher levels extended to 999 and beyond. In Spanish, differentiation took the form of variation in reading difficulty and writing tasks, ranging from basic sentence construction to more complex text production. Content was also distributed more gradually across instructional days relative to 2024 (for details on materials and notebooks see Appendix C).

2.3. Experimental Design

This subsection describes the sampling strategy used to identify eligible students in each cohort, followed by the randomization procedure used to assign students to Invited and Comparison groups.

2.3.1. Sampling

The study draws on two cohorts of public primary school students in the Dominican Republic, sampled in 2024 and 2025. In both years, the Ministry of Education (MINERD) selected participating schools based on indicators of academic need, including the number of overage students, grade repetition rates, and average test scores. However, the process for identifying eligible students within schools differed across cohorts.

In 2024, the eligible population consisted of students in grades 3–5 across 29 educational districts (out of 122). Within selected schools, teachers were asked to nominate students with the lowest academic performance in their classrooms. This process yielded a study sample of 5,482 students from 302 schools (out of 1,167 in the participating districts). Approximately 75 percent of nominated students were at least one year behind their age-appropriate grade, though overage status was not a formal eligibility requirement.

The 2025 sampling process differed from 2024 in two important ways, alongside an expansion from 29 to 49 districts. First, MINERD replaced teacher nomination with an administrative identification procedure. Using records from the national student information system (SIGERD), the ministry identified students who were at least one year behind their age-appropriate grade. Selection followed a prioritized rule: all students with three or more years of overage were included first, followed by those with two years of overage, and finally a random sample of students with one year of overage until district-level quotas were met. Second, MINERD incorporated the distance between schools and camp sites into the school selection process. In 2024, some students were enrolled in schools located up to 25 kilometers from their assigned camp, contributing to low take-up. In 2025, nearly all participating students attended schools within 5 kilometers of the camp. Together, these changes produced a larger study sample of 11,311 students from 320 schools (out of 2,141 in the participating districts), of whom approximately 38 percent were one year overage, 32 percent were two years overage, and 30 percent were three or more years overage.

Because identification relied on teacher nomination in 2024 and administrative records in 2025, the two cohorts are not drawn from identical populations. Results for each cohort should be interpreted as treatment effects for the respective eligible population. We account for this by presenting results separately by cohort in addition to pooled estimates.

2.3.2. Randomization Procedure

In both years, eligible students were assigned to receive an invitation to attend the summer camp or to a comparison group, with approximately 60 percent receiving an invitation to participate. Randomization was at the student level, stratified to ensure balance on observable characteristics. Stratification variables differed slightly across cohorts to reflect the available data at the time of randomization and operational constraints.

In 2024, randomization was stratified by region, district, school, grade, and an age bin defined by within-district quantiles, producing 1,166 strata.[1] Of these, 279 contained only a single student and therefore lack within-stratum treatment variation. This yielded 3,369 students invited to the camp and 2,113 in the comparison group.

In 2025, the stratification procedure was more elaborate to accommodate a larger number of covariates while maintaining adequate cell sizes. The initial stratification was defined by the interaction of district, assigned camp center, school, baseline tests attendance, and nationality. A minimum cell size of 10 students was imposed, and the procedure iteratively collapsed stratification dimensions — first nationality, then baseline tests attendance — in cells that fell below this threshold. The resulting stratification grouped students by district, camp center, and school, with nationality and baseline attendance retained only where cell sizes permitted. This process produced 476 final strata, within which 6,793 students were invited and 4,518 are part of the comparison group.

Table 1 summarizes the key design differences between the 2024 and 2025 implementations. While core features of the program — including daily schedule, class size, and treatment share — remained largely unchanged, the 2025 cohort introduced several important modifications, including a 30-min longer instructional period, explicit curriculum differentiation by performance level, transportation provision, and the use of administrative records for student identification.

[1] 279 strata contained only a single student and therefore lack within-stratum treatment variation. These students are excluded from all specifications by construction since since we use stratum fixed effects.

TABLE 1. Program Design Comparison: 2024 vs. 2025

Feature	2024 Cohort	2025 Cohort
Districts (camp sites)	29 (29)	49 (52)
Instructional days	15	19
Daily schedule	8:30 a.m.–3:30 p.m.	8:00 a.m.–3:30 p.m.
Student grouping	By grade	By performance level (3 tiers)
Curriculum differentiation	None	Explicit (3 leveled workbooks)
Student identification	Teacher nomination	Administrative records (SIGERD)
Transportation provided	No	Yes
Max distance to camp	25 km	5 km
Meals per day	3	3 + afternoon snack
Socio-emotional component	No	Yes
Counselor/psychology support	No	Yes (select sites)
Class size	≈25 (2 teachers)	≈25 (2 teachers)
Study sample	5,482 students	11,311 students
Participating schools	302 schools	320 schools
Treatment share	61%	60%

3. Data

Our analysis relies on three main sources of data: administrative records, camp's daily attendance, and assessments. We use these records to create the main outcomes of interest and the main variable that represent the dosage that students received during the remedial summer camp. Additionally, from administrative records we obtain student demographics and baseline covariates, including age, overage status, nationality, and gender.

We collected baseline and endline assessments in mathematics and Spanish to measure students' foundational numeracy and literacy progress. Baseline tests were administered to participating students in April 2024 and May 2025, prior to the summer camp in each cohort. The instruments were developed by MINERD to measure foundational numeracy and literacy skills aligned with the national primary curriculum. For the 2025 cohort, the tests replicated the 2024 baseline assessments with minor modifications informed by IRT analysis, and expanded with items drawn from the LAYS Item-Level Design Bank.

In mathematics, the 2025 baseline retained the core item set from 2024, with minor updates to numerical values and contextual details, and introduced new items covering place value decomposition, number ordering, time calculation, and graphical data interpretation. In Spanish, the 2025 baseline retained the main skill domains with minor wording adjustments, introduced a new listening comprehension item following guidelines from the *Foundational Reading Assessment Tool* (UNICEF, J-PAL, Pratham).

Endline tests for both subjects were administered in September 2024 and November 2025, approximately two and four months after the conclusion of the camp, respectively. This timing means our estimates capture short-term persistence of learning gains rather than immediate post-treatment effects. Within each year, the endline mirrors the structure and skill domains of the corresponding baseline instrument, with parallel items designed to measure the same underlying constructs. A detailed item-by-item comparison across rounds and cohorts is provided in Appendix D.

For each subject, cohort, and time of the test (baseline/endline), we estimate a two-parameter logistic (2PL) IRT model and recover student *latent ability scores* (θ_i). These ability measures, standardized to the comparison group mean within each cohort, are our primary academic outcome variables. We report results for mathematics, Spanish, and an aggregate measure constructed as the simple average of the standardized math and Spanish scores. Because we standardize outcomes within cohort, differences in instrument composition across years do not affect comparability of treatment effect estimates. Details on model estimation, item screening, and reliability analysis are provided in Appendix F.

To measure grade progression after the remedial camps, we use administrative records. For that, we use information on enrollment for two cycles: the academic year (AY) starting immediately after the camp, and the AY starting one year after the camp. Given the timeline, we only present results for 2024 cohort, we hope to add the 2025 cohort at the end of 2026. We create three variables of interest for this cohort. First, a variable ($progressed_i$) equal to 1 if student enrolled in a higher grade the academic years after the camp, and 0 otherwise. Second, a variable ($dropout_i$) that is equal to 1 if students did not enroll in any school, and third, a variable ($switch_i$) if student switched school from the one enrolled in the previous academic year.

Finally, daily student attendance was recorded in the classroom throughout the summer camp. In 2024, attendance sheets were completed on paper at each camp site and subsequently digitized by the enumerator team. In 2025, attendance sheets were maintained by camp staff and digitized directly by them. From these records, we construct the total number of days each student attended and a normalized dosage measure — the share of camp days attended ($shat15_i$), defined as total days divided by 15.

TABLE 2. Balance Table (2025)

	Comparison		Invited		Difference	
	Mean (1)	SE (2)	Mean (3)	SE (4)	(4)-(2)	P-value
Age (in years)	11.242	0.016	11.254	0.011	0.012	0.651
1 year overaged	0.379	0.005	0.373	0.003	-0.006	0.421
2 year overaged	0.317	0.005	0.317	0.004	0.000	0.991
At least 3 years overaged	0.304	0.005	0.310	0.003	0.006	0.464
Underlying ability in math	-0.000	0.009	0.013	0.006	0.013	0.416
Underlying ability in spanish	-0.000	0.010	0.008	0.006	0.009	0.579
Haitian	0.205	0.004	0.208	0.003	0.003	0.595
Missing nationality	0.011	0.001	0.010	0.001	-0.001	0.516
Female	0.366	0.005	0.365	0.004	-0.000	0.961
Number students	4,518		6,793			
p-value from joint significance test						0.983

Note: This table shows baseline characteristics by treatment status. We report means, standard errors, and pairwise differences between treatment and comparison groups. Estimates are from an OLS regression of each baseline variable on the treatment indicator, including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study. For students with missing baseline assessments, we impute the within-cohort mean (zero by construction since baseline scores are standardized within cohort). Standard errors are clustered at the school level. We also present p-values from a joint significance test of all variables included in the balance test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Go to Tables A1 and A2.

Table 2 presents baseline characteristics by treatment status pooling results across both cohorts, while Table A1 and Table A2 show results separately for the 2024 and 2025 cohorts. Differences between invited and comparison groups are small in magnitude and statistically insignificant across all covariates, including age, overage status, baseline test scores in math and Spanish, nationality, and gender. A joint F-test of the null hypothesis that all covariates simultaneously predict treatment assignment fails to reject in the pooled sample and in each cohort separately (p-values of 0.99, 0.97, and 0.98, respectively).

Table 3 examines attrition across cohorts, while Table A3 and Table A4 do the same analysis by cohorts. Overall, 37 percent of students in the comparison group do not have endline test scores. This attrition rate is driven primarily by the 2024 cohort (52 percent for Spanish and 55 percent for math); the 2025 cohort has a substantially lower attrition rate which is very similar across test (around 30 percent). In the pooled sample, differential attrition between invited and comparison groups is small and insignificant for both subjects. In the 2024 cohort, treatment students are significantly less likely to attrit in both subjects (3.0 percentage points for Spanish and 4.8 percentage points for math). In the 2025 cohort, differential attrition is significant for both subjects and around 2 percentage points. The direction of differential attrition — lower in invited group and students with higher baseline scores — implies that the remaining treatment group may be positively selected, biasing estimates upward. We do sensitivity analysis using Lee (2009) bounds in Appendix I.

TABLE 3. Attrition in Endline Tests (Pooled)

	Spanish		Math	
	(1)	(2)	(1)	(2)
Treatment	-0.010	0.004	-0.015**	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)
Age (in years)		0.030***		0.026***
		(0.005)		(0.005)
No overaged		0.032		0.046*
		(0.022)		(0.025)
1 year overaged		[base]		[base]
2 year overaged		0.002		0.007
		(0.012)		(0.011)
At least 3 years overaged		0.021		0.035*
		(0.018)		(0.018)
Underlying ability in spanish		-0.008		-0.007
		(0.006)		(0.006)
Underlying ability in math		-0.021***		-0.025***
		(0.005)		(0.005)
Hatian		0.013		0.016
		(0.013)		(0.013)
Female		-0.004		-0.006
		(0.008)		(0.008)
Control mean	0.368	0.368	0.378	0.378
Number observations	16,395	16,395	16,395	16,395
Number schools	478	478	478	478

Note: This table examines differential attrition by treatment status. The dependent variable is an indicator for whether the student's endline test score is missing by subject. Columns (1) and (2) report results for Spanish; Columns (3) and (4) report results for mathematics. Odd-numbered columns report the unconditional difference in attrition between treatment and comparison groups; even-numbered columns add baseline covariates as controls. Estimates are from an OLS regression including stratum and cohort fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study during 2024 and 2025 for whom baseline test scores are available. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Go to Tables A3 and A4.

4. Estimation Framework

The primary empirical question of this paper is whether attending the summer improves test scores. We assume that the impact of the camps scales approximately linearly with attendance. A naive approach to estimating the impact of the camp would be to use the variation in the number of days that the student attended the camp in

$$(1) \qquad Y_i = \alpha + \beta D_i + \varepsilon_i \,,$$

where y_i is the test score of student i, $D_i \in \{0, 1, ..., \bar{D}\}$, is the number of days that the student attends the camp (with $\bar{D}$ in our case being either 15 in 2024 or 19 in 2025), and ε_i represents the influence of unobserved factors. Estimation of β in 1 by ordinary least squares can lead to biased estimates if D_i is correlated with ε_i. Lower performing students may disproportionately attend the camps if they can benefit more from them. Meanwhile, the skills and supportive structures at home that allow students to attend the camp regularly may be more prevalent among high-performing students, generating positive selection among attendees.

The analysis presented in this study is centered around the impact of the invitation to attend the camp. As we will show, invitations to the camp increased camp attendance D_i and are uncorrelated with the error term ε because they were randomly assigned. Specifically, we estimate the following model of test scores as a function of invitations.

$$(2) \qquad Y_i = \delta Z_i + X_i'\Gamma + \eta_i$$

where X_i contains randomization strata fixed effects, as well as controls for grade and baseline test scores. The estimand δ in Equation 6 is referred to as the intent-to-treat (ITT), insofar as the offers to the camp were meant to ration access to attend the camp (D_i).

Using Z_i as an instrument for D_i allows us to estimate the average causal response of D_i on test scores. The ITT effect and the IV estimate are connected by the first stage effect of invitations on attendance

$$(3) \qquad D_i = \pi Z_i + X_i'\Omega + \upsilon_i \,.$$

The estimand π reflects the increase in attendance induced by offering an invitation.

Dropping the controls in X_i for expositional simplicity, the Wald estimand identifies a weighted average effect of the effect of increasing D_i by one unit (Theorem 4.5.3 Angrist and Pischke 2009). Specifically,

$$\frac{E[Y_i|Z=1]-E[Y_i|Z=0]}{E[D_i|Z=1]-E[D_i|Z=0]}=\sum_{d=1}^{\bar{D}}\omega_d E[Y_i(d)=Y_i(d-1)|D_i(1)\geq d>D_i(0)]\,, \tag{4}$$

where $Y_i(d)$ represents the potential outcome when attending d days of the camp, $D_i(z)$ represents the potential attendance level for individual i with invitation status $z \in \{0,1\}$, and $\omega_d = \frac{P[D_i(0)\geq d>D_i(0)]}{\sum_{j=1}^{\bar{D}} P[D_i(1)\geq j>D_i(0)]}$. We report estimates of the ACR in Equation 4 scaling D by 15, so that the estimate reflects the impact of attending 15 days, the full length of the 2024 program.

Identification of the ACR requires four assumptions. The first three are the standard assumptions for binary endogenous variable: relevance, independence and the exclusion restriction. The fourth assumption is monotonicity, which in the variable-intensity case is more demanding than its binary counterpart. In our setting, the first two assumptions are satisfied by construction. The exclusion restriction and monotonicity assumptions are discussed below.

The exclusion restriction requires that the random invitation affects student learning outcomes only through its effect on camp attendance. This assumption would be violated if, for example, the invitation itself conveyed information to parents — such as a signal that their child was academically at risk — that prompted changes in home learning behaviors independent of camp participation. We view these concerns as limited because invitation letters did not disclose individual performance information, and the camp took place during the school vacation period when no regular instruction was scheduled.

Monotonicity in the variable-treatment-intensity setting is more demanding than in the binary case. Rather than simply requiring that the instrument shifts all individuals in the same direction (into or out of treatment), monotonicity here requires that the invitation weakly increases the attendance dosage for every student: $D_i(1) \geq D_i(0)$ for all i. This rules out the existence of "defiers at any margin" — students who would attend fewer days if invited than if not invited. We view this assumption as plausible in our setting.

The invitation provides access to a free program with no associated costs or penalties; there is no obvious mechanism through which receiving an invitation would reduce a student's attendance at the camp. The main threat would arise if invited students were displaced from alternative learning opportunities by the invitation, but since the camp operated during the school vacation period when no regular instruction was scheduled, this concern is minimal.

Under a further assumptions that the effect scales linearly with each unit and that the average marginal effects over the full population, the ACR from Equation 4 and Equation 1 reflect the same structural estimand. We compare OLS estimates of Equation 1 to the identified estimates of the ACR in section **??**. The better identification of the estimates in Equation 4 come at the cost of less precision because the ACR is identified only from variation in attendance that arises from the invitation. OLS estimates of Equation 1 are more precise, using all variation in attendance, at the cost of potentially biased estimates due to selection effects. Comparison of OLS and ACR estimates yields a test of the selection effect under the assumption of linearity.

5. Results

5.1. Effects on Attendance

A central implementation challenge for voluntary remedial programs is whether students actually attend. We estimate the effect of receiving an invitation to participate in the remedial camp using the following empirical stratregy:

$$A_{iks} = \alpha_0 + \alpha_T T_i + \alpha_g G_{ig} + \alpha_{Y_0} Y_{iks0} + \sum_g \alpha_{g,Y_0} G_{ig} \cdot Y_{iks0} + \delta_k + \varepsilon_{iks}^{FS} \tag{5}$$

where A_{iks} is the attendance outcome of student i, in randomization stratum k, and subject s; T_i is an indicator equal to one if student i was randomly assigned to receive an invitation to the summer camp; G_{ig} is an indicator for student i being in grade $g \in \{3, 4, 5\}$ at baseline, and Y_{iks0} is the student's baseline test scores; δ_k is a vector of randomization stratum fixed effects unique for each cohort; and ε_{iks1}^{FS} is the error term. The coefficient of interest is α_T that measure the causal effect of being invited to the camp on attendance. For students with missing baseline test scores, we impute the cohort-specific mean and include indicator variables for missingness as additional controls. Standard errors are clustered at the school level to account for within-school correlation in student outcomes.[2]

Table 4 presents the effect of receiving a camp invitation on attendance. In the pooled sample, being invited increased the share of camp days attended (over 15 core days) by 0.44 (Column 1), equivalent to 6.7 additional days (Column 2), and raised the probability of attending at least one day by 42.6 percentage points (Column 3) and at least 15 days by 30 percentage points (Column 6).[3]

Compliance rates and program utilization differed markedly across cohorts. In 2024 (Panel B), being invited increased the share of camp days attended by 0.26 (Column 1) and the probability of attending at least 15 days by 7.4 percentage points (Column 6). However, 4.2 percent of comparison-group students attended at least 15 days, as enforcement of treatment assignment at camp sites was limited and uninvited students who arrived could not be turned away. In 2025 (Panel C), stricter enforcement at camp sites substantially reduced the percent of comparison-group students attending at least 15 days to less than 2 percent. For 2025, the corresponding effect of the invitation on the share of camp days attended rose to 0.526, while the probability of attending at least 15 days rose by 40.2 percentage points, representing a 100 and 600% increase from 2024 take up respectively.

The higher compliance in 2025 likely reflects both the reduced contamination and the design improvements summarized in Table 1, including the provision of transportation, shorter distances to camp sites, reorganization of classrooms by student level, and the addition of meals. Overall, the utilization rate — defined as the share of total student-days

[2]The inclusion of stratum fixed effects drops strata in which all students are assigned to the same treatment arm. This occurs in 279 of the 1,166 strata in 2024; no strata are dropped in 2025, where the sequential aggregation procedure ensured a minimum cell size of 10 students per stratum. Additionally, in 2024 some students were assigned to a different school during the randomization process. We dropped a total of 398 observations. Appendix H presents main results with the full sample, including balance tables. Results are almost identical to the results in the main text.

[3]Appendix G presents the results for the sample of students whose schools are up to 9 kilometers from the camp, which allows to have the same support in distance in both cohorts. Our results are very similar between the two samples.

of capacity actually attended — rose from 52 percent in 2024 to 61 percent in 2025. Among invited students only, the increase was sharper, from 42 to 59 percent, suggesting the program improved from one year to another.

TABLE 4. Effect of Being Invited to the Remedial Camp on Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Share att.	Total Days	>= 1 days	>= 5 days	>= 10 days	>= 15 days	= 19 days
***Panel A**: Pooled years*							
Treatment	0.445***	6.670***	0.426***	0.407***	0.380***	0.300***	0.275***
	(0.01)	(0.21)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	16,395	16,395	16,395	16,395	16,395	16,395	11,311
Control mean	0.061	0.911	0.071	0.066	0.058	0.025	0.012
***Panel B**: 2024*							
Treatment	0.264***	3.955***	0.345***	0.319***	0.264***	0.074***	
	(0.02)	(0.30)	(0.02)	(0.02)	(0.02)	(0.01)	
Observations	5,084	5,084	5,084	5,084	5,084	5,084	
Control mean	0.139	2.091	0.176	0.161	0.140	0.042	
***Panel C**: 2025*							
Treatment	0.526***	7.894***	0.463***	0.447***	0.432***	0.402***	0.275***
	(0.02)	(0.23)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311
Control mean	0.025	0.373	0.023	0.023	0.021	0.017	0.012

Note: This table shows the effect of being invited to the remedial camp on the share of days attended out of 15 core days (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 3–7). Estimates are from an OLS regression including stratum and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. We also include cohort fixed effects in the pooled analysis. The sample includes all students selected by the Ministry of Education to participate in the study across the 2024 and 2025 remedial camps, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

While the invitation was a strong predictor of attendance overall, take-up was not uniform across student subgroups. Three characteristics show significant differential attendance: nationality, overage status, and last grade attended.

Figure 1 presents the interaction coefficients for gender, nationality, and distance to the camp. Students of Haitian origin who were invited to the camp have a lower attendance share by 0.08 than non-Haitian invitees. The negative attendance pattern for invited Haitian student is similar and statistically significant for other attendance measures (see Table A5). This differential attendance rate is completely driven by the 2025 cohort, where the share of attended days for Haitian students invited to the camp drops by 0.09 (Table A7), likely reflecting fear among their parents to let their children participate given the heightened enforcement of immigration policies during this period.[4]

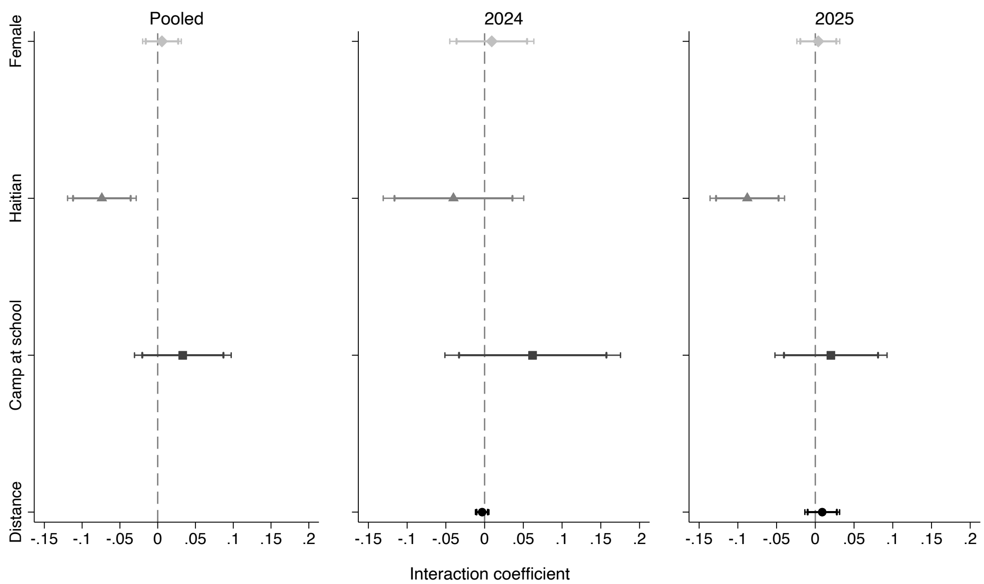

FIGURE 1. Heterogeneous Effect of Being Invited to the Remedial Camp on Attendance

Note: This figure shows the interaction coefficient between treatment assignment and each student characteristic on the share of days attended out of 15 core days. Each point represents the coefficient on the interaction term from a separate OLS regression of attendance on treatment, the student characteristic, and their interaction, with stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan. The left panel reports pooled estimates across both cohorts, the middle panel reports estimates for the 2024 cohort, and the right panel reports estimates for the 2025 cohort. The four characteristics are gender (female), nationality (Haitian), whether the camp is located at the student's school from the previous academic year, and distance in kilometers between the student's school and the camp site. Distance is omitted from the pooled panel because its support differs across cohorts (0–25 km in 2024 vs. 0–10 km in 2025). Standard errors are clustered at the school level. Thin bars represent 95% confidence intervals; thick bars represent 90% confidence intervals. Go to Tables A6 and A7.

Figure 2 presents the share of days attended by students overage status in each cohort.[5] Students with severe overage (three or more years behind) invited to the camp attended

[4]In October 2024, the Dominican government ordered the deportation of up to 10,000 Haitian migrants per week. Over 276,000 people were deported in 2024, with over 310,000 deported in the first ten months of 2025 alone. The enforcement campaign included raids on streets, workplaces, public transportation, and homes, creating a climate of fear among Haitian families that likely discouraged camp attendance. See Inter-American Commission on Human Rights (IACHR), "IACHR concerned about rights of individuals in human mobility contexts in the Dominican Republic," Press Release No. 023/2026, February 6, 2026; Loiseau, M. (2026, February 26), "How Far Will the Dominican Republic Go in Deporting Haitians?," *Inkstick Media*.

[5]We do not present pooled estimates for this category since the sample of eligible students in 2025 did not consider students with no overage as it did in 2024.

systematically fewer days in 2025 than invited students with two or one year overage. On average, their share of days attended decreased by .09 relative to students with one year overage (Table A12). This pattern for students with 3+ years of overage is also present in the 2024 cohort, where the difference is 0.05, although only statistically significant at the 10 percent level. Similar pattern occurs in 2024 for students with 2 years of overage (Table A11).

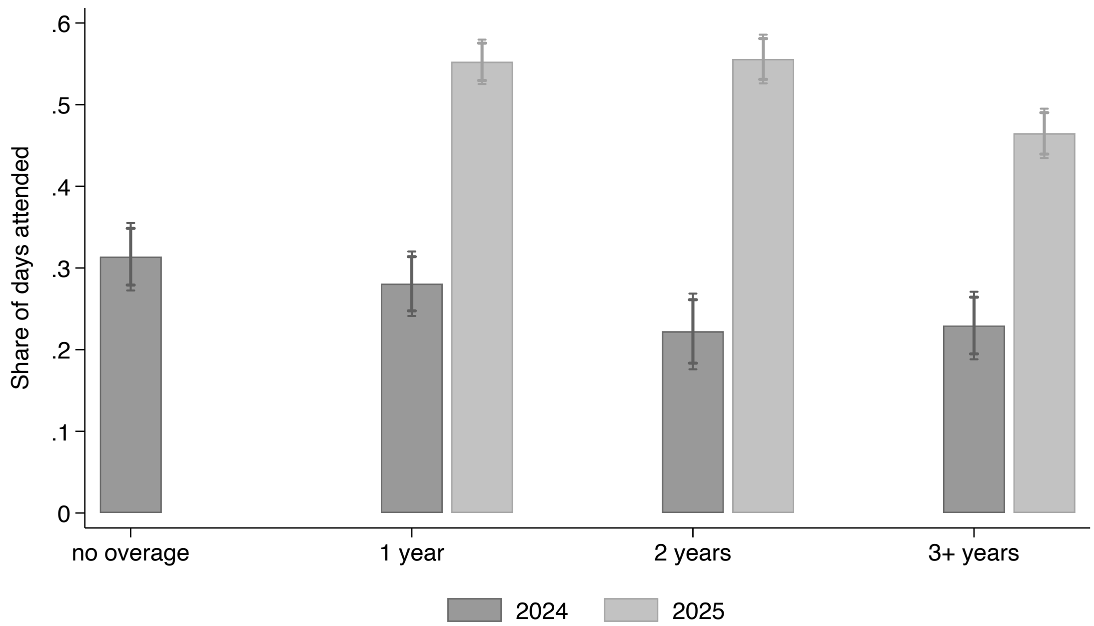

FIGURE 2. Total Effect of Being Invited to the Remedial Camp on Attendance by Students' Overage Status

Note: This figure plots the effect of being invited to attend the remedial camp on the probability of attending at least 15 days, by overage status. Estimates are from an OLS regression of attendance on the treatment indicator interacted with overage status indicators, using one year of overage as the omitted category. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The sample includes students selected by the Ministry of Education separately for the 2024 camp and the 2025 camp, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. Thin bars represent 95% confidence intervals; thick bars represent 90% confidence intervals. The no-overage category is only present in the 2024 cohort.Go to Tables A11 and A12.

Finally, students in higher grades have a lower share of days attended than students in 3rd grade. Fourth and fifth graders alike in the pooled estimation have a lower share by a magnitude of 0.05, (both statistically significant). Qualitative information collected through focus groups suggests the material and classroom environment — including music, decorations, and games during breaks — were less engaging for older students. These three attendance gaps likely reflect distinct mechanisms. Lower attendance among Haitian students appears only in 2025 and is tied to the political environment against migrants; the overage and grade attendance gap, by contrast, are present in both cohorts and appears related to the age-appropriateness of the summer camp's pedagogical approach. Along the remaining dimensions — gender, baseline test score decile, grade, and distance to the camp site — the effect of an invitation on attendance is broadly uniform (see Tables A5 to A14).

TABLE 5. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Grade (Pooled)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
Treatment (β_T)	0.482***	7.226***	0.457***	0.439***	0.409***	0.332***	0.309***
	(0.02)	(0.25)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Treatment × Grade 4 ($\beta_{T,4}$)	-0.053***	-0.789***	-0.041***	-0.045***	-0.044***	-0.046***	-0.047***
	(0.02)	(0.26)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Treatment × Grade 5 ($\beta_{T,5}$)	-0.051***	-0.762***	-0.044***	-0.048***	-0.037**	-0.042***	-0.046***
	(0.02)	(0.25)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Joint F-test	3.63	3.63	3.31	3.30	2.86	4.14	5.08
	0.006	0.006	0.011	0.011	0.023	0.003	0.007
$\beta_T + \beta_{T,4}$	0.429	6.436	0.416	0.394	0.364	0.286	0.262
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,5}$	0.431	6.464	0.412	0.391	0.372	0.290	0.264
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	16,395	16,395	16,395	16,395	16,395	16,395	11,311

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). All specifications include cohort and stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. Column 6 is available only for the 2025 cohort, as the 2024 camp ran for 15 days. The sample includes all students selected by the Ministry of Education to participate in the study, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01.Go to Tables A8, A9, A10, A13 and A14.

5.2. Effects on Learning of Being Invited to the Camp

To estimate the average effect of being invited to the program on endline test scores (ITT effects), we estimate the following specification:

$$Y_{iks1} = \beta_0 + \beta_T T_i + \beta_g G_{ig} + \beta_{Y_0} Y_{iks0} + \sum_g \beta_{g,Y_0} G_{ig} \cdot Y_{iks0} + \delta_k + \varepsilon_{iks}^{ITT} \tag{6}$$

where Y_{iks1} is the standardized test score of student i, in randomization stratum k, in subject s, measured at endline; T_i is an indicator equal to one if student i was randomly assigned to receive an invitation to the summer camp; G_{ig} is an indicator for student i being in grade $g \in \{3, 4, 5\}$ at baseline, and Y_{iks0} is the student's baseline test score in the same subject; δ_k is a vector of randomization stratum fixed effects unique for each cohort; and ε_{iks1}^{ITT} is the error term. All outcomes are standardized to the comparison group mean within each cohort. The coefficient of interest, β, captures the average effect of being invited to the program on endline test scores, regardless of whether the student actually attended the camp.

Table 6 presents intent-to-treat estimates of the effect of receiving a camp invitation on endline test scores. In the pooled sample (Panel A), the invitation increased aggregate test scores by 0.06 standard deviations, with effects of 0.05 SD in mathematics and 0.05 SD in Spanish, all significant at the 1 percent level.

Results differ across cohorts in overall magnitude. In 2024 (Panel B), the aggregate effect of 0.05 SD is driven entirely by Spanish (0.07 SD, significant at the 5 percent level), with a near-zero effect on mathematics. In 2025 (Panel C), both subjects show significant gains, with a stronger effect in mathematics (0.06 SD) than in Spanish (0.04 SD). Despite these differences in the subject-specific patterns, formal tests of equality across cohorts fail to reject the null for all three outcomes (p-values of 0.98, 0.16, and 0.41 for aggregate, math, and Spanish, respectively).

The absence of a math effect in 2024 may reflect several factors. The 2024 curriculum did not differentiate by performance level, potentially limiting the program's ability to target instruction to students' skill levels in a subject where foundational gaps vary widely.

Further, teachers in focus groups mentioned they had to spend more time in Spanish than in math given than some students did not know how to read. The introduction of ability-based grouping and leveled workbooks in 2025, along with the additional instructional week, may have contributed to the improved math results, though as noted in Section 2.2, we cannot isolate the contribution of any single design change.

TABLE 6. Effect on Learning of Being Invited to the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.057***	0.050***	0.048***
	(0.01)	(0.02)	(0.02)
Observations	10,632	10,185	10,309
Control mean	-0.000	-0.000	-0.000
***Panel B**: 2024*			
Treatment	0.054*	0.005	0.072**
	(0.03)	(0.03)	(0.03)
Observations	2,633	2,282	2,368
Control mean	-0.000	-0.001	-0.000
***Panel C**: 2025*			
Treatment	0.056***	0.061***	0.040**
	(0.02)	(0.02)	(0.02)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.98	0.16	0.41

Note: This table shows intent-to-treat estimates of the effect of being invited to the remedial camp on endline test scores. Outcomes are standardized to the comparison group mean within each cohort. Column (1) reports results for aggregate test scores, Column (2) for mathematics, and Column (3) for Spanish. Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The row "2024 vs 2025" reports the p-value from a test of equality of treatment effects across cohorts, obtained from a regression that interacts the treatment indicator with a cohort indicator. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15 examines heterogeneous treatment effects by nationality, gender, and distance to the camp site. None of these interaction terms are statistically significant, indicating that the effect of the camp invitation does not vary meaningfully along these dimensions. Haitian students show a negative but insignificant interaction with treatment across all three outcomes (−0.04 SD for aggregate test scores, −0.04 for math, −0.03 for Spanish), suggesting that the camp was similarly effective for Haitian and non-Haitian students despite the differential attendance documented in Section 5.1. The pattern is similar for both cohorts (Table A16 and Table A17), although no treatment effects are statistically significant in 2024. In terms of heterogeneities by overage status, students with 3+ years overage have lower effects in Spanish during 2025 relatively to students with one year overage, -0.09 SD (Table A22).

Table A23 presents pooled estimates of heterogeneous effects of being invited to the camp, whjile Table A24 and Table A25 present the heterogeneous effects by cohorts. Fifth graders have a significant effect on math in 2025. Other heterogeneities and cohort-specific results are reported in Appendix Table A18 to Table A25.

Effect on Non-academic Outcomes of Being Invited to the Remedial Camp in 2025. We find no statistically significant effects of the summer camp invitation on any socioemotional outcome measure in 2025 (see Table 7).[6] Point estimates are close to zero across all dimensions, though the signs are suggestive: emotional regulation shows the largest positive (non-significant) point estimate (0.023 SD), consistent with qualitative evidence from teacher focus groups highlighting improved self-regulation among participants. Bullying victimization carries a small negative sign (−0.015), pointing in the desired direction of reduced victimization.

[6] see Appendix E for details on the questionnaire and variables' constructions.

TABLE 7. Effects on Non-Academic Outcomes of Being Invited to the Remedial Camp

	(1)	(2)	(3)	(4)	(5)	(6)
	PCA Index	Empathy	Emotional Regulation	Academic Self-Concept	Growth Mindset	Bullying Victimization
Treatment	-0.003	-0.003	0.023	-0.019	0.001	-0.015
	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	5,423	7,781	7,581	7,545	7,511	5,665
Control mean	0.000	-0.005	-0.011	-0.003	-0.002	0.018

Note: This table reports intent-to-treat (ITT) estimates of the effect of being invited to the remedial camp on socioemotional outcomes for the 2025 cohort. Column (1) reports effects on the first principal component index of five socioemotional indices (with bullying victimization reverse-coded so that higher values indicate better outcomes for all indices). Columns (2)–(6) report effects on each index separately. Each index is the mean of its standardized component items: Empathy (items 1, 2, 4, 5, 15), Emotional Regulation (items 3, 6), Academic Self-Concept (items 8, 9), Growth Mindset (items 11, 12), and Bullying Victimization (items 13, 14). Items 1 and 12 are reverse-coded. All regressions include randomization stratum fixed effects, grade fixed effects, and baseline math and Spanish test scores interacted with grade (with missing value imputation and flags) as controls. Standard errors clustered at the school level are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

5.3. Effects on Learning of Attending the Camp

Having established that the camp invitation significantly improved learning outcomes, we now turn to the average causal response (ACR) estimates, which recover the impact of camp attendance for compliers. To do so, we estimate the following system of equations:

$$shatt15_{iks} = \pi_0 + \pi_1 T_i + +\pi_g G_{ig} + \pi_{Y_0} Y_{iks0} + \sum_g \pi_{g,Y_0} G_{ig} \cdot Y_{iks0} + \delta_k + \nu^1_{iks} \tag{7}$$

$$Y_{iks1} = \gamma_0 + \gamma_1 \widehat{shatt15}_{iks} + \gamma_g G_{ig} + \gamma_{Y_0} Y_{iks0} + \sum_g \beta_{g,Y_0} G_{ig} \cdot Y_{iks0} + \delta_k + \nu^{2SLS}_{iks1} \tag{8}$$

The coefficient θ estimates the ACR of a one-unit increase in *AttendShare* — equivalent to 15 instructional days of camp — on endline test scores.

Table 8 and Figure 3 present the ACR estimates. In the pooled sample (Panel A), a full dose of 15 instructional days increases aggregate test scores by 0.10 SD, with effects of 0.09 SD in mathematics and 0.086 SD in Spanish, all significant at the 1 percent level. These magnitudes are roughly double the corresponding ITT estimates, consistent with the observed compliance rates (see Section 5.2).

TABLE 8. Effects on Learning of Attending the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.103***	0.090***	0.086***
	(0.03)	(0.03)	(0.03)
Observations	10,632	10,185	10,309
Control mean	-0.000	-0.000	-0.000
***Panel B**: 2024*			
Treatment	0.166*	0.016	0.222**
	(0.09)	(0.11)	(0.11)
Observations	2,633	2,282	2,368
Control mean	-0.000	-0.001	-0.000
***Panel C**: 2025*			
Treatment	0.090***	0.098***	0.066**
	(0.03)	(0.03)	(0.03)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.37	0.48	0.15

Note: This table shows average causal response estimates of the effect of attending the remedial camp on endline test scores. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15. The instrument is the randomly assigned invitation to attend the camp. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The last line reports the p-value from a test of equality of treatment effects across cohorts. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In 2024 (Panel B), the aggregate effect of 0.17 SD is driven by Spanish (0.22 SD), significant at the 5 percent level, with a near-zero effect on mathematics. The 2024 estimates are imprecise — standard errors are three to four times larger than in 2025 — reflecting both the smaller sample and the weaker first stage. In 2025 (Panel C), both subjects show significant effects: 0.1 SD in mathematics and 0.07 SD in Spanish. Tests of equality across cohorts fail to reject for all three outcomes (p-values of 0.37, 0.48, and 0.15).

These cross-cohort comparisons should be interpreted with care. Because comparison-group contamination differed markedly between years, the complier population is not identical across cohorts: in 2024, compliers are defined relative to a counterfactual in which 17.5 percent of non-invited students also attend at least one day, while in 2025 the counterfactual attendance rate is near zero. If the dose-response function is non-linear, the ACR estimates in each cohort may reflect causal effects at different margins of the dosage distribution rather than changes in program effectiveness. The failure to reject equality across cohorts is consistent with a roughly linear dose-response, under which the ACR does not depend on the distribution of compliers across dosage margins (Angrist and Imbens 1995), though the wide confidence intervals in 2024 limit the power of this test.

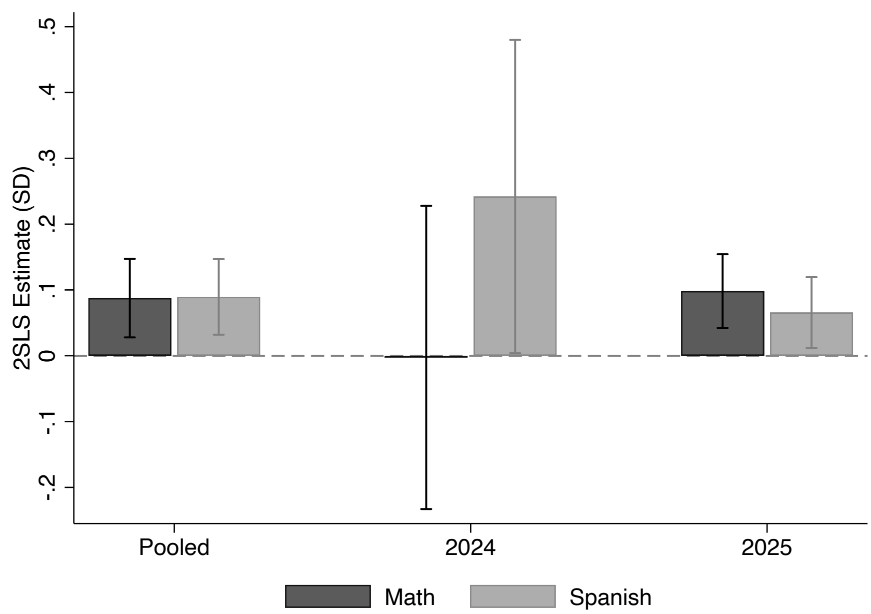

FIGURE 3. Effect on Learning of Attending the Remedial Camp

Note: This figure plots the average causal response of attending the remedial camp on endline test scores, separately for math and Spanish. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15, instrumented by the randomly assigned invitation. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification additionally includes cohort fixed effects. Results are shown for both remedial camps pooled, and separately for the 2024 and 2025 cohorts. The sample includes all students selected by the Ministry of Education for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Vertical bars represent 95 percent confidence intervals. Standard errors are clustered at the school level.

Effects of attending the camp on learning follow a similar pattern to the ITT estimates: the most overage students benefit less from Spanish instruction, with a significant negative effect for the 3+ years category in 2025 (Table A33). The grade 5 advantage in math also persists in the 2025 cohort (Table A34). No significant heterogeneity appears by nationality, gender, distance to camp, or baseline test score decile (Tables A26 to A29).

5.4. Attendance and Selection

Comparing OLS estimates from Equation 1 with the ACR estimates from Equation 4 provides an informal test of selection bias in the OLS relationship between camp attendance and test scores. Under the linearity assumption already imposed by OLS, and the further assumption that the average per-day causal effect for compliers equals the average per-day effect for the broader population of attendees, OLS and ACR identify the same structural parameter; under those conditions, any systematic difference between them reflects correlation between attendance and unobserved determinants of learning. Table 9 reports the two sets of estimates side-by-side, alongside the corresponding ITT.

In the pooled sample, OLS and ACR estimates are remarkably similar, leaving little room for OLS to be substantially biased by selection on unobservables. The pooled OLS estimate on aggregate test scores is 0.109 standard deviations, compared with an ACR of 0.103. A Wu–Hausman test of the equality between the OLS and 2SLS estimators fails to reject exogeneity of attendance in the pooled sample for average test scores ($\chi^2 = 0.03$, $p = 0.86$), math, and Spanish.[7]

A similar picture holds within each cohort. OLS estimates of the per-15-day effect are stable across cohorts, and although the ACR point estimates differ in level they cannot be statistically distinguished from one another. The 2024 cohort yields wider confidence intervals on both estimators, reflecting the weaker first-stage effect of the invitation that year, so the cohort-level comparison is less informative than the pooled one. Within each

[7] Two caveats temper a strong selection interpretation of the OLS–ACR comparison. First, the equivalence of OLS and ACR rests on homogeneity of per-day returns across student types: if compliers differ from always-takers in their per-day returns, the OLS–ACR gap conflates selection on attendance with treatment-effect heterogeneity across the complier and always-taker populations, and a small gap is consistent with selection bias and complier-population differences offsetting one another. Second, the OLS estimand under non-linear marginal returns averages a misspecified model over the full attendance distribution, while the ACR remains a properly weighted average of margin-specific effects; in that case, the OLS–ACR gap reflects functional-form misspecification in addition to any selection on unobservables.

cohort, Wu–Hausman tests likewise fail to reject exogeneity of attendance for any of the three outcomes.

Under the assumptions of linear marginal returns and homogeneity of per-day returns for different student types, the close correspondence between OLS and ACR is informative for the broader literature on voluntary remedial programs. Voluntary participation is a defining feature of most summer and after-school remedial programs, and observational estimates of their effects are routinely qualified by appeals to selection on unobservables. The pooled OLS–ACR similarity in this setting suggests that, for a government-led remedial program, observational estimates of the per-day return to camp attendance would have led to very similar conclusions to those obtained under random assignment. However, extrapolating this finding to remedial programs operated under very different recruitment, teacher training, enforcement, or participation regimes should be done carefully.

TABLE 9. Estimations of Remedial Camp Effects

	(1)	(2)	(3)
	OLS	ITT	ACR
***Panel A**: Pooled years*			
Average test scores	0.109***	0.057***	0.103***
	(0.01)	(0.01)	(0.03)
Math	0.104***	0.050***	0.090***
	(0.02)	(0.02)	(0.03)
Spanish	0.095***	0.048***	0.086***
	(0.01)	(0.02)	(0.03)
Observations	10,632	10,632	10,632
Control mean	-0.000	-0.000	-0.000
***Panel B**: 2024*			
Average test scores	0.102***	0.054*	0.166*
	(0.04)	(0.03)	(0.09)
Math	0.057	0.005	0.016
	(0.04)	(0.03)	(0.11)
Spanish	0.113***	0.072**	0.222**
	(0.04)	(0.03)	(0.11)
Observations	2,633	2,633	2,633
Control mean	-0.000	-0.000	-0.000
***Panel C**: 2025*			
Average test scores	0.110***	0.056***	0.090***
	(0.02)	(0.02)	(0.03)
Math	0.110***	0.061***	0.098***
	(0.02)	(0.02)	(0.03)
Spanish	0.092***	0.040**	0.066**
	(0.02)	(0.02)	(0.03)
Observations	7,999	7,999	7,999
Control mean	0.000	0.000	0.000
2024 vs 2025 (p-value)	0.84	0.98	0.37

Note: This table shows the different estimators for the remedial camp effects. Column (1) reports OLS estimates of endline test scores on camp attendance. Column (2) reports intent-to-treat (ITT) estimates of the effect of being randomly assigned to the remedial camp invitation. Column (3) reports the average causal response (ACR) using treatment assignment as an instrument for camp attendance. All specifications include stratification fixed effects, year fixed effects, and baseline test score controls interacted with grade and year. Baseline test scores set to zero when missing, with missing indicators included as controls. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5. Camp Centers Effectiveness

The pooled average causal response masks wide variation in effectiveness across camps. It is possible that some camps moved their students far more than others—and because we ran 52 separate camps in 2025, each with its own teachers and coordinators, we examine whether such differences exist and where the potential spread comes from. To do so, we follow Walters (2015), who studies the same kind of cross-site variation in the Head Start Impact Study, and we work at the camp level throughout.

The variation across camps is real, not sampling noise. When we let the invitation's effect differ by camp and test the differences jointly, they are overwhelmingly significant ($F \approx 85$, $p < 0.001$). Figure A1 shows the test graphically: if every camp shared the same effect, each camp's reduced form would line up with its first stage along a single ray through the origin, but the camps scatter far off it. Most importantly, the difference is not about attendance. A camp's first stage—how many days the invitation buys—tells us almost nothing about its reduced form ($r \approx 0.04$); what separates camps is not how many days students attend but how much each of those days is worth.

How much they differ is harder to measure than whether they differ. The natural estimate—the variance of each camp's average causal response—falls apart with so few students per camp, where the per-camp ratios have no finite moments and the implied variance is implausible, the same pathology Walters (2015) runs into and sets aside.[8] His fix, which we adopt, models the whole cross-camp distribution at once through a random-coefficients selection model (Table 10). It puts the standard deviation at 0.046 in intent-to-treat terms, about 0.074 per day attended—large next to a mean per-day return near 0.09, so a camp one standard deviation above average teaches nearly twice as much per day as the typical one.

What drives the gap is harder still, because almost nothing we observe explains it. Camp composition, size, attendance, and the surveyed measures of teacher practice leave the spread essentially untouched, and of all twenty candidate characteristics tested one at a time in Appendix Table A35, not one clears five percent significance, against the one expected by chance. Even Walters, who explains about a third of his variation, finds that the usual suspects do not drive it: teacher education and class size are unrelated to Head Start effectiveness, and only full-day scheduling and whether a center drew children from

[8]Applied to our data, the moment estimator of Walters (2015, Eq. 4) yields a cross-camp standard deviation of about 0.085 (unweighted) to 0.063 (precision-weighted) in intent-to-treat units, but its bootstrap confidence interval includes zero. We report it only as the benchmark Walters rejects.

TABLE 10. Cross-camp heterogeneity in effectiveness

Approach (estimand)	Cross-camp effect SD	Share expl. by inputs
Moment deconvolution, dofile 24 (ITT)	0.085 / 0.063[a]	0.12[b]
Structural random-coeff. mixed (ITT)	0.046[c]	≈0
same, rescaled to ACR units	0.074	–
Walters binary-Heckit benchmark (attend)	0.099 (0.020)	0.26 (0.204)
Walters (2015), cognitive (reference)	0.184	0.34

[a] unweighted / N-weighted deconvolved SD; 95% bootstrap CI touches 0 (magnitude imprecise). [b] R^2 of shrunken estimates on inputs (descriptive lower bound). [c] random invitation-slope SD; heterogeneity significant (LR $p = 0.039$), but inputs do not reduce the residual slope variance (share ≈0). Structural estimates use 52 camps; Walters used ~290 sites, so his variance/share are far better powered.

Notes: 2025 cohort, 52 camps; the outcome is aggregate ability in standard deviations. Each row reports a cross-camp standard deviation of effects and the share of its variance explained by observable inputs, under a different estimator; units are indicated in the row label (ITT, per invited child; ACR, per day of attendance; attend, the binary attend-versus-not margin). The moment deconvolution applies the Walters (2015) estimator (Eq. 4); applied to the per-camp IV estimates it returns a negative (hence uninformative) variance, ≈ -12, so it is computed on the reduced form instead, unweighted and precision-weighted. The random-coefficients estimate is a linear mixed model with a camp random intercept and a random invitation slope (restricted maximum likelihood); the binary-take-up row applies the Walters (2015) selection model to our data as a robustness check (take-up defined as attending at least one day). The final row reproduces Walters (2015)'s Head Start figures (cognitive skills, Spring 2003; 286 centers). Inputs span camp size, composition, attendance, teacher proactivism and practices, supervision, and survey coverage; superscript definitions appear in the table footnote.

home care rather than other preschools predict it. We cannot point even to that much; what makes a camp effective lies almost entirely in what we did not measure.

6. Cost-Effectiveness

The camp's learning gains invite a natural comparison: the learning a child gains from a year of regular schooling, the resource that grade retention consumes. We construct that benchmark, set the camp against it on both effect and cost, and express the two in common per-dollar terms, using the 2025 cohort and aggregate ability throughout.

The benchmark comes from the children who were not invited. Comparing their endline ability across grades, each additional grade corresponds to 0.47 standard deviations of aggregate ability—the progress a typical student makes over a school year without the camp.[9] At a public cost of about $2,700 per pupil-year, that year of schooling buys roughly

[9]This treats the cross-grade gradient among uninvited students as the gain from an additional year of schooling. It assumes baseline and endline ability share a common vertical scale, and that the gradient reflects schooling rather than age or cross-grade differences in the student population.

0.017 standard deviations of learning per $100 spent. This is the number the camp has to beat.

The program clears that bar comfortably. Nineteen days of camp raise aggregate ability by 0.06 standard deviations for every invited child and by 0.11 for those who attend—between an eighth and a quarter of a full school year, in a tenth of the school days. Matching each effect to its own cost—the intent-to-treat effect to the $234 spent per invited student, the attendance effect to the $450 spent per attending

How much of that advantage survives depends on how the camp's costs are reckoned. The $234 per invited student reflects under-subscription at some sites; costed instead at the $324 that full enrollment would imply, the camp returns about 0.018 per $100—roughly level with a year of schooling. The ranking is therefore robust in direction if not in size: the camp is at least as cost-effective as the schooling it would displace, and up to roughly 45 percent more so at the costs actually incurred.

Limitations: This analysis has three limitations. First, it assumes baseline and endline ability lie on a common vertical scale, which the cross-grade benchmark requires. Second, it sets a short, one-time intervention against a full year of schooling, so the per-dollar ranking speaks to the marginal value of remediation, not to the value of schooling as a whole. And finally, it presumes the two cost figures are measured on the same basis; if the camp's $234 captures only marginal delivery while the $2,700 per pupil-year is fully loaded, the comparison flatters the camp, and the true gap is narrower than the headline figure.

7. Conclusion

We show that a government-designed and government-led remedial summer camp can produce meaningful learning gains for low-performing primary students. Public school teachers, working with ministry materials and without specialized training beyond standard camp orientation, generated effects comparable to those of similar programs elsewhere—our pooled intent-to-treat effect of 0.06 SD sits within the range of Pratham's TaRL summer camps in Bihar, the closest existing comparator (Banerjee et al. 2016), and near the median

TABLE 11. Cost-effectiveness of the camp relative to a year of schooling (2025 cohort, aggregate ability).

	Year of school	Camp (per invited)	Camp (full capacity)	Camp (per attended)
Learning gain (SD, agg.)	0.466	0.059	0.059	0.109
Duration (days)	190	19	19	19
Cost (USD/student)	2700	233.58	323.82	450.26
CE (SD per 100 USD)	**0.017**	**0.025**	**0.018**	**0.024**

Year of school = control cross-grade theta gradient. Camp (per invited / full capacity) = ITT; camp (per attended) = ACR (full-camp completion) × avg. attendance share (≈0.90). Share = days/19.

Notes: 2025 cohort; the outcome is endline aggregate ability in standard deviations. Cost-effectiveness (bottom row) is the learning gain per $100 spent, matching each effect to its cost base: the year-of-schooling gain (the cross-grade ability gradient among uninvited students) to the per-pupil cost of a school year; the intent-to-treat effect to the cost per invited student; and the attendance effect to the cost per student who attended at least one day. The "full capacity" column recosts the per-invited effect at the per-student cost that full enrollment would imply. Costs are in USD per student; further definitions appear in the table footnote.

for education trials in low- and middle-income countries (Evans and Yuan 2022; Kraft 2020). The gains are smaller than those of high-dosage tutoring in the United States, but at a small fraction of the cost.

The effects are broadly uniform across the baseline test-score distribution, reaching the lowest- and highest-performing students the program serves alike. OLS and instrumented estimates of the per-day return coincide closely, pointing to a limited role for selection on unobservables—suggesting, under the assumptions of Section **??**, that observational evaluations of similar centrally administered programs may reach conclusions close to those under random assignment.

Effectiveness nonetheless varied widely across camps, and in ways observable inputs could not explain. Camps differed not in how much attendance they generated but in how productively each day of attendance translated into learning, and no measured camp, teacher, or student characteristic robustly predicted which camps worked—a caution for efforts to scale by replicating observable features rather than the unmeasured quality that appears to drive effectiveness.

The two-cohort design shows that implementation itself is improvable from within a government system. Between 2024 and 2025 the ministry introduced ability-based grouping, leveled workbooks, transportation, and shorter travel distances, and compliance rose sharply while comparison-group contamination nearly vanished.

Taken together, the results suggest that short, intensive remedial camps are a viable complement to the regular school year: they run during vacation without displacing instructional time, use existing teachers and infrastructure, and scale through the existing administrative apparatus. Several limitations remain. The one substantive exception to the otherwise uniform effects—smaller Spanish gains among students three or more years overage—marks a subgroup that may warrant curriculum adaptation. Attrition was high in 2024 and differentially favored the treated group, though this makes our estimates conservative. The many simultaneous changes between cohorts preclude attributing improvement to any single feature, and endline assessments fell at different post-camp intervals, so effects are measured over different persistence windows. Finally, we do not yet observe whether the gains persist into the following school year or translate into grade progression.

References

Angrist, Joshua D. and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association* 90 (430): 431–442.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, Noam and Rachael Meager. 2023. "Implementation matters: Generalizing treatment effects in education." *Available at SSRN 4487496* .

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India." *NBER Working Paper No. 22746* .

Borman, Geoffrey D and N Maritza Dowling. 2006. "Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial." *Educational Evaluation and Policy Analysis* 28 (1): 25–48.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas. 2020. "Experimental evidence on alternative policies to increase learning at scale." Tech. rep., National Bureau of Economic Research.

Evans, David K. and Fei Yuan. 2022. "How Big Are Effect Sizes in International Education Studies?" *Educational Evaluation and Policy Analysis* 44 (3): 532–540.

Gomes-Neto, Joao Batista and Eric A Hanushek. 1994. "Causes and consequences of grade repetition: Evidence from Brazil." *Economic development and cultural change* 43 (1): 117–148.

Goos, Mieke, Joana Pipa, and Francisco Peixoto. 2021. "Effectiveness of grade retention: A systematic review and meta-analysis." *Educational Research Review* 34: 100401.

Jacob, Brian A. and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *Review of Economics and Statistics* 86 (1): 226–244.

Kim, James S and David M Quinn. 2013. "The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions." *Review of Educational Research* 83 (3): 386–431.

Koppensteiner, Martin Foureaux. 2014. "Automatic Grade Promotion and Student Performance: Evidence from Brazil." *Journal of Development Economics* 107: 277–290.

Kraft, Matthew A. 2020. "Interpreting effect sizes of education interventions." *Educational researcher* 49 (4): 241–253.

LLECE. 2021. "Estudio Regional Comparativo y Explicativo (ERCE 2019): Reporte Nacional de Resultados – República Dominicana." Tech. rep., UNESCO Office Santiago, Regional Bureau for Education in Latin America and the Caribbean, Santiago. `https://unesdoc.unesco.org/ark:/48223/pf0000380257`.

Manacorda, Marco. 2012. "The Cost of Grade Retention." *Review of Economics and Statistics* 94 (2): 596–606.

Mariano, Louis T and Paco Martorell. 2013. "The academic effects of summer instruction and retention in New York City." *Educational Evaluation and Policy Analysis* 35 (1): 96–117.

Matsudaira, Jordan D. 2008. "Mandatory summer school and student achievement." *Journal of Econometrics* 142 (2): 829–850.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2024. "Adapting for Scale? Experimental Evidence on Technology-Aided Instruction in India." Working paper — VERIFY YEAR, AUTHORS, AND OUTLET.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2024. "The Promise of Tutoring for PreK–12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." *American Educational Research Journal* 61 (1): 74–107.

Walters, Christopher R. 2015. "Inputs in the production of early childhood human capital: Evidence from Head Start." *American Economic Journal: Applied Economics* 7 (4): 76–102.

World Bank. 2019. "Dominican Republic Learning Poverty Brief." Tech. rep., World Bank Group, Washington, DC. `https://documents1.worldbank.org/curated/en/994181637665163043/pdf/Dominican-Republic-Learning-Poverty-Brief-2019.pdf`.

Appendix A. Additional Tables and Figures

TABLE A1. Balance Table (2024)

	Comparison		Invited		Difference	
	Mean	SE	Mean	SE		
	(1)	(2)	(3)	(4)	(4)-(2)	P-value
Age (in years)	10.437	0.013	10.438	0.009	0.000	0.988
No overaged	0.259	0.005	0.250	0.003	-0.009	0.293
1 year overaged	0.278	0.006	0.287	0.004	0.010	0.357
2 year overaged	0.208	0.005	0.210	0.004	0.002	0.795
At least 3 years overaged	0.256	0.005	0.253	0.003	-0.003	0.653
Underlying ability in math	-0.002	0.010	0.006	0.007	0.007	0.650
Underlying ability in spanish	-0.001	0.011	-0.017	0.007	-0.016	0.373
Haitian	0.106	0.004	0.102	0.003	-0.004	0.591
Missing nationality	0.104	0.004	0.098	0.003	-0.006	0.390
Female	0.332	0.008	0.335	0.005	0.003	0.803
Missing gender	0.102	0.004	0.096	0.003	-0.006	0.361
Number students	2,061		3,023			
p-value from joint significance test						0.968

Note: This table shows baseline characteristics by treatment status. We report means, standard errors, and pairwise differences between treatment and comparison groups. Estimates are from an OLS regression of each baseline variable on the treatment indicator, including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study in 2024. For students with missing baseline assessments, we impute the within-cohort mean (zero by construction since baseline scores are standardized within cohort). Standard errors are clustered at the school level. We also present p-values from a joint significance test of all variables included in the balance test. * p < 0.10, ** p < 0.05, *** p < 0.01. Go to Table 2.

TABLE A2. Balance Table (Pooled)

	Comparison		Invited		Difference	
	Mean (1)	SE (2)	Mean (3)	SE (4)	(4)-(2)	P-value
Age (in years)	10.992	0.012	11.001	0.008	0.008	0.668
No overaged	0.080	0.002	0.078	0.001	-0.003	0.293
1 year overaged	0.348	0.004	0.346	0.003	-0.001	0.830
2 year overaged	0.283	0.004	0.284	0.003	0.001	0.907
At least 3 years overaged	0.289	0.004	0.292	0.003	0.003	0.627
Underlying score in math	-0.001	0.007	0.010	0.005	0.011	0.357
Underlying score in spanish	-0.000	0.008	0.001	0.005	0.001	0.927
Haitian	0.174	0.003	0.175	0.002	0.001	0.822
Missing nationality	0.040	0.001	0.037	0.001	-0.003	0.292
Female	0.355	0.004	0.356	0.003	0.001	0.926
Missing gender	0.032	0.001	0.030	0.001	-0.002	0.363
Number students	6,579		9,816			
p-value from joint significance test						0.991

Note: This table shows baseline characteristics by treatment status.We report means, standard errors, and pairwise differences between treatment and comparison groups. Estimates are from an OLS regression of each baseline variable on the treatment indicator, including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study across the 2024 and 2025 cohorts. For students with missing baseline assessments, we impute the within-cohort mean (zero by construction since baseline scores are standardized within cohort). Standard errors are clustered at the school level.We also present p-values from a joint significance test of all variables included in the balance test. * p < 0.10, ** p < 0.05, *** p < 0.01. Go to Table 2.

TABLE A3. Attrition in Endline Tests (2024)

	Spanish		Math	
	(1)	(2)	(1)	(2)
Treatment	-0.032**	-0.030**	-0.047***	-0.048***
	(0.014)	(0.014)	(0.012)	(0.013)
Age (in years)		0.059***		0.049***
		(0.019)		(0.015)
No overaged		0.050		0.060*
		(0.031)		(0.030)
1 year overaged		[base]		[base]
2 year overaged		-0.075**		-0.047
		(0.032)		(0.034)
At least 3 years overaged		-0.162***		-0.088*
		(0.056)		(0.053)
Underlying ability in math		0.001		-0.007
		(0.012)		(0.013)
Underlying ability in spanish		-0.020		-0.038***
		(0.012)		(0.014)
Hatian		-0.000		-0.010
		(0.026)		(0.026)
Female		-0.025		-0.016
		(0.015)		(0.016)
Control mean	0.522	0.522	0.547	0.547
Number observations	5,084	5,084	5,084	5,084
Number schools	279	279	279	279

Note: This table examines differential attrition by treatment status. The dependent variable is an indicator for whether the student's endline test score is missing. Columns (1) and (2) report results for Spanish; Columns (3) and (4) report results for mathematics. Odd-numbered columns report the unconditional difference in attrition between treatment and comparison groups; even-numbered columns add baseline covariates as controls. Estimates are from an OLS regression including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study during 2024, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Go to Table 3.

TABLE A4. Attrition in Endline Tests (2025)

	Spanish		Math	
	(1)	(2)	(1)	(2)
Treatment	-0.000	0.019**	-0.001	0.019**
	(0.008)	(0.008)	(0.008)	(0.008)
Age (in years)		0.027***		0.024***
		(0.005)		(0.005)
No overaged		0.000		0.000
		(.)		(.)
1 year overaged		[base]		[base]
2 year overaged		0.041**		0.049**
		(0.019)		(0.020)
At least 3 years overaged		-0.009		-0.006
		(0.006)		(0.006)
Underlying ability in math		-0.020***		-0.022***
		(0.006)		(0.006)
Underlying ability in spanish		0.015		0.020
		(0.015)		(0.015)
Hatian		0.002		-0.003
		(0.009)		(0.009)
Female		0.094***		0.096***
		(0.022)		(0.021)
Control mean	0.298	0.298	0.302	0.302
Number observations	11,311	11,311	11,311	11,311
Number schools	320	320	320	320

Note: This table examines differential attrition by treatment status. The dependent variable is an indicator for whether the student's endline test score is missing. Columns (1) and (2) report results for Spanish; Columns (3) and (4) report results for mathematics. Odd-numbered columns report the unconditional difference in attrition between treatment and comparison groups; even-numbered columns add baseline covariates as controls. Estimates are from an OLS regression including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study during 2025, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Go to Table 3.

TABLE A5. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance (Pooled)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
***Panel A**: Heterogeneity by Gender*							
Female	-0.000	-0.007	0.000	-0.003	-0.002	-0.006	0.003
	(0.02)	(0.25)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Treatment	0.451***	6.759***	0.431***	0.411***	0.385***	0.303***	0.272***
	(0.01)	(0.20)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Female	0.003	0.049	-0.003	0.001	-0.001	0.014	0.008
	(0.01)	(0.19)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.017	0.248	0.040	0.047	0.055	-0.045	0.085
Observations	15,859	15,859	15,859	15,859	15,859	15,859	11,311
***Panel B**: Heterogeneity by Nationality*							
Haitian	0.015	0.232	-0.002	0.019	0.021	0.015	0.001
	(0.03)	(0.46)	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)
Treatment	0.469***	7.029***	0.446***	0.427***	0.400***	0.320***	0.284***
	(0.01)	(0.20)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Haitian	-0.079***	-1.185***	-0.069***	-0.073***	-0.072***	-0.050***	-0.031*
	(0.02)	(0.33)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.001	0.019	0.039	0.028	0.035	-0.066	0.085
Observations	15,730	15,730	15,730	15,730	15,730	15,730	11,191
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*							
Treatment	0.440***	6.594***	0.420***	0.402***	0.375***	0.295***	0.276***
	(0.01)	(0.20)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Zero Distance	0.033	0.500	0.038	0.032	0.029	0.038*	-0.008
	(0.03)	(0.48)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Constant	0.011	0.162	0.036	0.041	0.044	-0.052	0.086
Observations	16,395	16,395	16,395	16,395	16,395	16,395	11,311

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), and whether the camp is located in the student's school from the previous academic year (Panel C). Distance between the student's school and the camp site is excluded from this specification because its support differs across cohorts (0–25 km in 2024 vs. 0–10 km in 2025); heterogeneity by distance is examined separately for each cohort (see Tables A6 and A7) and in a pooled sample restricted to students within 9 km in Appendix G. Columns report results for the share of days attended out of 15 (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 3–7). All specifications include stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The main effect of camp at student's school is absorbed by the stratum fixed effects as this variable does not vary within strata. The sample includes all students selected by the Ministry of Education to participate in the study across the 2024 and 2025 remedial camps, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Go to Figure 1.

TABLE A6. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance (2024)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) = 15 days
***Panel A**: Heterogeneity by Gender*						
Female	-0.000	-0.002	0.000	-0.003	-0.001	-0.006
	(0.02)	(0.28)	(0.02)	(0.02)	(0.02)	(0.01)
Treatment	0.267***	3.998***	0.355***	0.321***	0.268***	0.067***
	(0.02)	(0.35)	(0.03)	(0.03)	(0.03)	(0.01)
Treatment x Female	0.001	0.012	-0.014	0.003	-0.001	0.026
	(0.03)	(0.42)	(0.03)	(0.03)	(0.03)	(0.02)
Constant	0.198	2.964	0.234	0.222	0.210	0.059
Observations	4,582	4,582	4,582	4,582	4,582	4,582
***Panel B**: Heterogeneity by Nationality*						
Haitian	0.015	0.218	-0.003	0.018	0.019	0.015
	(0.03)	(0.52)	(0.04)	(0.04)	(0.04)	(0.02)
Treatment	0.274***	4.108***	0.355***	0.329***	0.276***	0.080***
	(0.02)	(0.33)	(0.03)	(0.03)	(0.02)	(0.01)
Treatment x Haitian	-0.056	-0.839	-0.040	-0.051	-0.067	-0.027
	(0.05)	(0.69)	(0.06)	(0.06)	(0.05)	(0.03)
Constant	0.196	2.940	0.234	0.218	0.208	0.056
Observations	4,574	4,574	4,574	4,574	4,574	4,574
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*						
Treatment	0.256***	3.835***	0.338***	0.312***	0.257***	0.062***
	(0.02)	(0.32)	(0.03)	(0.02)	(0.02)	(0.01)
Treatment x Zero Distance	0.062	0.928	0.055	0.053	0.055	0.091***
	(0.06)	(0.92)	(0.07)	(0.07)	(0.06)	(0.03)
Constant	0.187	2.798	0.222	0.211	0.196	0.052
Observations	5,084	5,084	5,084	5,084	5,084	5,084
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*						
Treatment	0.274***	4.114***	0.361***	0.330***	0.276***	0.080***
	(0.03)	(0.39)	(0.03)	(0.03)	(0.03)	(0.01)
Treatment x Distance	-0.003	-0.048	-0.005	-0.003	-0.004	-0.002
	(0.00)	(0.07)	(0.01)	(0.01)	(0.00)	(0.00)
Constant	0.187	2.798	0.222	0.210	0.196	0.052
Observations	5,084	5,084	5,084	5,084	5,084	5,084

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance for the 2024 cohort. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for the share of days attended out of 15 (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, or 15 days (Columns 3–6). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 1.

TABLE A7. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance (2025)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
***Panel A**: Heterogeneity by Gender*							
Female	0.004	0.054	-0.000	0.002	0.004	0.004	0.003
	(0.01)	(0.12)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment	0.525***	7.873***	0.462***	0.447***	0.433***	0.399***	0.272***
	(0.02)	(0.25)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Female	0.004	0.058	0.002	-0.001	-0.002	0.009	0.008
	(0.01)	(0.22)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.170	2.546	0.145	0.148	0.142	0.130	0.085
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311
***Panel B**: Heterogeneity by Nationality*							
Haitian	-0.005	-0.072	-0.007	-0.005	-0.004	-0.005	0.001
	(0.02)	(0.23)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment	0.548***	8.215***	0.482***	0.466***	0.450***	0.417***	0.284***
	(0.02)	(0.25)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Haitian	-0.088***	-1.317***	-0.081***	-0.082***	-0.074***	-0.059***	-0.031
	(0.03)	(0.38)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.173	2.588	0.147	0.151	0.145	0.133	0.085
Observations	11,191	11,191	11,191	11,191	11,191	11,191	11,191
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*							
Treatment	0.522***	7.835***	0.457***	0.443***	0.429***	0.399***	0.276***
	(0.02)	(0.26)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment x Zero Distance	0.020	0.304	0.030	0.022	0.017	0.014	-0.008
	(0.04)	(0.57)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Constant	0.172	2.573	0.145	0.149	0.144	0.132	0.086
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*							
Treatment	0.513***	7.698***	0.455***	0.439***	0.424***	0.393***	0.249***
	(0.02)	(0.34)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment x Distance	0.009	0.134	0.006	0.005	0.005	0.006	0.018*
	(0.01)	(0.18)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.171	2.571	0.145	0.149	0.144	0.132	0.086
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance for the 2025 cohort. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for the share of days attended out of 15 (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 3–7). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. In 2025, the reduction in maximum distance to camp sites to less than 5 kilometers substantially reduced variation in the distance variable. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Figure 1.

TABLE A8. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Baseline Test Score Decile (Pooled)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
Treatment	0.588***	8.825***	0.539***	0.528***	0.502***	0.428***	0.340***
	(0.03)	(0.48)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Treatment × D2	0.010	0.143	0.026	0.015	0.006	-0.026	0.030
	(0.03)	(0.52)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Treatment × D3	-0.004	-0.056	-0.001	-0.004	-0.005	-0.013	-0.001
	(0.04)	(0.55)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Treatment × D4	-0.040	-0.599	-0.026	-0.029	-0.032	-0.049	-0.032
	(0.04)	(0.57)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Treatment × D5	0.018	0.270	0.033	0.016	0.009	0.010	0.017
	(0.04)	(0.60)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Treatment × D6	-0.051	-0.764	-0.037	-0.047	-0.048	-0.060*	-0.018
	(0.04)	(0.56)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Treatment × D7	-0.047	-0.712	-0.040	-0.051	-0.052	-0.022	-0.016
	(0.04)	(0.55)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Treatment × D8	-0.059	-0.892	-0.044	-0.059*	-0.055	-0.049	-0.027
	(0.04)	(0.57)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Treatment × D9	-0.055	-0.819	-0.037	-0.045	-0.045	-0.051	-0.033
	(0.04)	(0.59)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Treatment × D10	-0.055	-0.818	-0.035	-0.048	-0.037	-0.085**	-0.034
	(0.04)	(0.64)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Joint F-test (D2–D10)	1.59	1.59	1.55	1.65	1.57	1.39	0.72
	0.059	0.059	0.068	0.046	0.063	0.130	0.692
Observations	11,217	11,217	11,217	11,217	11,217	11,217	8,497

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by baseline test score. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the within-cohort baseline test score distribution. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). The specification includes stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. The sample is restricted to students for whom baseline test scores are available. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 5.

TABLE A9. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Baseline Test Score Decile (2024)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) = 15 days
Treatment	0.334***	5.014***	0.426***	0.406***	0.352***	0.123***
	(0.05)	(0.81)	(0.07)	(0.06)	(0.06)	(0.05)
Treatment × D2	0.031	0.470	0.076	0.054	0.014	-0.070
	(0.07)	(1.03)	(0.08)	(0.08)	(0.08)	(0.06)
Treatment × D3	0.044	0.662	0.050	0.039	0.007	0.041
	(0.06)	(0.97)	(0.08)	(0.07)	(0.07)	(0.05)
Treatment × D4	-0.060	-0.893	-0.056	-0.071	-0.080	-0.036
	(0.07)	(1.05)	(0.08)	(0.08)	(0.08)	(0.06)
Treatment × D5	0.024	0.366	0.045	0.022	-0.002	0.007
	(0.07)	(1.05)	(0.09)	(0.09)	(0.08)	(0.06)
Treatment × D6	-0.020	-0.297	-0.054	-0.043	-0.011	-0.041
	(0.07)	(1.06)	(0.08)	(0.08)	(0.08)	(0.06)
Treatment × D7	-0.150*	-2.245*	-0.169*	-0.172**	-0.170*	-0.046
	(0.08)	(1.15)	(0.09)	(0.09)	(0.09)	(0.06)
Treatment × D8	-0.051	-0.758	-0.049	-0.094	-0.072	-0.010
	(0.07)	(1.02)	(0.08)	(0.08)	(0.08)	(0.06)
Treatment × D9	0.030	0.449	0.026	0.030	0.023	0.009
	(0.07)	(1.03)	(0.08)	(0.08)	(0.08)	(0.06)
Treatment × D10	0.090	1.349	0.091	0.073	0.106	-0.045
	(0.08)	(1.18)	(0.09)	(0.09)	(0.09)	(0.06)
Joint F-test (D2–D10)	1.72	1.72	1.64	1.82	1.47	1.20
	0.086	0.086	0.106	0.067	0.163	0.296
Observations	2,720	2,720	2,720	2,720	2,720	2,720

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by baseline test score for the 2024 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the baseline test score distribution. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, or 15 days (Columns 2–5). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. The sample is restricted to students for whom baseline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 5.

TABLE A10. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Baseline Test Score Decile (2025)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
Treatment	0.670***	10.045***	0.575***	0.567***	0.550***	0.525***	0.340***
	(0.04)	(0.56)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Treatment × D2	0.003	0.038	0.011	0.002	0.003	-0.013	0.030
	(0.04)	(0.62)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Treatment × D3	-0.019	-0.286	-0.017	-0.018	-0.009	-0.030	-0.001
	(0.04)	(0.64)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D4	-0.034	-0.505	-0.017	-0.016	-0.017	-0.053	-0.032
	(0.04)	(0.66)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D5	0.016	0.239	0.030	0.014	0.013	0.011	0.017
	(0.05)	(0.69)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D6	-0.061	-0.913	-0.032	-0.048	-0.059*	-0.066*	-0.018
	(0.04)	(0.66)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D7	-0.015	-0.221	0.001	-0.012	-0.015	-0.014	-0.016
	(0.04)	(0.64)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D8	-0.062	-0.935	-0.043	-0.047	-0.050	-0.061	-0.027
	(0.04)	(0.67)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D9	-0.082*	-1.225*	-0.058	-0.069*	-0.067*	-0.070*	-0.033
	(0.05)	(0.70)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Treatment × D10	-0.101**	-1.512**	-0.075*	-0.087**	-0.083**	-0.098**	-0.034
	(0.05)	(0.74)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Joint F-test (D2–D10)	1.31	1.31	1.31	1.31	1.44	1.67	0.72
	0.229	0.229	0.229	0.229	0.171	0.095	0.692
Observations	8,497	8,497	8,497	8,497	8,497	8,497	8,497

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by baseline test score for the 2025 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the baseline test score distribution. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. The sample is restricted to students for whom baseline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 5.

TABLE A11. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Overage Status (2024)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) = 15 days
Treatment (β_T)	0.281***	4.210***	0.364***	0.340***	0.280***	0.059***
	(0.02)	(0.30)	(0.02)	(0.02)	(0.02)	(0.01)
Treatment × No overage ($\beta_{T,0}$)	0.033	0.497	0.057	0.042	0.026	0.034
	(0.03)	(0.45)	(0.04)	(0.04)	(0.03)	(0.02)
Treatment × 2 years overage ($\beta_{T,2}$)	-0.058*	-0.876*	-0.072*	-0.067*	-0.059*	0.016
	(0.03)	(0.47)	(0.04)	(0.04)	(0.04)	(0.02)
Treatment × 3+ years overage ($\beta_{T,3}$)	-0.051*	-0.767*	-0.070**	-0.070**	-0.038	0.014
	(0.03)	(0.44)	(0.03)	(0.03)	(0.03)	(0.02)
Joint F-test	4.12	4.12	5.89	4.77	2.37	0.79
	0.006	0.006	0.001	0.003	0.069	0.501
$\beta_T + \beta_{T,0}$	0.314	4.707	0.421	0.382	0.306	0.092
	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,2}$	0.222	3.334	0.292	0.273	0.221	0.074
	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,3}$	0.230	3.443	0.294	0.270	0.242	0.072
	0.000	0.000	0.000	0.000	0.000	0.000
Observations	5,084	5,084	5,084	5,084	5,084	5,084

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by overage status for the 2024 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for overage categories: not overage, one year overage, two years overage, and three or more years overage. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, or 15 days (Columns 2–5). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 2.

TABLE A12. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Overage Status (2025)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
Treatment (β_T)	0.552***	8.287***	0.489***	0.474***	0.457***	0.420***	0.259***
	(0.01)	(0.21)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Treatment × 2 years overage ($\beta_{T,2}$)	0.003	0.052	0.000	-0.001	-0.002	0.008	0.039**
	(0.02)	(0.31)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment × 3+ years overage ($\beta_{T,3}$)	-0.088***	-1.314***	-0.085***	-0.085***	-0.077***	-0.066***	0.012
	(0.02)	(0.31)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Joint F-test	11.49	11.49	14.64	14.61	11.94	10.08	3.12
	0.000	0.000	0.000	0.000	0.000	0.000	0.044
$\beta_T + \beta_{T,2}$	0.556	8.339	0.489	0.472	0.455	0.428	0.298
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,3}$	0.465	6.972	0.405	0.389	0.379	0.354	0.272
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by overage status for the 2025 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for overage categories: one year overage, two years overage, and three or more years overage. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 2.

TABLE A13. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Grade (2024)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) = 15 days
Treatment (β_T)	0.280***	4.204***	0.365***	0.343***	0.279***	0.083***
	(0.02)	(0.33)	(0.03)	(0.03)	(0.02)	(0.01)
Treatment × Grade 4 ($\beta_{T,4}$)	-0.046*	-0.687*	-0.054*	-0.053*	-0.044	-0.028
	(0.03)	(0.38)	(0.03)	(0.03)	(0.03)	(0.02)
Treatment × Grade 5 ($\beta_{T,5}$)	-0.013	-0.190	-0.015	-0.033	-0.009	-0.002
	(0.03)	(0.44)	(0.04)	(0.03)	(0.03)	(0.02)
Joint F-test	1.86	1.86	1.71	1.68	1.42	1.35
	0.158	0.158	0.184	0.189	0.242	0.262
$\beta_T + \beta_{T,4}$	0.234	3.517	0.311	0.290	0.235	0.055
	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,5}$	0.268	4.014	0.350	0.310	0.271	0.080
	0.000	0.000	0.000	0.000	0.000	0.000
Observations	5,084	5,084	5,084	5,084	5,084	5,084

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. Column 6 is available only for the 2025 cohort, as the 2024 camp ran for 15 days. The sample includes all students selected by the Ministry of Education to participate in the study, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 5.

TABLE A14. Heterogeneous Effects of Being Invited to the Remedial Camp on Attendance by Grade (2025)

	(1) Share att.	(2) Total Days	(3) >= 1 days	(4) >= 5 days	(5) >= 10 days	(6) >= 15 days	(7) = 19 days
Treatment (β_T)	0.573***	8.600***	0.499***	0.484***	0.469***	0.446***	0.309***
	(0.02)	(0.32)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment × Grade 4 ($\beta_{T,4}$)	-0.058***	-0.867***	-0.037**	-0.043**	-0.047**	-0.056***	-0.047***
	(0.02)	(0.33)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Treatment × Grade 5 ($\beta_{T,5}$)	-0.069***	-1.035***	-0.058***	-0.056***	-0.052***	-0.062***	-0.046***
	(0.02)	(0.30)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Joint F-test	6.21	6.21	5.54	5.47	5.09	7.69	5.08
	0.002	0.002	0.004	0.005	0.007	0.001	0.007
$\beta_T + \beta_{T,4}$	0.516	7.733	0.462	0.440	0.422	0.389	0.262
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_T + \beta_{T,5}$	0.504	7.565	0.441	0.428	0.417	0.384	0.264
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311

Note: This table shows heterogeneous effects of being invited to the remedial camp on attendance by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators. Columns report results for total days attended (Column 1) and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 2–6). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. Column 6 is available only for the 2025 cohort, as the 2024 camp ran for 15 days. The sample includes all students selected by the Ministry of Education to participate in the study, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 5.

TABLE A15. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp (Pooled)

	(1) Aggregate Ability	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	-0.042	-0.087	0.120***
	(0.07)	(0.07)	(0.03)
Treatment	0.047**	0.043*	0.038*
	(0.02)	(0.02)	(0.02)
Treatment x Female	0.029	0.021	0.029
	(0.03)	(0.03)	(0.04)
Constant	-0.217	-0.315	-0.216
Observations	10,439	10,029	10,132
***Panel B**: Heterogeneity by Nationality*			
Haitian	0.111***	-0.085	-0.045
	(0.04)	(0.09)	(0.11)
Treatment	0.064***	0.059***	0.054***
	(0.02)	(0.02)	(0.02)
Treatment x Haitian	-0.036	-0.040	-0.034
	(0.04)	(0.05)	(0.05)
Constant	-0.290	-0.198	-0.114
Observations	10,395	9,985	10,092
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Treatment	0.051***	0.042**	0.046***
	(0.02)	(0.02)	(0.02)
Treatment x Zero Distance	0.038	0.043	0.025
	(0.05)	(0.05)	(0.05)
Constant	-0.300	-0.381	-0.217
Observations	10,795	10,357	10,477

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 6.

TABLE A16. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp (2024)

	(1) Aggregate Ability	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	-0.039	-0.090	0.056
	(0.07)	(0.08)	(0.08)
Treatment	0.064	0.010	0.090
	(0.05)	(0.06)	(0.06)
Treatment x Female	-0.009	0.010	-0.037
	(0.10)	(0.10)	(0.12)
Constant	0.099	0.053	0.079
Observations	2,437	2,123	2,188
***Panel B**: Heterogeneity by Nationality*			
Haitian	-0.089	-0.073	-0.052
	(0.11)	(0.10)	(0.13)
Treatment	0.067	0.018	0.086*
	(0.04)	(0.05)	(0.05)
Treatment x Haitian	-0.059	-0.034	-0.096
	(0.13)	(0.15)	(0.14)
Constant	0.091	0.023	0.106
Observations	2,433	2,119	2,184
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Treatment	0.038	-0.010	0.055
	(0.04)	(0.04)	(0.04)
Treatment x Zero Distance	0.127	0.126	0.135
	(0.10)	(0.13)	(0.12)
Constant	0.063	0.013	0.067
Observations	2,793	2,451	2,533
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*			
Treatment	0.030	-0.034	0.040
	(0.04)	(0.05)	(0.05)
Treatment x Distance	0.008	0.014	0.012
	(0.01)	(0.01)	(0.01)
Constant	0.061	0.011	0.066
Observations	2,793	2,451	2,533

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores for the 2024 cohort. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are largely absorbed by the stratum fixed effects as these variables do not vary within strata. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 6.

TABLE A17. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp (2025)

	(1) Aggregate Ability	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	0.037	-0.052*	0.119***
	(0.03)	(0.03)	(0.03)
Treatment	0.041*	0.051**	0.024
	(0.02)	(0.02)	(0.02)
Treatment x Female	0.040	0.025	0.048
	(0.03)	(0.04)	(0.03)
Constant	-0.126	-0.132	-0.109
Observations	8,002	7,906	7,944
***Panel B**: Heterogeneity by Nationality*			
Haitian	0.110***	0.167***	0.041
	(0.04)	(0.04)	(0.04)
Treatment	0.062***	0.070***	0.045**
	(0.02)	(0.02)	(0.02)
Treatment x Haitian	-0.031	-0.045	-0.018
	(0.04)	(0.05)	(0.04)
Constant	-0.137	-0.194	-0.069
Observations	7,962	7,866	7,908
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Treatment	0.054***	0.057***	0.043**
	(0.02)	(0.02)	(0.02)
Treatment x Zero Distance	0.008	0.020	-0.011
	(0.05)	(0.05)	(0.05)
Constant	-0.109	-0.153	-0.057
Observations	8,002	7,906	7,944
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*			
Treatment	0.038	0.050*	0.012
	(0.03)	(0.03)	(0.03)
Treatment x Distance	0.012	0.007	0.020
	(0.01)	(0.01)	(0.01)
Constant	-0.109	-0.154	-0.057
Observations	8,002	7,906	7,944

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores for the 2025 cohort. Each panel reports estimates from a separate OLS regression in which the treatment indicator is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. In 2025, the reduction in maximum distance to camp sites to less than 5 kilometers substantially reduced variation in the distance variables. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 6.

TABLE A18. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Baseline Test Score Decile (Pooled)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Treatment	0.074	0.093*	0.075
	(0.05)	(0.05)	(0.06)
Treatment × D2	0.030	-0.029	0.003
	(0.07)	(0.08)	(0.08)
Treatment × D3	0.006	-0.005	0.001
	(0.07)	(0.08)	(0.08)
Treatment × D4	0.029	-0.071	0.150*
	(0.07)	(0.07)	(0.09)
Treatment × D5	0.036	0.028	-0.013
	(0.08)	(0.09)	(0.08)
Treatment × D6	0.028	0.024	-0.035
	(0.07)	(0.09)	(0.08)
Treatment × D7	0.024	0.013	0.011
	(0.07)	(0.08)	(0.08)
Treatment × D8	0.032	-0.108	0.049
	(0.07)	(0.08)	(0.08)
Treatment × D9	-0.008	0.043	-0.033
	(0.07)	(0.08)	(0.07)
Treatment × D10	-0.049	-0.028	-0.072
	(0.07)	(0.08)	(0.08)
Joint F-test (D2–D10)	0.39	0.89	1.05
p-value	0.989	0.588	0.402
Observations	8,315	7,462	7,800

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by baseline test score. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the within-cohort baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores are not included as a separate control as it enters the specification through the decile interaction terms. The bottom panel reports a joint F-test of equality across all decile interactions (D2–D10) and the corresponding p-value. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 6.

TABLE A19. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Baseline Test Score Decile (2024)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Treatment	0.134	0.062	-0.022
	(0.10)	(0.15)	(0.17)
Treatment × D2	-0.193	-0.030	-0.010
	(0.19)	(0.22)	(0.23)
Treatment × D3	-0.002	0.035	0.133
	(0.18)	(0.21)	(0.20)
Treatment × D4	-0.026	-0.039	0.162
	(0.14)	(0.20)	(0.23)
Treatment × D5	-0.016	-0.141	0.249
	(0.15)	(0.24)	(0.21)
Treatment × D6	-0.089	-0.328	-0.012
	(0.17)	(0.22)	(0.22)
Treatment × D7	-0.174	-0.128	0.328
	(0.17)	(0.23)	(0.24)
Treatment × D8	-0.069	-0.276	0.154
	(0.18)	(0.24)	(0.23)
Treatment × D9	-0.083	0.160	-0.005
	(0.16)	(0.24)	(0.23)
Treatment × D10	-0.154	0.078	0.031
	(0.18)	(0.23)	(0.23)
Joint F-test (D2–D10)	0.32	0.88	0.66
p-value	0.966	0.546	0.746
Observations	1,619	1,160	1,277

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by baseline test score for the 2024 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores is not included as a separate control as it enters the specification through the decile interaction terms. The bottom panel reports a joint F-test of equality across all decile interactions (D2–D10) and the corresponding p-value. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 6.

TABLE A20. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Baseline Test Score Decile (2025)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Treatment	0.059	0.098*	0.094
	(0.06)	(0.06)	(0.07)
Treatment × D2	0.084	-0.029	0.006
	(0.08)	(0.08)	(0.09)
Treatment × D3	0.008	-0.012	-0.024
	(0.08)	(0.09)	(0.09)
Treatment × D4	0.042	-0.077	0.147
	(0.08)	(0.08)	(0.10)
Treatment × D5	0.049	0.060	-0.065
	(0.09)	(0.09)	(0.09)
Treatment × D6	0.056	0.089	-0.040
	(0.08)	(0.09)	(0.09)
Treatment × D7	0.072	0.039	-0.051
	(0.08)	(0.09)	(0.08)
Treatment × D8	0.056	-0.077	0.029
	(0.08)	(0.09)	(0.08)
Treatment × D9	0.010	0.022	-0.038
	(0.07)	(0.08)	(0.08)
Treatment × D10	-0.023	-0.048	-0.092
	(0.08)	(0.08)	(0.08)
Joint F-test (D2–D10)	0.41	0.78	1.37
p-value	0.927	0.636	0.200
Observations	6,696	6,302	6,523

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by baseline test score for the 2025 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for each decile of the baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores is not included as a separate control as it enters the specification through the decile interaction terms. The bottom panel reports a joint F-test of equality across all decile interactions (D2–D10) and the corresponding p-value. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 6.

TABLE A21. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Students' Overage Status (2024)

	(1) Aggregate	(2) Math	(3) Spanish
Treatment	0.038	0.045	0.045
	(0.06)	(0.07)	(0.06)
Treatment × No overage	0.113	-0.009	0.132
	(0.08)	(0.08)	(0.09)
Treatment × 2 years overage	-0.032	-0.097	-0.022
	(0.09)	(0.12)	(0.09)
Treatment × 3+ years overage	-0.044	-0.089	-0.020
	(0.09)	(0.11)	(0.10)
Joint F-test	1.85	0.44	1.17
	0.139	0.721	0.323
Observations	2,633	2,282	2,368

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by overage status for the 2024 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for overage categories, with one year of overage as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 6.

TABLE A22. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Students' Overage Status (2025)

	(1) Aggregate	(2) Math	(3) Spanish
Treatment	0.076***	0.074***	0.065**
	(0.03)	(0.03)	(0.03)
Treatment × 2 years overage	-0.005	-0.011	0.004
	(0.04)	(0.04)	(0.04)
Treatment × 3+ years overage	-0.067	-0.039	-0.092**
	(0.04)	(0.05)	(0.04)
Joint F-test	1.51	0.40	2.99
	0.223	0.670	0.052
Observations	7,999	7,903	7,941

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by overage status for the 2025 cohort. Estimates are from an OLS regression in which the treatment indicator is interacted with indicators for overage categories, with one year of overage as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The sample is restricted to overage students, as no on-age students were included in the 2025 cohort. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 6.

TABLE A23. Effect on Learning of Being Invited to the Remedial Camp by Grade (Pooled)

	(1) Aggregate	(2) Math	(3) Spanish
Treatment (β_T)	0.020	0.005	0.020
	(0.03)	(0.03)	(0.03)
Treatment × Grade 4 ($\beta_{T,4}$)	0.046	0.041	0.054
	(0.04)	(0.04)	(0.04)
Treatment × Grade 5 ($\beta_{T,5}$)	0.053	0.081**	0.026
	(0.04)	(0.04)	(0.04)
Joint F-test	0.66	1.14	0.48
	0.618	0.337	0.749
$\beta_T + \beta_{T,4}$	0.066	0.046	0.074
	0.008	0.103	0.007
$\beta_T + \beta_{T,5}$	0.073	0.086	0.046
	0.003	0.001	0.084
Observations	10,632	10,185	10,309

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators, with grade 3 as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Go to Tables A24 and A25.

TABLE A24. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Grade (2024)

	(1) Aggregate	(2) Math	(3) Spanish
Treatment (β_T)	0.054	-0.012	0.056
	(0.04)	(0.05)	(0.05)
Treatment × Grade 4 ($\beta_{T,4}$)	-0.013	-0.007	0.030
	(0.07)	(0.08)	(0.08)
Treatment × Grade 5 ($\beta_{T,5}$)	0.013	0.069	0.027
	(0.07)	(0.08)	(0.08)
Joint F-test	0.06	0.45	0.10
	0.946	0.639	0.904
$\beta_T + \beta_{T,4}$	0.041	-0.019	0.086
	0.472	0.764	0.161
$\beta_T + \beta_{T,5}$	0.067	0.056	0.083
	0.248	0.396	0.203
Observations	2,633	2,282	2,368

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators, with grade 3 as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table A23.

TABLE A25. Heterogeneous Effects on Learning of Being Invited to the Remedial Camp by Grade (2025)

	(1) Aggregate	(2) Math	(3) Spanish
Treatment (β_T)	0.005	0.003	0.010
	(0.04)	(0.03)	(0.04)
Treatment × Grade 4 ($\beta_{T,4}$)	0.069	0.063	0.061
	(0.04)	(0.05)	(0.05)
Treatment × Grade 5 ($\beta_{T,5}$)	0.070	0.093**	0.025
	(0.05)	(0.05)	(0.05)
Joint F-test	1.41	2.05	0.87
	0.245	0.131	0.421
$\beta_T + \beta_{T,4}$	0.073	0.065	0.071
	0.007	0.035	0.015
$\beta_T + \beta_{T,5}$	0.074	0.095	0.034
	0.004	0.000	0.210
Observations	7,999	7,903	7,941

Note: This table shows heterogeneous intent-to-treat effects of being invited to the remedial camp on endline test scores by grade. Estimates are from an OLS regression in which the treatment indicator is interacted with grade indicators, with grade 3 as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table A23.

TABLE A26. Heterogeneous Effects on Learning of Attending the Remedial Camp (Pooled)

	(1) Aggregate	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	-0.043	-0.092***	0.071
	(0.05)	(0.02)	(0.18)
Share of days attended	0.098**	0.075*	0.088**
	(0.04)	(0.04)	(0.04)
Share of days attended x Female	0.045	0.034	0.036
	(0.07)	(0.07)	(0.07)
Observations	10,257	9,844	9,942
***Panel B**: Heterogeneity by Nationality*			
Haitian	0.113	-0.074	0.043
	(2.06)	(1.83)	(0.12)
Share of days attended	0.123***	0.100***	0.110***
	(0.03)	(0.04)	(0.03)
Share of days attended x Haitian	-0.068	-0.067	-0.079
	(0.10)	(0.10)	(0.10)
Observations	10,212	9,799	9,900
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Share of days attended	0.100***	0.068*	0.095***
	(0.03)	(0.04)	(0.03)
Share of days attended x Zero Distance	0.072	0.085	0.053
	(0.09)	(0.09)	(0.09)
Observations	10,632	10,185	10,309

Note: This table shows heterogeneous average causal response estimates of attending the remedial camp on endline test scores. Each panel reports estimates from a separate two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

TABLE A27. Heterogeneous Effects on Learning of Attending the Remedial Camp (2024)

	(1) Aggregate Ability	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	-0.039	-0.097	0.070
	(0.10)	(0.10)	(0.11)
Share of days attended	0.197	0.030	0.278*
	(0.15)	(0.16)	(0.17)
Share of days attended x Female	-0.020	0.034	-0.109
	(0.25)	(0.25)	(0.28)
Observations	2,258	1,941	2,001
***Panel B**: Heterogeneity by Nationality*			
Haitian	-0.064	-0.056	-0.008
	(0.14)	(0.13)	(0.15)
Share of days attended	0.205*	0.056	0.266**
	(0.11)	(0.13)	(0.12)
Share of days attended x Haitian	-0.180	-0.114	-0.311
	(0.41)	(0.43)	(0.42)
Observations	2,253	1,936	1,995
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Share of days attended	0.122	-0.032	0.177
	(0.10)	(0.12)	(0.12)
Share of days attended x Zero Distance	0.263	0.301	0.284
	(0.22)	(0.26)	(0.27)
Observations	2,633	2,282	2,368
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*			
Share of days attended	0.095	-0.113	0.128
	(0.12)	(0.15)	(0.14)
Share of days attended x Distance	0.023	0.043	0.031
	(0.03)	(0.03)	(0.03)
Observations	2,633	2,282	2,368

Note: This table shows heterogeneous average causal response estimates of attending the remedial camp on endline test scores for the 2024 cohort. Each panel reports estimates from a separate two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are largely absorbed by the stratum fixed effects as these variables do not vary within strata. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

TABLE A28. Heterogeneous Effects on Learning of Attending the Remedial Camp (2025)

	(1) Aggregate Ability	(2) Math	(3) Spanish
***Panel A**: Heterogeneity by Gender*			
Female	0.034	-0.054*	0.116***
	(0.03)	(0.03)	(0.03)
Share of days attended	0.069**	0.084**	0.040
	(0.03)	(0.04)	(0.04)
Share of days attended x Female	0.061	0.037	0.074
	(0.05)	(0.05)	(0.05)
Observations	7,999	7,903	7,941
***Panel B**: Heterogeneity by Nationality*			
Haitian	0.113***	0.170***	0.043
	(0.04)	(0.04)	(0.04)
Share of days attended	0.099***	0.110***	0.072**
	(0.03)	(0.03)	(0.03)
Share of days attended x Haitian	-0.043	-0.066	-0.023
	(0.07)	(0.07)	(0.07)
Observations	7,959	7,863	7,905
***Panel C**: Heterogeneity by Camp Being Host at Student's Regular School*			
Share of days attended	0.088***	0.092***	0.069**
	(0.03)	(0.03)	(0.03)
Share of days attended x Zero Distance	0.012	0.030	-0.019
	(0.08)	(0.07)	(0.07)
Observations	7,999	7,903	7,941
***Panel D**: Heterogeneity by Distance between Camp and Student's Regular School*			
Share of days attended	0.065	0.084*	0.023
	(0.04)	(0.04)	(0.04)
Share of days attended x Distance	0.017	0.009	0.029
	(0.02)	(0.02)	(0.02)
Observations	7,999	7,903	7,941

Note: This table presents average causal response estimates of attending the remedial camp on endline test scores for the 2025 cohort. Each panel reports estimates from a separate two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with a student characteristic: nationality (Panel A), gender (Panel B), whether the camp is located in the student's school from the previous academic year (Panel C), and distance in kilometers between the student's school and the camp site (Panel D). Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The main effects of distance and camp at student's school are absorbed by the stratum fixed effects as these variables do not vary within strata. In 2025, the reduction in maximum distance to camp sites to less than 5 kilometers substantially reduced variation in the distance variables. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

TABLE A29. Effects on Learning of Attending the Remedial Camp by Baseline Test Score Decile (Pooled)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Share of days attended	0.145	0.133	0.079
	(0.09)	(0.09)	(0.11)
Share of days attended × D2	-0.015	-0.043	0.023
	(0.14)	(0.14)	(0.15)
Share of days attended × D3	-0.000	0.021	0.058
	(0.13)	(0.14)	(0.14)
Share of days attended × D4	0.033	-0.096	0.286*
	(0.12)	(0.13)	(0.16)
Share of days attended × D5	0.034	-0.033	0.057
	(0.13)	(0.17)	(0.14)
Share of days attended × D6	0.024	0.008	-0.026
	(0.13)	(0.15)	(0.14)
Share of days attended × D7	-0.025	0.019	0.124
	(0.16)	(0.14)	(0.15)
Share of days attended × D8	0.041	-0.197	0.181
	(0.14)	(0.15)	(0.17)
Share of days attended × D9	-0.033	0.101	-0.023
	(0.12)	(0.14)	(0.14)
Share of days attended × D10	-0.109	-0.016	-0.085
	(0.13)	(0.14)	(0.15)
Joint F-test (D2–D10)	0.34	0.81	1.03
p-value	0.995	0.690	0.420
Observations	8,315	7,462	7,800

Note: This table presents average causal response estimates of attending the remedial camp on endline test scores by baseline test score. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with indicators for each decile of the within-cohort baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum and cohort fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores is not included as a separate control as it enters the specification through the decile interaction terms. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

TABLE A30. Effects on Learning of Attending the Remedial Camp by Baseline Test Score Decile (2024)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Share of days attended	0.407	0.099	-0.160
	(0.29)	(0.38)	(0.52)
Share of days attended × D2	-0.575	-0.054	0.054
	(0.52)	(0.62)	(0.63)
Share of days attended × D3	-0.063	0.186	0.439
	(0.47)	(0.54)	(0.58)
Share of days attended × D4	-0.113	-0.043	0.464
	(0.40)	(0.57)	(0.66)
Share of days attended × D5	-0.104	-0.552	0.804
	(0.42)	(0.82)	(0.64)
Share of days attended × D6	-0.261	-0.758	0.058
	(0.49)	(0.60)	(0.63)
Share of days attended × D7	-0.581	-0.273	1.055
	(0.65)	(0.56)	(0.73)
Share of days attended × D8	-0.182	-0.732	0.703
	(0.58)	(0.64)	(0.87)
Share of days attended × D9	-0.291	0.387	0.013
	(0.41)	(0.59)	(0.63)
Share of days attended × D10	-0.469	0.192	0.107
	(0.45)	(0.53)	(0.75)
Joint F-test (D2–D10)	0.30	0.76	0.65
p-value	0.974	0.652	0.751
Observations	1,619	1,160	1,277

Note: This table presents average causal response estimates of attending the remedial camp on endline test scores by baseline test score for the 2024 cohort. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with indicators for each decile of the baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores is not included as a separate control as it enters the specification through the decile interaction terms. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A31. Effects on Learning of Attending the Remedial Camp by Baseline Test Score Decile (2025)

	(1) Aggregate Ability	(2) Math	(3) Spanish
Share of days attended	0.082	0.139*	0.126
	(0.08)	(0.08)	(0.09)
Share of days attended × D2	0.121	-0.041	0.017
	(0.11)	(0.11)	(0.13)
Share of days attended × D3	0.015	-0.010	-0.017
	(0.11)	(0.12)	(0.13)
Share of days attended × D4	0.068	-0.106	0.252*
	(0.11)	(0.11)	(0.14)
Share of days attended × D5	0.068	0.063	-0.089
	(0.12)	(0.12)	(0.12)
Share of days attended × D6	0.093	0.150	-0.042
	(0.12)	(0.14)	(0.12)
Share of days attended × D7	0.110	0.073	-0.058
	(0.11)	(0.13)	(0.11)
Share of days attended × D8	0.095	-0.099	0.079
	(0.11)	(0.13)	(0.12)
Share of days attended × D9	0.030	0.048	-0.031
	(0.11)	(0.12)	(0.11)
Share of days attended × D10	-0.022	-0.054	-0.123
	(0.12)	(0.13)	(0.11)
Joint F-test (D2–D10)	0.35	0.72	1.37
p-value	0.956	0.688	0.200
Observations	6,696	6,302	6,523

Note: This table presents average causal response estimates of attending the remedial camp on endline test scores by baseline test score for the 2025 cohort. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with indicators for each decile of the baseline test score distribution, with the lowest decile as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including grade. Baseline test scores is not included as a separate control as it enters the specification through the decile interaction terms. The sample is restricted to students for whom both baseline and endline test scores are available. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A32. Effects on Learning of Attending the Remedial Camp by Overage Status (2024)

	(1) Aggregate	(2) Math	(3) Spanish
Share of days attended	0.108	0.127	0.134
	(0.18)	(0.19)	(0.19)
Share of days attended × No overage	0.271	-0.035	0.327
	(0.21)	(0.23)	(0.27)
Share of days attended × 2 years overage	-0.089	-0.336	-0.047
	(0.34)	(0.42)	(0.33)
Share of days attended × 3+ years overage	-0.121	-0.276	-0.035
	(0.30)	(0.34)	(0.35)
Joint F-test (heterogeneity)	1.26	0.40	0.78
p-value	0.290	0.755	0.507
Observations	2,633	2,282	2,368

Note: This table shows heterogeneous average causal response estimates of attending the remedial camp on endline test scores by overage status for the 2024 cohort. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with indicators for overage categories, with one year of overage as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The sample includes students selected by the Ministry of Education to participate in the 2024 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A33. Effects on Learning of Attending the Remedial Camp by Overage Status (2025)

	(1) Aggregate	(2) Math	(3) Spanish
Share of days attended	0.122***	0.121***	0.104**
	(0.04)	(0.04)	(0.05)
Share of days attended × 2 years overage	-0.013	-0.022	0.002
	(0.06)	(0.07)	(0.07)
Share of days attended × 3+ years overage	-0.106	-0.060	-0.148**
	(0.07)	(0.07)	(0.07)
Joint F-test (heterogeneity)	1.33	0.34	2.77
p-value	0.267	0.712	0.064
Observations	7,999	7,903	7,941

Note: This table shows heterogeneous average causal response estimates of attending the remedial camp on endline test scores by overage status for the 2025 cohort. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with indicators for overage categories, with one year of overage as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). The specification includes stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The sample is restricted to overage students, as no on-age students were included in the 2025 cohort. The sample includes students selected by the Ministry of Education to participate in the 2025 remedial camp for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

TABLE A34. Heterogeneous Effects on Learning of Attending the Remedial Camp by Grade

	(1)	(2)	(3)
	Aggregate	Math	Spanish
***Panel A**: Pooled years*			
Share of days attended	0.052	0.006	0.050
	(0.05)	(0.05)	(0.05)
Share of days attended × Grade 4	0.077	0.060	0.109
	(0.08)	(0.08)	(0.08)
Share of days attended × Grade 5	0.093	0.156**	0.050
	(0.08)	(0.08)	(0.08)
Joint F-test	0.77	1.29	0.54
	0.544	0.273	0.706
Observations	10,632	10,185	10,309
***Panel B**: 2024*			
Share of days attended	0.157	-0.035	0.165
	(0.12)	(0.14)	(0.15)
Share of days attended × Grade 4	-0.007	-0.036	0.142
	(0.24)	(0.27)	(0.26)
Share of days attended × Grade 5	0.037	0.200	0.071
	(0.22)	(0.24)	(0.23)
Joint F-test	0.02	0.44	0.16
	0.981	0.645	0.855
Observations	2,633	2,282	2,368
***Panel C**: 2025*			
Share of days attended	0.008	0.005	0.015
	(0.05)	(0.05)	(0.06)
Share of days attended × Grade 4	0.113*	0.104	0.101
	(0.07)	(0.07)	(0.07)
Share of days attended × Grade 5	0.118	0.157**	0.042
	(0.07)	(0.07)	(0.07)
Joint F-test	1.69	2.39	0.98
	0.186	0.094	0.376
Observations	7,999	7,903	7,941

Note: This table shows heterogeneous average causal response estimates of attending the remedial camp on endline test scores by grade. Estimates are from a two-stage least squares regression in which the share of camp days attended (instrumented by the randomly assigned invitation) is interacted with grade indicators, with grade 3 as the omitted category. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. Grade is not included as a separate control as it enters the specification through the interaction terms. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Figure 3.

Appendix B. Camp Staff Focus Groups

To document implementation processes during both years of the summer camp, we conducted focus groups with camp staff in 2024 and 2025. In 2024, focus groups were conducted with teachers only. In 2025, we conducted separate focus groups with teachers,

TABLE A35. Registry of camp inputs tested as univariate predictors of camp effectiveness. Every input tested is shown, none selected post hoc, sorted by R^2.

Camp input (univariate vs. camp effect)	Coef.	SE	p	R^2
camp$_size$	-0.000	(0.000)	0.057	0.070
response$_rate$	0.053	(0.034)	0.126	0.055
pct$_attended$	0.071	(0.042)	0.097	0.054
sh$_levels_recorded$	0.115	(0.071)	0.114	0.049
pct$_female$	-0.192	(0.137)	0.168	0.038
att$_intensity$	0.003	(0.002)	0.259	0.025
supervision	-0.007	(0.008)	0.375	0.019
lvl$_accuracy$	0.005	(0.007)	0.504	0.011
sh$_multilevel$	-0.015	(0.022)	0.517	0.010
sh$_mismatch$	-0.014	(0.026)	0.586	0.006
mat$_sufficient$	-0.006	(0.015)	0.686	0.004
proactivism	-0.003	(0.012)	0.776	0.002
avg$_age$	-0.002	(0.010)	0.821	0.001
sh$_group_level$	0.003	(0.023)	0.909	0.000
tr$_complete$	-0.002	(0.013)	0.908	0.000
camp$_pilot24$	0.002	(0.014)	0.901	0.000
pct$_haitian$	0.003	(0.040)	0.943	0.000
sh$_returning$	-0.001	(0.031)	0.978	0.000
pct$_overage$	0.000	(0.000)	.	0.000
tr$_train$	0.000	(0.000)	.	0.000

0 of 20 significant at the 5% level, against 1 expected by chance.

Notes: 2025 cohort. Each row regresses the empirical-Bayes–shrunken camp effect (in intent-to-treat standard deviations) on a single camp input across camps; columns report the coefficient, robust standard error, *p*-value, and R^2, sorted by R^2. Inputs span camp size, age and demographic composition, attendance intensity, teacher proactivism and practices, technician supervision, placement, and survey coverage. The table lists every input tested, not a selected subset; the final line reports how many are significant at the 5% level against the number expected by chance.

camp coordinators, and district-level technicians.

B.1. 2024 Focus Groups

Focus groups were conducted with 28 teachers. The discussion guide covered preparation and logistics, the use of instructional materials, classroom organization, and student participation.

Teachers described receiving the teacher guides shortly before the start of the camp, in some cases with content for the initial days still incomplete. This limited their ability to plan ahead and required additional work outside scheduled hours to organize and adapt the material. Several also noted the limited availability of physical classroom materials

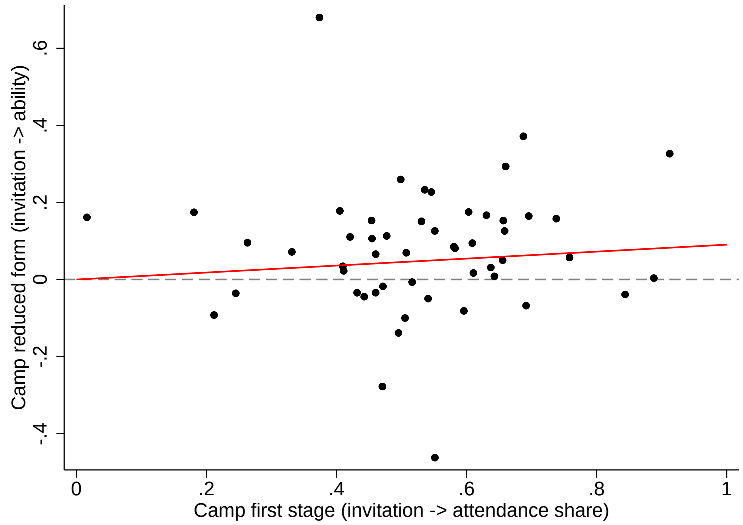

FIGURE A1. Camp-specific reduced forms against first stages.

Notes: Each point is one of the 52 camps in the 2025 cohort. The horizontal axis is the camp's first stage — the within-camp coefficient of attendance (share of the program attended) on the invitation; the vertical axis is the camp's reduced form — the within-camp coefficient of endline aggregate ability (in standard deviations) on the invitation. Both come from separate per-camp regressions with stratum fixed effects and robust standard errors. The solid line is the pooled average causal response drawn through the origin, the proportional relationship that would hold for every camp if effects were homogeneous. Scatter about the line beyond sampling error reflects cross-camp heterogeneity in effectiveness.

such as manipulatives and art supplies, which constrained the variety of activities that could be implemented.

When asked about the alignment between the instructional guides and students' observed learning levels, teachers reported classrooms in which students' reading abilities were below the level assumed by the daily activities. In response, they described modifying activities, reducing the number of exercises completed per day, omitting components, or reorganizing tasks. Several reported that the volume of planned activities exceeded what could be covered within the camp schedule given students' pace.

Teachers also described variation in attendance patterns across the three weeks, including students joining after the first week and others discontinuing attendance. Transportation barriers were frequently mentioned, particularly for students living farther from camp locations, and were described as affecting consistent attendance at some sites. Teachers reported that late arrivals disrupted classroom dynamics, requiring repetition of prior material or individual catch-up instruction.

Finally, teachers described classroom management challenges, particularly during the initial weeks and in classrooms with mixed age groups. Engagement varied across students, and teachers reported making instructional adjustments for students who were not yet literate, including providing individualized support within the classroom.

B.2. 2025 Focus Groups

Focus groups were conducted with 34 teachers, 39 camp coordinators, and 37 district-level technicians. Separate discussion guides were developed for each group to reflect differences in responsibility across classroom instruction, camp-level organization, and district oversight.

Teachers and coordinators highlighted the formal organization of students into three performance levels as a central change relative to 2024. Classrooms were structured accordingly, and both instructional guides and student workbooks were aligned to these levels. They described this structure as reducing within-class heterogeneity compared to the prior year, and the availability of level-specific materials as facilitating daily instructional planning and implementation.

Logistical arrangements were also described as improved. In several locations, counseling and psychology support staff were assigned to assist with implementation, providing additional help with student discipline and well-being. Teachers noted that transportation was more consistently available and that meal and snack provision had improved relative to the prior year. These changes were described as supporting attendance and student participation.

Teachers reported that monitoring visits were frequent at some sites, noting that repeated classroom observations could at times interrupt instructional flow.

Classroom management remained a recurring theme. Teachers described discipline challenges and lower initial engagement during the first week, particularly among older students, but reported that engagement improved as socio-emotional and recreational activities were introduced. Several teachers described students who were often characterized negatively during the regular school year as responding positively to the camp's structured and affirming environment. Participants also noted limited strategies for supporting students with different or special learning needs within the camp structure.

Camp coordinators discussed operational aspects of implementation that were less visible at the classroom level. In particular, they described the process of organizing students into classrooms and coordinating teachers across levels during the first days of the camp. Because the distribution of students across learning levels was finalized only after the camp had begun, the initial organization of classrooms required adjustments during the first days. Several coordinators noted that this process was particularly challenging in camps without counseling or psychology support, as they were often expected to assist teachers in managing behavioral or learning-related issues.

Coordinators also described a substantial operational workload, including coordinating materials, conducting pedagogical classroom visits and providing feedback to teachers, communicating with families, and addressing day-to-day logistical issues. Those who had also participated in the 2024 camp reported that the workload was greater than expected under the single-coordinator model and suggested that having two coordinators — one focused on logistics and another on pedagogical support, as in the previous year — would facilitate operations. This concern was particularly salient for coordinators responsible for more than one camp site, as occurred in three districts where a single coordinator oversaw two camps.

At the same time, coordinators highlighted several positive aspects of the program, including improved transportation arrangements, the provision of snacks for students, and the structure of the instructional materials by level. Some also noted that end-of-day debrief sessions with teachers provided a useful space to discuss challenges and coordinate adjustments for subsequent sessions.

District-level technicians emphasized delays in the delivery of student level-assignment lists and instructional materials in some locations, which complicated the initial planning and organization of classrooms. Technicians also discussed their role in monitoring implementation and providing feedback to camp coordinators and teachers during classroom visits. While monitoring was generally viewed as important for maintaining implementation quality, its frequency at some sites was noted as potentially disruptive to instructional flow — a concern echoed by teachers and coordinators. At the same time, technicians highlighted that regular monitoring helped identify implementation challenges early and facilitated communication between schools, districts, and the central team, allowing logistical issues to be addressed more quickly during the camp.

Appendix C. Summer Camp Material

This appendix describes the instructional materials used in the 2024 and 2025 summer camps, including teacher guides, student notebooks, and supplementary classroom materials.

C.1. Teacher Guides

2024. A single teacher guide was used across all classrooms, covering both Spanish and mathematics. The guide provided daily instructional sequences organized by week, with each week structured around a theme, a specific text type in Spanish, and a defined mathematics focus.

In the first week (cooperative games and technology), Spanish instruction focused on instructive texts — understanding structure, identifying purpose, and following written instructions — while mathematics covered natural number sequences, place value, and number recognition and ordering. In the second week (arts and their different expressions), Spanish shifted to narrative texts, including reading comprehension, narrative structure, and basic written production, and mathematics addressed addition, subtraction, and problem-solving using basic operations. In the third week (environmental care), Spanish covered informational and news texts through reading and discussion, and mathematics introduced multiplication and division.

The daily schedule also included a one-hour ludic activities block at the end of the day, devoted to recreational and complementary activities for the whole group. Suggested activities included cooperative games, group dynamics, artistic and creative activities such as drawing, dramatization, or music, and outdoor or movement-based games. Teachers could also use this time to work with smaller groups of students, allowing for differentiated support or additional practice.

2025. The teacher guide was redesigned in two main ways: it was extended to cover a four-week camp, and it incorporated explicit differentiation by performance level. Students were organized into three groups — Exploradores, Creadores, and Innovadores — and the guide included a table describing expected competencies by level in both subjects. It also integrated socio-emotional components, a bank of recreational activities, structured daily time allocation, and a daily 15-minute reading block.

The weekly progression followed a similar thematic structure. The first week (socio-emotional education and technology) focused on identity documents and personal identity in Spanish, and numeration, recognition, writing, and ordering of numbers in mathematics. The second week (cooperative games) covered shopping lists and functional texts in Spanish, and place value and addition in mathematics. The third week (arts and imagination) addressed short stories and narrative comprehension and production in Spanish, and addition and subtraction in mathematics. The fourth week (environmental care) introduced children's songs and lyrical texts in Spanish, and multiplication and division in mathematics.

Content differed across levels in terms of complexity. In mathematics, differentiation operated primarily through number range and operational complexity. For example, in the first week, Exploradores worked with numbers 1–20 on ordering, identifying previous and next numbers, and writing numbers; Creadores covered the same skills but extended the range to 1–999; and Innovadores worked within the same number range but focused on solving everyday problems using mathematical operations. In Spanish, content varied in reading and writing complexity across levels.

The 2025 guide also included a daily ludic activities block devoted to structured recreational activities designed to promote participation, collaboration, and well-being. Activities included cooperative games, movement-based activities, music and rhythm exercises, and short group dynamics aimed at strengthening social interaction and classroom climate. In some cases, these activities were also used to introduce socio-emotional themes, such as recognizing emotions, building self-esteem, and fostering respectful relationships among students.

C.2. Student Notebooks and Other Materials

In 2024, no separate student notebooks were provided. All instruction was delivered through the teacher guide, with activities implemented based on the daily instructional sequences described therein. In 2025, three separate student notebooks were developed, one for each performance level: Exploradores, Creadores, and Innovadores. Each notebook was organized by week and included explicit weekly learning goals (*Metas de aprendizaje*), separate sections for Spanish and mathematics, a student identification page, socio-emotional elements, and weekly self-assessment sections.

In both years, the activities described in the guides relied on simple classroom materials to implement instructional and recreational activities. These included materials for literacy activities (e.g., cards, short printed texts, or posters to support reading and writing exercises), manipulatives for mathematics (e.g., objects or visual representations for number recognition, ordering, and basic operations), and basic materials for the artistic and movement-based activities in the ludic workshops.

Appendix D. Tests Comparison (2024-2025)

The Summer Camp evaluation administered four rounds of student assessments. Within each cohort, the Baseline and Endline follow a *mirror design*, same structure and skill domains. The tables below compare instruments *across cohorts*: Baseline 2024 against Baseline 2025 (left pair), and Endline 2024 against Endline 2025 (right pair).

TABLE A36. Spanish Items — Change Summary (2024 vs. 2025)

Question Type	**Task**	**Change type**	**What changed**
L1	Listening: identify dictated word among 4 options (3 real + 2 nonsense words)	*New in 2025*	Item added in 2025; not present in 2024 (BL or EL)
L2	Word decoding: complete word with correct missing letter (image-supported, 4 items)	*Minor*	One target word replaced: *montaña* → *paloma* (BL); *elefante* → *mesa* (EL)
L3	Syllable assembly: rearrange syllables into the correct word (4 items)	*None*	Identical across years (BL and EL)
L4	Reading comprehension: read passage and answer 3 MC questions	*Modified*	BL: same story; Q(c) replaced with inferential question. EL: new story (*Tomás el gato*); Q(b) and Q(c) changed
L5	Open-ended justification: write justification for previous answer	*Dropped in 2025*	Item removed in 2025 (BL and EL)
L6	Sentence reconstruction: rewrite the scrambled sentence correctly	*Dropped in 2025*	Item removed in 2025 (BL and EL)
L7	Fill in the blanks: complete 3 sentences using words from a box	*Minor*	BL word box identical. EL word box changed: *amigos/humilde/gigante* → *dormido/curioso/ratón*
L8	Spelling (dictation): write four words dictated by the teacher	*Dropped in 2025*	Item removed in 2025 (BL and EL)

Note: BL = Baseline; EL = Endline. *None* = identical across years; *Minor* = single word or numeral changed; *Modified* = structure or content substantively changed.

TABLE A37. Mathematics Items — Change Summary (2024 vs. 2025): Part I

Question Type	Task	Change type	What changed
M1	Arithmetic operations: vertical addition and subtraction	*Modified*	Expanded from 4 to 6 operations; two simpler 2-digit problems added
M2	Number recognition: identify correct written form of a numeral	*Minor*	BL numeral identical (708). EL numeral changed: 678 → 748
M3	Number conversion: words to digits	*Minor*	BL identical (1,844). EL number and context changed: 1,949 → 1,963
M4	Comparing numbers	*Modified*	Format changed: open-ended with justification (2024) → MC only (2025)
M5	Number ordering on a number line: order values from least to greatest	*Minor*	BL identical. EL price range adjusted: RD$ 130–170 → RD$ 120–160
M6	Addition word problem	*None*	Identical across years (BL and EL)
M7	Subtraction word problem	*Minor*	Numbers changed (BL: 55/17 → 63/25; EL: 65/13 → 74/22); structure unchanged
M8	Division word problem	*Dropped in 2025*	Item removed in 2025 (BL and EL)

Note: BL = Baseline; EL = Endline. *None* = identical across years; *Minor* = single word or numeral changed; *Modified* = structure or content substantively changed.

TABLE A38. Mathematics Items — Change Summary (2024 vs. 2025): Part II

Question Type	Task	Change type	What changed
M9	Multiplication word problem — single-digit multiplier	*None*	Identical across years (BL and EL)
M10	Multiplication word problem — two-digit multiplier	*Dropped in 2025*	Item removed in 2025 (BL and EL)
M11	Place value decomposition — identify hundreds, tens, and units digits (open-ended)	*New in 2025*	Item added in 2025. Number differs by round: 679 (BL) vs. 842 (EL)
M12	Time calculation — add one hour to a given time	*New in 2025*	Item added in 2025. Start time differs by round: 7:00 (BL) vs. 8:00 (EL)
M13	Bar chart interpretation — identify two categories with equal values	*New in 2025*	Item added in 2025. Context differs by round: erasers sold (BL) vs. bottles sold (EL)

Note: None = identical across years; *Minor* = single word or numeral changed; *Modified* = structure or content substantively changed.

The tables below compare instruments *across cohorts*: Baseline 2024 against Baseline 2025 (left pair), and Endline 2024 against Endline 2025 (right pair).

TABLE A39. Spanish Items by Round (Part 1)

	Baseline		Endline	
	2024	**2025**	**2024**	**2025**
L1 Listening — Identify the dictated word among 4 options (3 real words + 2 nonsense words)				
L1	—	**3 real: *moto, paloma, cocodrilo.* Nonsense: *lopame, trusal***	—	**3 real: *luna, camisa, elefante.* Nonsense: *misola, frusal***
L2 Word decoding — Complete the word with the correct letter (image-supported, 4 items)				
L2	*casa, tigre, escuela, montaña*	*casa, tigre, escuela,* ***paloma***	*zapato, canasta, pluma, elefante*	*zapato, canasta, pluma,* ***mesa***
L3 Syllable Assembly — Rearrange syllables into the correct word (4 items)				
L3	*bri-lla-som* (sombrilla), *tra-es-ma* (maestra), *co-bar* (barco), *po-sa-ri-ma* (mariposa)	Same as BL 2024	*na-za-ria-ho* (zanahoria), *za-pi-rra* (pizarra), *bro-li* (libro), *chi-la-mo* (mochila)	Same as EL 2024
L4 Reading Comprehension — Read the passage and answer 3 MC questions				
L4	Story: *El gato y los ratones*. MC: (a) cat's name; (b) who was scared; (c) did they resolve the problem?	Same story; same (a) and (b). **(c) replaced: inferential question on why mice applauded.**	Story: *Nito and Leo*. MC: (a) lion's name; (b) who was scared; (c) did they resolve the problem?	**Story: *Tomás el gato*.** MC: (a) cat's name; **(b) what mice were doing; (c) what mouse proposed to the cat.**

Note: BL = Baseline; EL = Endline. Bold text indicates content that changed relative to 2024. '—' = item not included in that round.

TABLE A40. Spanish Items by Round (Part 2)

	Baseline		Endline	
	2024	**2025**	**2024**	**2025**
L5	*Open-ended Justification — Write justification for the previous answer*			
L5	Write justification for Q3	—	Write justification for Q3	—
L6	*Sentence Reconstruction — Rewrite the scrambled sentence correctly*			
L6	"*Los ratones y ratas se reunieron para buscar una solución*"	—	"*valiente era un Nito muy ratoncito*"	—
L7	*Fill in the Blanks — Complete sentences using words from the box (3 items)*			
L7	*cascabel / astuto / madrigueras*	Same as BL 2024	***amigos / humilde / gigante***	***dormido / curioso / ratón***
L8	*Spelling (dictation) — Write four words dictated by the teacher*			
L8	Four dictated words; party supply list context	—	Four dictated words; party supply list context	—

Note: BL = Baseline; EL = Endline. Bold text indicates content that changed relative to 2024. '—' = item not included in that round.

TABLE A41. Mathematics Items by Round (Part 1)

	Baseline		Endline	
	2024	**2025**	**2024**	**2025**
M1	*Arithmetic Operations — Vertical addition and subtraction*			
M1	4 ops: 234+45, 347+367, 465−54, 546−537	**6 ops:** 34+12**,** 234+45**,** 347+367**,** 79−56**,** 465−54**,** 546−537	4 ops: 345+56, 458+478, 576−65, 657−648	**6 ops:** 42+14**,** 345+56**,** 458+478**,** 89−45**,** 576−65**,** 657−648
M2	*Number recognition — Identify the correct written form of a number (MC, 4 options)*			
M2	Written form of 708	Same as BL 2024	Written form of 678	**Written form of 748**
M3	*Number conversion (words to digits) — Identify the correct numeral (MC, 4 options)*			
M3	*Mil ochocientos cuarenta y cuatro* → 1,844 (Independence year)	Same as BL 2024	*Mil novecientos cuarenta y nueve* → 1,949 (grandmother's birth year)	***Mil novecientos sesenta y tres*** → **1,963 (school inauguration)**
M4	*Comparing Numbers*			
M4	Compare 8493 vs. 9834; justify; construct smallest 4-digit number from {8,3,4,9}	**MC only: greatest among {2735, 2537, 2573, 2753}**	Compare 7561 vs. 7651; justify; construct smallest 4-digit number from {5,7,1,6}	**MC only: greatest among {3648, 3468, 3684, 3486}**
M5	*Number Line — Order prices from least to greatest (MC, 4 options)*			
M5	Five prices, RD$ 140–160	Same as BL 2024	Five prices, RD$ 130–170	**Five prices, RD$ 120–160**
M6	*Addition Word Problem (MC)*			
M6	Wilson has RD$575; dad gives RD$360. Total?	Same as BL 2024	Juan has RD$352; dad gives RD$465. Total?	Same as EL 2024

Note: BL = Baseline; EL = Endline. Bold text indicates content that changed relative to 2024. '—' = item not included in that round.

TABLE A42. Mathematics Items by Round (Part 2)

	Baseline		Endline	
	2024	**2025**	**2024**	**2025**
M7 Subtraction Word Problem (MC)				
M7	Manuel has 55 lives; gives sister 17. Remaining?	**Manuel has 63 lives; gives sister 25. Remaining?**	Juan has 65 bottles; gives cousin 13. Remaining?	**Juan has 74 bottles; gives cousin 22. Remaining?**
M8 Division Word Problem (MC)				
M8	Masks factory: 1,740 units in boxes of 6. How many boxes?	—	Ice cream factory: 1,456 units in boxes of 4. How many boxes?	—
M9–M10 Multiplication Word Problems (MC)				
M9	Restaurant: 54 tables × 6 plates. Total?	Same as BL 2024	Birthday: 42 tables × 5 plates. Total?	Same as EL 2024
M10	Juan: 35 candy bags × 24 candies. Total?	—	Juan: 45 candy bags × 6 candies. Total?	—
M11 Place value decomposition — Identify digits by hundreds, tens, and units (open-ended)				
M11	—	**Hundreds, tens, units digits in 679**	—	**Hundreds, tens, units digits in 842**
M12 Time calculation — Add one hour to a given time (MC, 4 options)				
M12	—	**School starts at 7:00; arrives 1 hour late**	—	**School starts at 8:00; arrives 1 hour late**
M13 Bar chart interpretation — Identify equal values across categories (MC, 4 options)				
M13	—	**Bar chart: two days with same number of erasers sold**	—	**Bar chart: two days with same number of bottles sold**

Note: BL = Baseline; EL = Endline. Bold text indicates content that changed relative to 2024. '—' = item not included in that round.

Appendix E. Socio-Emotional Survey (2025)

The 2025 evaluation included a socio-emotional survey administered at both Baseline and Endline. This instrument was not part of the 2024 evaluation. The Baseline survey measured empathy and emotional awareness, while the Endline survey covered a broader set of constructs including emotion regulation, school belonging, academic self-concept, and growth mindset. Tables A43 and A44 list all items for each round.

E.1. Index Construction

For each Endline item, responses are first reverse-coded where applicable (items 1 and 12) so that higher values consistently reflect higher levels of the measured construct. Each item is then standardized to have mean zero and standard deviation one. Domain indices are constructed as the mean of the standardized items within each domain, following the groupings below:

Index	**Items**	**Reverse-coded items**
Emotional Awareness & Empathy	EL 1, 2, 4, 5, 15	Item 1
Emotion Regulation & Self-Management	EL 3, 6	—
School Belonging	EL 7	—
Academic Self-Concept & Effort	EL 8, 9	—
Classroom Norms & Discipline	EL 10	—
Growth Mindset	EL 11, 12	Item 12
Bullying Victimization	EL 13, 14	—

TABLE A43. Socio-Emotional Survey Items — Baseline 2025

#	Question	Response type	Reverse-coded
1	Children who cry because they are happy are dumb.	Binary (Yes/No)	Yes
2	I really like watching people when they open gifts, even when I don't receive any.	Binary (Yes/No)	No
3	Seeing a child crying makes me want to cry.	Binary (Yes/No)	No
4	Even when I don't know why someone is laughing, I laugh too.	Binary (Yes/No)	No
5	Sometimes I cry when I watch television.	Binary (Yes/No)	No
6	Girls who cry because they are happy are dumb.	Binary (Yes/No)	Yes
7	I find it difficult to understand why another person gets angry.	Binary (Yes/No)	Yes
8	I feel sad when I see a child who can't find anyone to play with.	Binary (Yes/No)	No
9	Some songs make me so sad that they make me want to cry.	Binary (Yes/No)	No
10	Adults sometimes cry even when they have no reason to feel bad.	Binary (Yes/No)	No
11	It is silly to treat dogs and cats as if they had feelings just like people.	Binary (Yes/No)	Yes
12	I get angry when I see a classmate who pretends to need the teacher's help all the time.	Binary (Yes/No)	Yes
13	Children who don't have friends probably don't want to have them.	Binary (Yes/No)	Yes
14	Seeing a girl crying makes me want to cry.	Binary (Yes/No)	No
15	I think it is ridiculous that some people cry during a sad movie or while reading a sad book.	Binary (Yes/No)	Yes
16	I am able to eat all my candy even when I see someone looking at me and wanting one.	Binary (Yes/No)	Yes
17	I do not get upset when I see a classmate punished for not obeying school rules.	Binary (Yes/No)	Yes

TABLE A44. Socio-Emotional Survey Items — Endline 2025

Construct / Domain	#	Question	Response type	Reverse-coded
Emotional Awareness & Empathy	1	Children who cry because they are happy are silly.	Binary (Yes/No)	Yes
	2	Seeing a child cry makes me feel like crying.	Binary (Yes/No)	No
	4	Sometimes, when I get angry, it is actually because I feel sad or hurt.	Binary (Yes/No)	No
	5	Sometimes people get angry because they feel left out.	Binary (Yes/No)	No
	15	When someone gets angry with me, I try to imagine what they are thinking or feeling.	Never / Sometimes / Several times / Always	No
Emotion Regulation & Self-Management	3	I find it easy to calm myself down when I feel upset.	Binary (Yes/No)	No
	6	When I get angry with someone, I try to calm myself down even if that person does not apologize.	Binary (Yes/No)	No
School Belonging & Institutional Support	7	My school cares about my well-being and about me doing well.	Binary (Yes/No)	No
Academic Self-Concept & Effort	8	I am the kind of student who makes an effort and works hard.	Binary (Yes/No)	No
	9	When I learn something new, I feel proud of myself.	Binary (Yes/No)	No
Classroom Norms & Discipline	10	It is important to stay quiet while the teacher is speaking.	Binary (Yes/No)	No
Growth Mindset	11	If I study a lot, I can improve in reading and math.	Binary (Yes/No)	No
	12	Some children are simply good at school and others are not.	Binary (Yes/No)	Yes
Bullying Exposure	13	During the last month: classmates laughed at me or insulted me.	Never / Sometimes / Many times / Every or almost every day	No
	14	During the last month: they left me out of games or activities.	Never / Sometimes / Many times / Every or almost every day	No

Appendix F. IRT and Cronbach's alpha Analysis

We use Item Response Theory (IRT) to construct latent ability measures for each subject (Mathematics and Spanish) and cohort (2024 and 2025). IRT has several advantages over simpler scoring approaches (see, Muralidharan, Singh, and Ganimian (2019)). Unlike percentage-correct or normalized raw scores, IRT allows items to contribute differently to the latent score depending on their estimated discrimination and difficulty, placing greater weight on items that are more informative for distinguishing students along the ability distribution. It also summarizes performance on the retained item set using a common latent scale within each subject and cohort. Finally, it provides item-level diagnostics (discrimination, difficulty, and ICC fit) that allow us to identify and remove poorly performing items before constructing the final scores.

F.1. 2PL Model

Under the two-parameter logistic (2PL) specification, the probability that student i answers item j correctly is:

$$P(Y_{ij} = 1 \mid \theta_i) = \frac{1}{1 + \exp\left[-a_j\left(\theta_i - b_j\right)\right]} \tag{A1}$$

where θ_i is the latent ability of student i, a_j is the *discrimination* parameter, and b_j is the *difficulty* parameter of item j. The difficulty parameter gives the ability level at which the probability of a correct response equals 50%; items with higher b_j require greater ability to answer correctly. The discrimination parameter a_j governs how steeply the probability of success increases with ability around b_j; higher values indicate that the item better differentiates between students of different ability levels. The logistic functional form in equation (A1) implies a monotonic S-shaped Item Characteristic Curve (ICC), mapping latent ability to the probability of a correct response.

For each subject and cohort, we estimate a two-parameter logistic (2PL) model using the retained item set and predict student-specific latent ability scores $\hat{\theta}_i$ from the fitted model. The resulting scores are standardized to mean zero and standard deviation one within cohort. Students who did not attend the endline assessment receive a missing score rather than an imputed value.

F.2. Empirical Distribution of IRT Scores

Figures A2 and A3 present the distribution of IRT scores in Mathematics and Spanish at baseline and endline for each cohort. The distributions are unimodal and well spread across the ability range, with little mass at the floor or ceiling of the distribution. This indicates that the tests provide informative measures of student achievement across a wide range of ability levels in this setting.

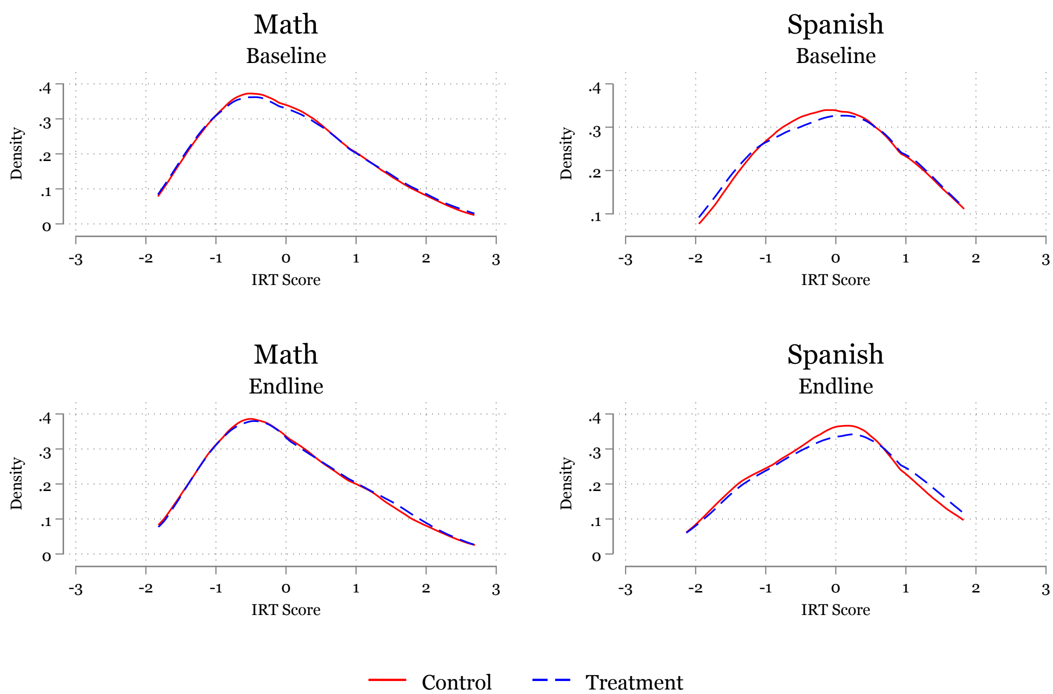

FIGURE A2. Distribution of IRT Scores by Treatment Status, 2024 Cohort

Each panel shows the kernel density of IRT scores separately for treatment and comparison students. The top panels correspond to the baseline assessment (Mathematics and Spanish) and the bottom panels to the endline assessment. IRT scores are estimated using a two-parameter logistic model on the retained item sets and standardized to have mean zero and standard deviation one within cohort

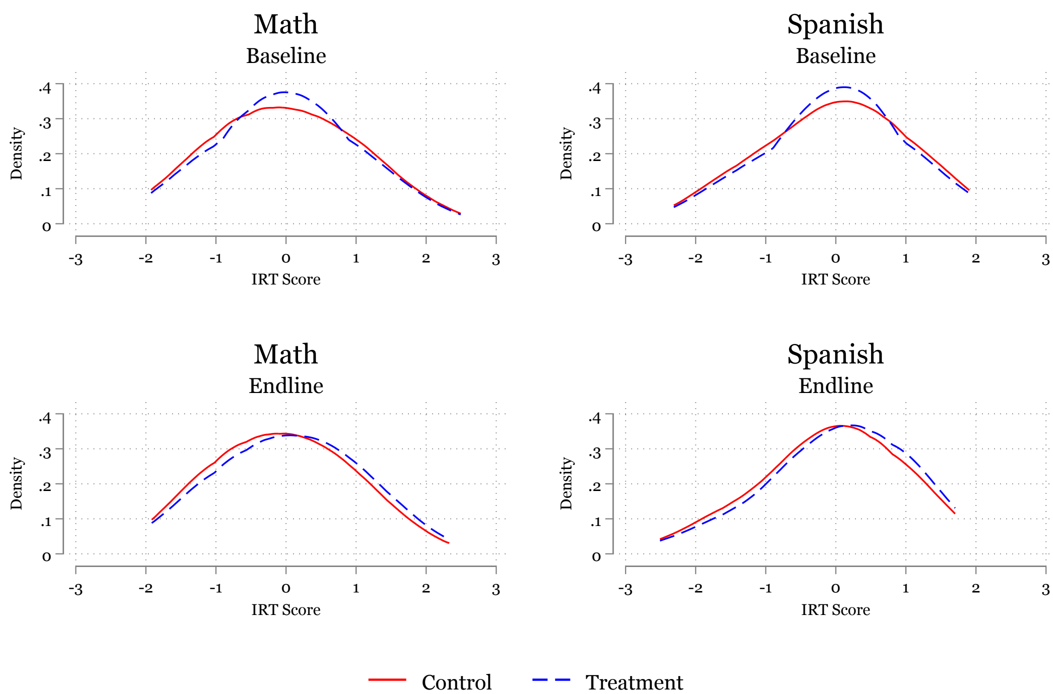

FIGURE A3. Distribution of IRT Scores by Treatment Status, 2025 Cohort

Each panel shows the kernel density of IRT scores separately for treatment and comparison students. The top panels correspond to the baseline assessment (Mathematics and Spanish) and the bottom panels to the endline assessment. IRT scores are estimated using a two-parameter logistic model on the retained item sets and standardized to have mean zero and standard deviation one within cohort.

F.3. Item Fit and Differential Item Functioning

A standard assumption in IRT is that item parameters are invariant across subgroups of respondents, what the psychometrics literature calls the absence of *differential item functioning* (DIF). In a randomized experiment, this assumption is particularly consequential: if the same item behaved differently for treatment and comparison students at the same ability level, the latent scores would not be on a common scale across arms, and treatment effect estimates could be confounded by the measurement model rather than reflecting genuine learning differences.

We assess DIF following the approach of Muralidharan, Singh, and Ganimian (2019): for each item we plot the estimated ICC together with the empirical proportion correct computed within quintiles of the estimated ability distribution, calculated separately for treatment and comparison students. If the model fits well and DIF is absent, the empirical points for both groups should lie close to the same curve. Figures A4–A7 show these plots for all four test specifications. Across items and cohorts, the empirical proportions for treatment and comparison students broadly track the same estimated curve, with no clear visual evidence of systematic differential item functioning across experimental arms. This suggests that the estimated treatment effects are unlikely to be driven by differential behavior of the measurement model between treatment and comparison students.

2024 Cohort

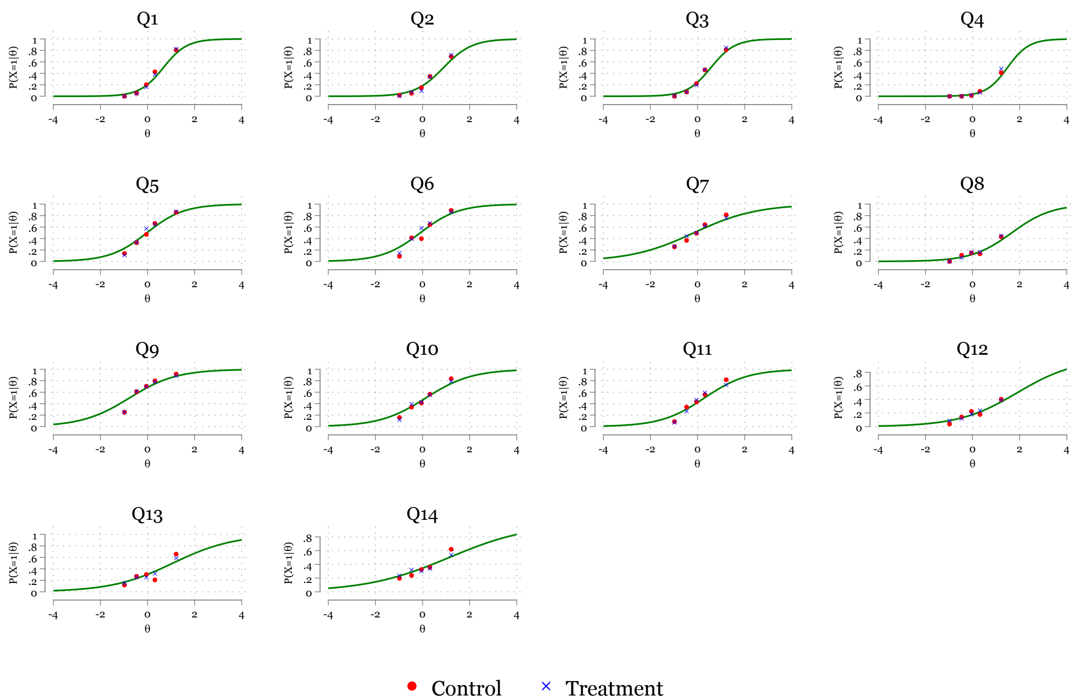

FIGURE A4. Item Characteristic Curves, Math Endline 2024

Each subplot shows the estimated Item Characteristic Curve (ICC) for one item from the 2024 Math endline assessment together with the empirical proportion of correct responses computed within quintiles of the estimated ability distribution, shown separately for treatment (crosses) and comparison (circles) students.

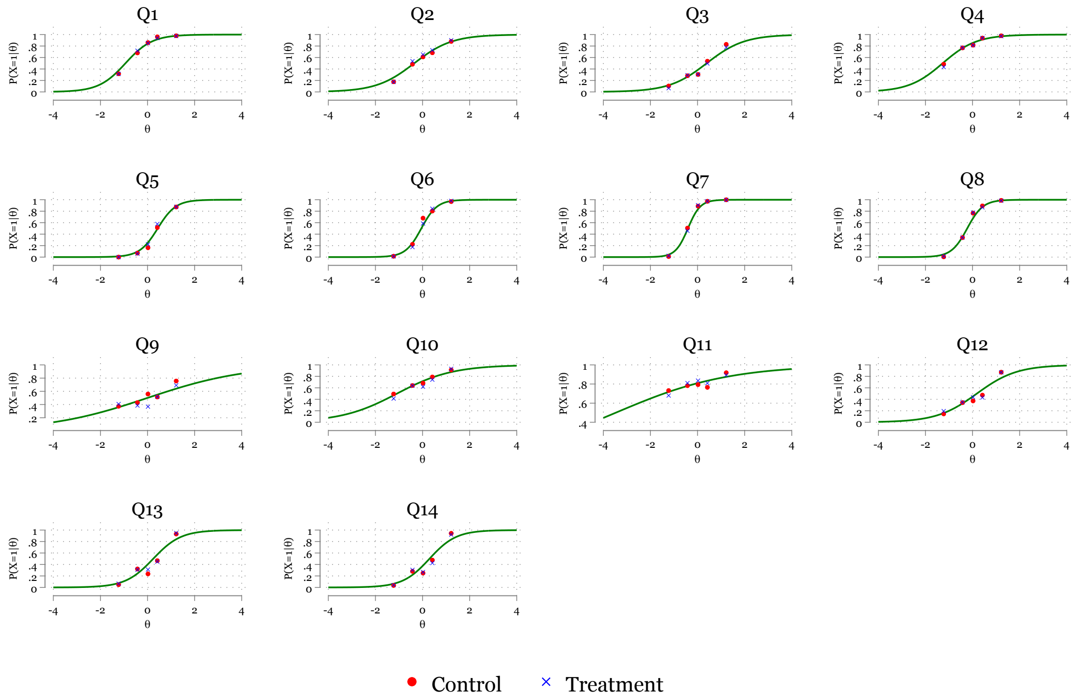

FIGURE A5. Item Characteristic Curves, Spanish Endline 2024

Each subplot shows the estimated Item Characteristic Curve (ICC) for one item from the 2024 Spanish endline assessment together with the empirical proportion of correct responses computed within quintiles of the estimated ability distribution, shown separately for treatment (crosses) and comparison (circles) students.

2025 Cohort

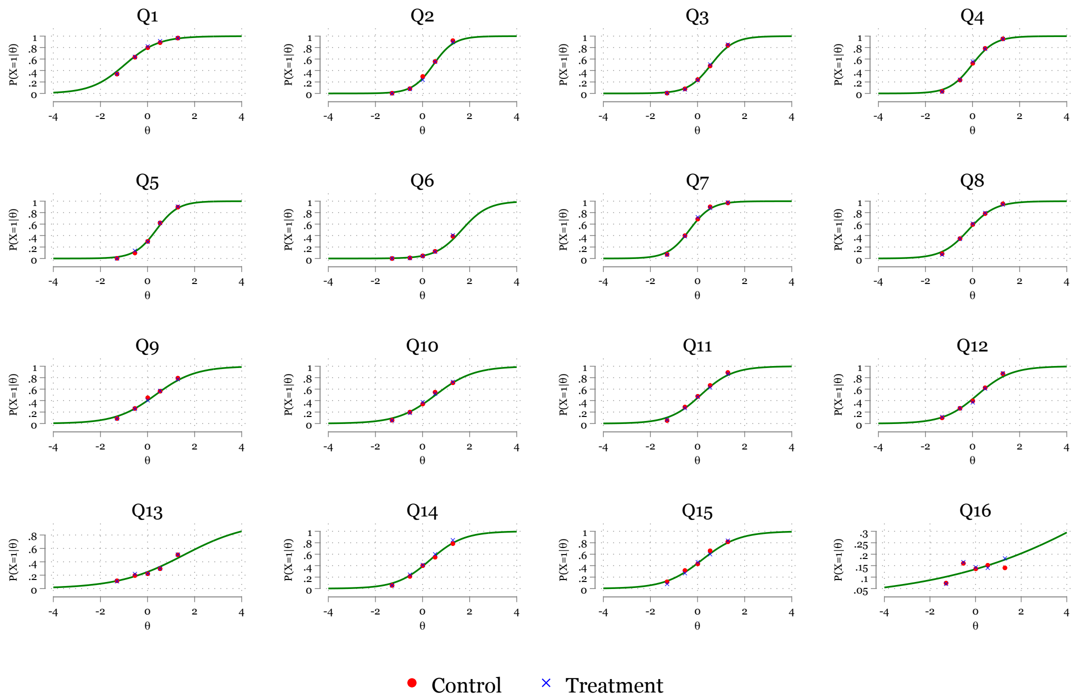

FIGURE A6. Item Characteristic Curves, Math Endline 2025

Each subplot shows the estimated Item Characteristic Curve (ICC) for one item from the 2025 Mathematics endline assessment together with the empirical proportion of correct responses computed within quintiles of the estimated ability distribution, shown separately for treatment (crosses) and comparison (circles) students.

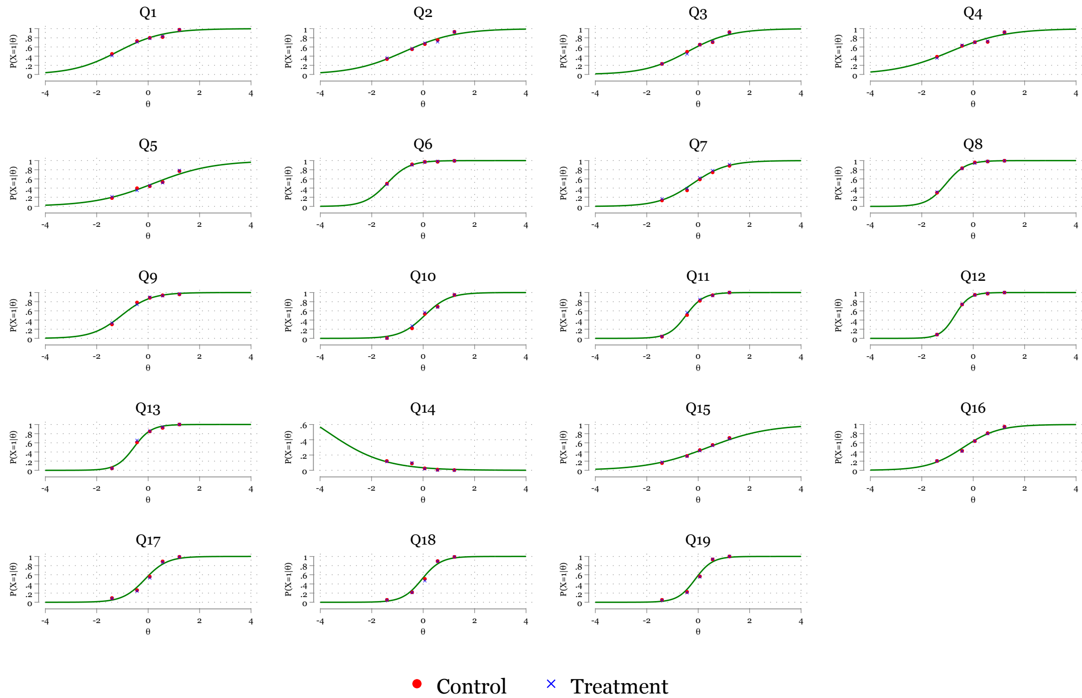

FIGURE A7. Item Characteristic Curves, Spanish Endline 2025

Each subplot shows the estimated Item Characteristic Curve (ICC) for one item from the 2025 Spanish endline assessment together with the empirical proportion of correct responses computed within quintiles of the estimated ability distribution, shown separately for treatment (crosses) and comparison (circles) students.

F.4. Item Screening and Exclusions

For each test specification, we examined each item's estimated discrimination parameter, difficulty parameter, ICC shape, item-rest correlation, and contribution to overall scale reliability. Items exhibiting flat or irregular ICC patterns, weak discrimination, or near-zero item-rest correlations were flagged and reviewed. After this diagnostic review, one item was excluded from each test:

- **2024 Mathematics:** Item Q7 was excluded.
- **2024 Spanish:** Item Q11 was excluded.
- **2025 Mathematics:** Item Q16 was excluded.
- **2025 Spanish:** Item Q14 was excluded.

The final ability scores $\hat{\theta}_i$ used as outcome variables were predicted from the 2PL model estimated on the retained item sets. The excluded items are shown in the reliability tables below (marked with an asterisk) for transparency.

F.5. Scale Reliability

Tables A45–A46 report Cronbach's alpha and item-level diagnostics for each test specification. Cronbach's alpha is defined as:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{j=1}^{K} \sigma_j^2}{\sigma_T^2}\right) \tag{A2}$$

where K is the number of items, σ_j^2 is the variance of item j, and σ_T^2 is the variance of the total score. Values above 0.70 are generally considered acceptable for research purposes. For each item we report: (i) difficulty, measured as the proportion of students answering correctly; (ii) the item-test correlation, i.e. the correlation of the item with the full scale score; (iii) the item-rest correlation, i.e. the correlation of the item with the sum of all other items; and (iv) the alpha that would result from deleting that item. Items marked with * were excluded from the final IRT estimation.

2024 Cohort

TABLE A45. Scale Reliability: Endline 2024

Panel A. Mathematics

Item	Difficulty	Item-Test Corr.	Item-Rest Corr.	α if Deleted
Q1	0.304	0.596	0.498	0.763
Q2	0.269	0.556	0.455	0.769
Q3	0.341	0.603	0.503	0.762
Q4	0.131	0.527	0.451	0.766
Q5	0.531	0.538	0.421	0.770
Q6	0.542	0.537	0.419	0.770
Q7*	0.548	0.437	0.309	0.781
Q9	0.204	0.475	0.378	0.774
Q10	0.678	0.448	0.327	0.778
Q11	0.487	0.495	0.369	0.774
Q12	0.452	0.525	0.404	0.772
Q13	0.211	0.397	0.286	0.782
Q14	0.343	0.454	0.331	0.778
Q15	0.364	0.388	0.257	0.783
Overall α		0.786		

Panel B. Spanish

Item	Difficulty	Item-Test Corr.	Item-Rest Corr.	α if Deleted
Q1	0.755	0.549	0.467	0.816
Q2	0.592	0.532	0.434	0.817
Q3	0.417	0.545	0.448	0.817
Q4	0.794	0.468	0.382	0.822
Q5	0.367	0.653	0.575	0.810
Q6	0.533	0.678	0.600	0.807
Q7	0.662	0.685	0.613	0.805
Q8	0.595	0.675	0.598	0.807
Q9	0.515	0.350	0.228	0.833
Q10	0.713	0.450	0.348	0.823
Q11*	0.811	0.247	0.145	0.837
Q14	0.496	0.530	0.426	0.819
Q15	0.479	0.648	0.562	0.811
Q16	0.471	0.651	0.566	0.811
Overall α		0.828		

Each panel reports item-level reliability diagnostics for the 2024 endline assessment. Difficulty is the proportion of students answering the item correctly. Item-test correlation is the correlation between the item score and the total scale score. Item-rest correlation is the correlation between the item score and the sum of all other items. Alpha if deleted reports the Cronbach's alpha of the scale after removing that item. Overall alpha is computed using the full set of items shown in the table. Items marked with * were excluded from the final IRT estimation due to poor diagnostic performance and are shown here for transparency.

2025 Cohort

TABLE A46. Scale Reliability: Endline 2025

Panel A. Mathematics

Item	Difficulty	Item-Test Corr.	Item-Rest Corr.	α if Deleted
Q1	0.732	0.494	0.406	0.834
Q2	0.360	0.648	0.572	0.824
Q3	0.332	0.620	0.542	0.826
Q4	0.509	0.639	0.558	0.825
Q5	0.388	0.641	0.562	0.825
Q6	0.115	0.437	0.371	0.835
Q7	0.606	0.655	0.579	0.824
Q8	0.551	0.625	0.542	0.826
Q9	0.423	0.524	0.429	0.832
Q10	0.372	0.532	0.440	0.832
Q11	0.465	0.597	0.510	0.828
Q12	0.449	0.576	0.486	0.829
Q13	0.270	0.377	0.279	0.840
Q14	0.420	0.571	0.482	0.829
Q15	0.457	0.561	0.469	0.830
Q16*	0.137	0.161	0.078	0.851
Overall α		0.840		

Panel B. Spanish

Item	Difficulty	Item-Test Corr.	Item-Rest Corr.	α if Deleted
Q1	0.747	0.514	0.444	0.882
Q2	0.647	0.493	0.414	0.883
Q3	0.595	0.542	0.465	0.881
Q4	0.670	0.475	0.395	0.884
Q5	0.464	0.453	0.366	0.885
Q6	0.869	0.575	0.526	0.879
Q7	0.560	0.571	0.496	0.880
Q8	0.813	0.651	0.601	0.877
Q9	0.774	0.583	0.522	0.879
Q10	0.488	0.620	0.551	0.878
Q11	0.671	0.702	0.648	0.875
Q12	0.750	0.719	0.671	0.874
Q13	0.694	0.704	0.651	0.875
Q14*	0.048	-0.160	-0.203	0.890
Q15	0.431	0.421	0.333	0.886
Q16	0.603	0.572	0.499	0.880
Q17	0.549	0.646	0.580	0.878
Q18	0.526	0.667	0.604	0.877
Q19	0.552	0.693	0.634	0.876
Overall α		0.886		

Each panel reports item-level reliability diagnostics for the 2025 endline assessment. Difficulty is the proportion of students answering the item correctly. Item-test correlation is the correlation between the item score and the total scale score. Item-rest correlation is the correlation between the item score and the sum of all other items. Alpha if deleted reports the Cronbach's alpha of the scale after removing that item. Overall alpha is computed using the full set of items shown in the table. Items marked with * were excluded from the final IRT estimation due to poor diagnostic performance and are shown here for transparency.

Appendix G. Analysis sample overlapping in distance across years

TABLE A47. Balance across samples

	Full Sample			Capped (≤ 9km)			Full vs Capped	
	Comparison	Invited	(2)-(1) P-value	Comparison	Invited	(5)-(4) P-value	(4)-(1) Diff.	(5)-(2) Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age (in years)	10.439	10.514	0.931	10.407	10.468	0.509	-0.032**	-0.046***
No overaged	0.257	0.242	0.355	0.261	0.247	0.523	0.004	0.006**
1 year overaged	0.275	0.275	0.353	0.278	0.279	0.349	0.003	0.004*
2 year overaged	0.212	0.215	0.878	0.214	0.216	0.936	0.002	0.001
At least 3 years overaged	0.256	0.269	0.641	0.247	0.258	0.486	-0.009***	-0.011***
Underlying ability in math	-0.000	0.005	0.851	-0.004	-0.003	0.986	-0.004	-0.007*
Underlying ability in spanish	0.000	-0.007	0.472	0.004	-0.012	0.297	0.004	-0.005
Haitian	0.106	0.106	0.595	0.105	0.104	0.613	-0.001	-0.003
Missing nationality	0.106	0.097	0.343	0.102	0.092	0.302	-0.003	-0.006***
Female	0.330	0.331	0.858	0.336	0.336	0.963	0.006*	0.005*
Missing gender	0.104	0.096	0.320	0.101	0.090	0.280	-0.003	-0.006***
Number students	2,113	3,369		1,913	3,040			

Note: This table compares baseline characteristics between the full 2024 sample and the subsample restricted to students whose assigned camp is within 9 kilometers of their school. The distance cap harmonizes the support of the distance variable across cohorts, as camp-to-school distance ranges up to 25 kilometers in 2024 but only up to 10 kilometers in 2025. Columns 1 and 2 report means for the comparison and invited groups in the full sample; column 3 reports the p-value from a test of equality between them. Columns 4 and 5 report means for the comparison and invited groups in the capped sample; column 6 reports the corresponding p-value. P-values in columns 3 and 6 are obtained from a regression of each baseline characteristic on a treatment indicator, absorbing stratum fixed effects and clustering standard errors at the school level. Columns 7 and 8 report the difference in means between the capped and full samples, separately for the comparison and invited groups, with significance stars indicating whether the difference is statistically significant based on a regression of each characteristic on an indicator for being within 9 kilometers with robust standard errors. Baseline characteristics include age, overage status categories, baseline IRT-scaled ability scores in mathematics and Spanish, nationality, and gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 2

TABLE A48. Attrition in Endline Tests (2024)

	Full Sample				Capped (≤ 9km)			
	Spanish		Math		Spanish		Math	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Treatment	-0.031**	-0.028*	-0.047***	-0.048***	-0.031**	-0.028*	-0.045***	-0.046***
	(0.014)	(0.015)	(0.013)	(0.014)	(0.015)	(0.016)	(0.014)	(0.015)
Age (in years)		0.058***		0.047***		0.053**		0.047***
		(0.021)		(0.017)		(0.022)		(0.018)
No overaged		0.044		0.057*		0.027		0.057
		(0.034)		(0.033)		(0.036)		(0.035)
1 year overaged		[base]		[base]		[base]		[base]
2 years overaged		-0.071**		-0.044		-0.070*		-0.051
		(0.034)		(0.037)		(0.037)		(0.040)
At least 3 years overaged		-0.152**		-0.078		-0.139**		-0.088
		(0.062)		(0.058)		(0.067)		(0.062)
Underlying ability in math		-0.000		-0.006		-0.010		-0.009
		(0.013)		(0.013)		(0.013)		(0.014)
Underlying ability in spanish		-0.021		-0.039***		-0.020		-0.038**
		(0.013)		(0.015)		(0.014)		(0.015)
Haitian		-0.003		-0.013		0.009		-0.011
		(0.028)		(0.028)		(0.029)		(0.030)
Female		-0.026		-0.017		-0.027		-0.014
		(0.017)		(0.018)		(0.017)		(0.019)
Control mean	0.518	0.518	0.544	0.544	0.504	0.504	0.530	0.530
Number observations	5,482	5,482	5,482	5,482	4,953	4,953	4,953	4,953
Number schools	302	302	302	302	257	257	257	257

Note: This table examines differential attrition by treatment status in the 2024 cohort, comparing the full sample (left block) with the subsample restricted to students whose assigned camp is within 9 kilometers of their school (right block). Within each block, the first pair of columns reports results for Spanish and the second pair reports results for mathematics; within each subject, column (1) reports the unconditional difference in attrition between treatment and comparison groups, and column (2) adds baseline covariates as controls. Estimates are from an OLS regression including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study during 2025 and 2024, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE A49. Effect of Being Invited to the Remedial Camp on Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Share att.	Total Days	>= 1 days	>= 5 days	>= 10 days	>= 15 days	= 19 days
***Panel A**: Pooled years*							
Treatment	0.450***	6.756***	0.432***	0.412***	0.384***	0.304***	0.275***
	(0.01)	(0.21)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	16,264	16,264	16,264	16,264	16,264	16,264	11,311
Control mean	0.058	0.873	0.068	0.063	0.055	0.024	0.012
***Panel B**: 2024*							
Treatment	0.269***	4.040***	0.358***	0.327***	0.269***	0.070***	
	(0.02)	(0.32)	(0.03)	(0.02)	(0.02)	(0.01)	
Observations	4,953	4,953	4,953	4,953	4,953	4,953	
Control mean	0.137	2.056	0.174	0.159	0.137	0.040	
***Panel C**: 2025*							
Treatment	0.526***	7.894***	0.463***	0.447***	0.432***	0.402***	0.275***
	(0.02)	(0.23)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311
Control mean	0.025	0.373	0.023	0.023	0.021	0.017	0.012

Note: This table shows the effect of being invited to the remedial camp on the share of days attended out of 15 core days (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 3–7), restricting the sample to students whose assigned camp is within 9 kilometers of their school. The distance cap harmonizes the support of the distance variable across cohorts, as camp-to-school distance ranges up to 25 kilometers in 2024 but only up to 9 kilometers in 2025. Estimates are from an OLS regression including stratum and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. We also include cohort fixed effects in the pooled analysis. The sample includes students selected by the Ministry of Education to participate in the study across the 2024 and 2025 remedial camps whose camp is within 9 kilometers, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Table 4

TABLE A50. Effect on Learning of Being Invited to the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.056***	0.048***	0.047***
	(0.02)	(0.02)	(0.02)
Observations	10,513	10,085	10,222
Control mean	-0.001	-0.003	0.001
***Panel B**: 2024*			
Treatment	0.051	-0.006	0.070*
	(0.03)	(0.04)	(0.04)
Observations	2,514	2,182	2,281
Control mean	-0.006	-0.014	0.003
***Panel C**: 2025*			
Treatment	0.056***	0.061***	0.040**
	(0.02)	(0.02)	(0.02)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.96	0.11	0.46

Note: This table shows intent-to-treat estimates of the effect of being invited to the remedial camp on endline test scores. Outcomes are standardized to the comparison group mean within each cohort. Column (1) reports results for aggregate test scores, Column (2) for mathematics, and Column (3) for Spanish. Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The row "2024 vs 2025" reports the p-value from a test of equality of treatment effects across cohorts, obtained from a regression that interacts the treatment indicator with a cohort indicator. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A51. Effects on Learning of Attending the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.102***	0.086***	0.085***
	(0.03)	(0.03)	(0.03)
Observations	10,513	10,085	10,222
Control mean	-0.001	-0.003	0.001
***Panel B**: 2024*			
Treatment	0.161	-0.020	0.221*
	(0.11)	(0.12)	(0.13)
Observations	2,514	2,182	2,281
Control mean	-0.006	-0.014	0.003
***Panel C**: 2025*			
Treatment	0.090***	0.098***	0.066**
	(0.03)	(0.03)	(0.03)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.44	0.34	0.19

Note: This table shows average causal response estimates of the effect of attending the remedial camp on endline test scores. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15. The instrument is the randomly assigned invitation to attend the camp. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The last line reports the p-value from a test of equality of treatment effects across cohorts. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix H. Balance check: full sample and dropping observations with different distance or singletons within a stratum

TABLE A52. Balance across samples

	Full Sample			No singletons			Full vs No singletons	
	Comparison	Invited	(2)-(1) P-value	Comparison	Invited	(5)-(4) P-value	(4)-(1) Diff.	(5)-(2) Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age (in years)	10.439	10.514	0.931	10.440	10.436	0.988	0.001	-0.078***
No overaged	0.257	0.242	0.355	0.258	0.250	0.293	0.001	0.009***
1 year overaged	0.275	0.275	0.353	0.277	0.288	0.357	0.002	0.013***
2 year overaged	0.212	0.215	0.878	0.208	0.210	0.795	-0.004***	-0.005**
At least 3 years overaged	0.256	0.269	0.641	0.258	0.252	0.653	0.002	-0.017***
Underlying ability in math	-0.000	0.005	0.851	-0.002	0.006	0.650	-0.002	0.001
Underlying ability in spanish	0.000	-0.007	0.472	0.001	-0.018	0.373	0.001	-0.011**
Haitian	0.106	0.106	0.595	0.106	0.102	0.591	-0.001	-0.004**
Missing nationality	0.106	0.097	0.343	0.104	0.098	0.390	-0.002	0.001
Female	0.330	0.331	0.858	0.331	0.335	0.803	0.001	0.004*
Missing gender	0.104	0.096	0.320	0.102	0.096	0.361	-0.002	0.000
Number students	2,113	3,369		2,061	3,023			

Note: This table compares baseline characteristics between the full 2024 sample and the subsample restricted to students whose assigned camp is within 9 kilometers of their school. The distance cap harmonizes the support of the distance variable across cohorts, as camp-to-school distance ranges up to 25 kilometers in 2024 but only up to 10 kilometers in 2025. Columns 1 and 2 report means for the comparison and invited groups in the full sample; column 3 reports the p-value from a test of equality between them. Columns 4 and 5 report means for the comparison and invited groups in the capped sample; column 6 reports the corresponding p-value. P-values in columns 3 and 6 are obtained from a regression of each baseline characteristic on a treatment indicator, absorbing stratum fixed effects and clustering standard errors at the school level. Columns 7 and 8 report the difference in means between the capped and full samples, separately for the comparison and invited groups, with significance stars indicating whether the difference is statistically significant based on a regression of each characteristic on an indicator for being within 9 kilometers with robust standard errors. Baseline characteristics include age, overage status categories, baseline IRT-scaled ability scores in mathematics and Spanish, nationality, and gender. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 2

TABLE A53. Attrition in Endline Tests (2024)

	Full Sample				No singletons			
	Spanish		Math		Spanish		Math	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Treatment	-0.031**	-0.028*	-0.047***	-0.048***	-0.032**	-0.030**	-0.047***	-0.048***
	(0.014)	(0.015)	(0.013)	(0.014)	(0.013)	(0.014)	(0.012)	(0.013)
Age (in years)		0.058***		0.047***		0.059***		0.049***
		(0.021)		(0.017)		(0.019)		(0.015)
No overaged		0.044		0.057*		0.050		0.060*
		(0.034)		(0.033)		(0.031)		(0.030)
1 year overaged		[base]		[base]		[base]		[base]
2 years overaged		-0.071**		-0.044		-0.075**		-0.047
		(0.034)		(0.037)		(0.032)		(0.034)
At least 3 years overaged		-0.152**		-0.078		-0.162***		-0.088*
		(0.062)		(0.058)		(0.056)		(0.053)
Underlying ability in math		-0.000		-0.006		0.001		-0.007
		(0.013)		(0.013)		(0.012)		(0.013)
Underlying ability in spanish		-0.021		-0.039***		-0.020		-0.038***
		(0.013)		(0.015)		(0.012)		(0.014)
Haitian		-0.003		-0.013		-0.000		-0.010
		(0.028)		(0.028)		(0.026)		(0.026)
Female		-0.026		-0.017		-0.025		-0.016
		(0.017)		(0.018)		(0.015)		(0.016)
Control mean	0.518	0.518	0.544	0.544	0.522	0.522	0.547	0.547
Number observations	5,482	5,482	5,482	5,482	5,084	5,084	5,084	5,084
Number schools	302	302	302	302	279	279	279	279

Note: This table examines differential attrition by treatment status in the 2024 cohort, comparing the full sample (left block) with the subsample that drops singleton strata and students assigned to a different school than their previous academic year (right block). Within each block, the first pair of columns reports results for Spanish and the second pair reports results for mathematics; within each subject, column (1) reports the unconditional difference, and column (2) adds baseline covariates. Estimates are from an OLS regression including stratum fixed effects. The sample includes all students selected by the Ministry of Education to participate in the study during 2025 and 2024, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A54. Effect of Being Invited to the Remedial Camp on Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Share att.	Total Days	>= 1 days	>= 5 days	>= 10 days	>= 15 days	= 19 days
***Panel A**: Pooled years*							
Treatment	0.443***	6.645***	0.426***	0.406***	0.378***	0.298***	0.275***
	(0.01)	(0.21)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	16,793	16,793	16,793	16,793	16,793	16,793	11,311
Control mean	0.061	0.919	0.071	0.067	0.059	0.025	0.012
***Panel B**: 2024*							
Treatment	0.263***	3.947***	0.346***	0.319***	0.262***	0.072***	
	(0.02)	(0.30)	(0.02)	(0.02)	(0.02)	(0.01)	
Observations	5,482	5,482	5,482	5,482	5,482	5,482	
Control mean	0.139	2.088	0.175	0.160	0.141	0.042	
***Panel C**: 2025*							
Treatment	0.526***	7.894***	0.463***	0.447***	0.432***	0.402***	0.275***
	(0.02)	(0.23)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	11,311	11,311	11,311	11,311	11,311	11,311	11,311
Control mean	0.025	0.373	0.023	0.023	0.021	0.017	0.012

Note: This table shows the effect of being invited to the remedial camp on the share of days attended out of 15 core days (Column 1), total days attended (Column 2), and the likelihood of attending at least 1, 5, 10, 15, or a total of 19 days (Columns 3–7), restricting the sample to students whose assigned camp is within 9 kilometers of their school. The distance cap harmonizes the support of the distance variable across cohorts, as camp-to-school distance ranges up to 25 kilometers in 2024 but only up to 9 kilometers in 2025. Estimates are from an OLS regression including stratum and baseline controls pre-specified in our pre-analysis plan, including baseline test scores and grade. We also include cohort fixed effects in the pooled analysis. The sample includes students selected by the Ministry of Education to participate in the study across the 2024 and 2025 remedial camps whose camp is within 9 kilometers, including students for whom baseline test scores are unavailable. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * p < 0.10, ** p < 0.05, *** p < 0.01. Back to Table 4

TABLE A55. Effect on Learning of Being Invited to the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.057***	0.048***	0.049***
	(0.02)	(0.02)	(0.02)
Observations	10,712	10,249	10,386
Control mean	0.000	0.000	-0.000
***Panel B**: 2024*			
Treatment	0.055*	-0.001	0.077**
	(0.03)	(0.04)	(0.04)
Observations	2,713	2,346	2,445
Control mean	0.000	-0.000	-0.000
***Panel C**: 2025*			
Treatment	0.056***	0.061***	0.040**
	(0.02)	(0.02)	(0.02)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.95	0.13	0.34

Note: This table shows intent-to-treat estimates of the effect of being invited to the remedial camp on endline test scores. Outcomes are standardized to the comparison group mean within each cohort. Column (1) reports results for aggregate test scores, Column (2) for mathematics, and Column (3) for Spanish. Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The row "2024 vs 2025" reports the p-value from a test of equality of treatment effects across cohorts, obtained from a regression that interacts the treatment indicator with a cohort indicator. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A56. Effects on Learning of Attending the Remedial Camp

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
***Panel A**: Pooled years*			
Treatment	0.104***	0.088***	0.089***
	(0.03)	(0.03)	(0.03)
Observations	10,712	10,249	10,386
Control mean	0.000	0.000	-0.000
***Panel B**: 2024*			
Treatment	0.170*	-0.003	0.242**
	(0.10)	(0.12)	(0.12)
Observations	2,713	2,346	2,445
Control mean	0.000	-0.000	-0.000
***Panel C**: 2025*			
Treatment	0.090***	0.098***	0.066**
	(0.03)	(0.03)	(0.03)
Observations	7,999	7,903	7,941
Control mean	0.000	0.000	-0.000
2024 vs 2025	0.36	0.40	0.12

Note: This table shows average causal response estimates of the effect of attending the remedial camp on endline test scores. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15. The instrument is the randomly assigned invitation to attend the camp. Columns report results for aggregate test scores (Column 1), mathematics (Column 2), and Spanish (Column 3). Panel A reports pooled results across both cohorts, Panel B reports results for the 2024 cohort, and Panel C reports results for the 2025 cohort. All specifications include stratum fixed effects and baseline controls pre-specified in our pre-analysis plan, including baseline test scores interacted with grade. The pooled specification in Panel A additionally includes cohort fixed effects. The last line reports the p-value from a test of equality of treatment effects across cohorts. The sample includes all students selected by the Ministry of Education to participate in the study for whom endline test scores are available. Missing baseline variables are imputed at the mean, with indicator variables for missingness included. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix I. Lee bound sensitivity analysis

TABLE A57. Lee Bounds on the Effects on Learning of Being Invited to the Remedial Camp (Residualized Ranking)

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
Panel A: Pooled years			
Invited	0.057***	0.050***	0.048***
	(0.015)	(0.016)	(0.015)
Lee Lower bound	0.046***	0.033**	0.045***
	(0.015)	(0.016)	(0.015)
Lee Upper bound	0.103***	0.103***	0.093***
	(0.014)	(0.016)	(0.015)
Imbens-Manski 95% CI	[0.022, 0.126]	[0.007, 0.129]	[0.020, 0.117]
Trim share (q)	0.021	0.027	0.019
Group trimmed	Invited	Invited	Invited
Observations	10,632	10,185	10,309
Panel B: 2024			
Invited	0.054*	0.005	0.072**
	(0.029)	(0.034)	(0.034)
Lee Lower bound	0.024	-0.033	0.072**
	(0.029)	(0.034)	(0.034)
Lee Upper bound	0.187***	0.182***	0.205***
	(0.029)	(0.032)	(0.034)
Imbens-Manski 95% CI	[-0.023, 0.235]	[-0.088, 0.234]	[0.016, 0.262]
Trim share (q)	0.074	0.097	0.067
Group trimmed	Invited	Invited	Invited
Observations	2,633	2,282	2,368
Panel C: 2025			
Invited	0.056***	0.061***	0.040**
	(0.016)	(0.018)	(0.017)
Lee Lower bound	0.055***	0.057***	0.039**
	(0.016)	(0.018)	(0.017)
Lee Upper bound	0.056***	0.065***	0.042**
	(0.016)	(0.018)	(0.017)
Imbens-Manski 95% CI	[0.023, 0.087]	[0.025, 0.096]	[0.007, 0.074]
Trim share (q)	0.000	0.002	0.001
Group trimmed	Comparison	Invited	Invited
Observations	7,999	7,903	7,941

Note: This table reports Lee (2009) sharp bounds on the intent-to-treat (ITT) effect of being invited to the remedial camp on standardized endline test scores, addressing differential attrition between invited and comparison groups. Columns (1)–(3) report results for aggregate test scores, mathematics, and Spanish, respectively. Panel A reports pooled estimates across both cohorts, Panel B reports the 2024 cohort, and Panel C reports the 2025 cohort. The row labeled *Invited* reports the conventional ITT estimate using the full survivor sample. The *Lee Lower bound* and *Lee Upper bound* rows report estimates from re-running the ITT regression after trimming a fraction q of the higher-survival group's observed outcomes, where q equals the difference in survival rates divided by the higher survival rate. The row *Trim share (q)* reports the trimming fraction applied in each specification. Trimming is based on the rank of the residualized endline score, where residuals come from a regression of the outcome on baseline controls and randomization-strata fixed effects estimated on the full survivor sample. Imbens-Manski (2004) 95% confidence intervals are reported in brackets. Because Lee bounds partially identify the treatment effect under monotonicity, the intervals in brackets constitute the relevant inference for the identified set. Standard errors clustered at the school level are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE A58. Lee Bounds on the Effects on Learning of Being Invited to the Remedial Camp (Raw Outcome Ranking)

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
Panel A: Pooled years			
Invited	0.057***	0.050***	0.048***
	(0.015)	(0.016)	(0.015)
Lee Lower bound	0.037**	0.020	0.029*
	(0.015)	(0.016)	(0.015)
Lee Upper bound	0.092***	0.090***	0.086***
	(0.014)	(0.016)	(0.015)
Imbens-Manski 95% CI	[0.013, 0.116]	[-0.006, 0.115]	[0.003, 0.110]
Trim share (q)	0.021	0.027	0.019
Group trimmed	Invited	Invited	Invited
Observations	10,632	10,185	10,309
Panel B: 2024			
Invited	0.054*	0.005	0.072**
	(0.029)	(0.034)	(0.034)
Lee Lower bound	-0.021	-0.092***	-0.000
	(0.029)	(0.035)	(0.035)
Lee Upper bound	0.136***	0.132***	0.171***
	(0.030)	(0.033)	(0.034)
Imbens-Manski 95% CI	[-0.069, 0.186]	[-0.150, 0.187]	[-0.057, 0.228]
Trim share (q)	0.074	0.097	0.067
Group trimmed	Invited	Invited	Invited
Observations	2,633	2,282	2,368
Panel C: 2025			
Invited	0.056***	0.061***	0.040**
	(0.016)	(0.018)	(0.017)
Lee Lower bound	0.055***	0.058***	0.039**
	(0.016)	(0.018)	(0.017)
Lee Upper bound	0.056***	0.062***	0.041**
	(0.016)	(0.018)	(0.017)
Imbens-Manski 95% CI	[0.023, 0.088]	[0.025, 0.096]	[0.007, 0.074]
Trim share (q)	0.000	0.002	0.001
Group trimmed	Comparison	Invited	Invited
Observations	7,999	7,903	7,941

Note: This table reports Lee (2009) sharp bounds on the intent-to-treat (ITT) effect of being invited to the remedial camp on standardized endline test scores, addressing differential attrition between invited and comparison groups. Columns (1)–(3) report results for aggregate test scores, mathematics, and Spanish, respectively. Panel A reports pooled estimates across both cohorts, Panel B reports the 2024 cohort, and Panel C reports the 2025 cohort. The row labeled *Invited* reports the conventional ITT estimate using the full survivor sample. The *Lee Lower bound* and *Lee Upper bound* rows report estimates from re-running the ITT regression after trimming a fraction q of the higher-survival group's observed outcomes, where q equals the difference in survival rates divided by the higher survival rate. The row *Trim share (q)* reports the trimming fraction applied in each specification. In this robustness specification, observations are ranked directly using the raw endline outcome in the trimming step, rather than the residualized outcome. Imbens-Manski (2004) 95% confidence intervals are reported in brackets. Because Lee bounds partially identify the treatment effect under monotonicity, the intervals in brackets constitute the relevant inference for the identified set. Standard errors clustered at the school level are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE A59. Lee Bounds on the Effects on Learning of Attending the Remedial Camp (Residualized Ranking)

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
Panel A: Pooled years			
Share of days attended	0.103***	0.090***	0.086***
	(0.026)	(0.029)	(0.027)
Lee Lower bound	0.085***	0.060**	0.081***
	(0.027)	(0.029)	(0.027)
Lee Upper bound	0.188***	0.186***	0.167***
	(0.025)	(0.028)	(0.027)
Imbens-Manski 95% CI	[0.041, 0.229]	[0.013, 0.232]	[0.036, 0.211]
Trim share (*q*)	0.021	0.027	0.019
Group trimmed	Invited	Invited	Invited
Observations	10,632	10,185	10,309
Panel B: 2024			
Share of days attended	0.166*	0.016	0.222**
	(0.092)	(0.105)	(0.106)
Lee Lower bound	0.075	-0.100	0.222**
	(0.089)	(0.103)	(0.106)
Lee Upper bound	0.576***	0.561***	0.625***
	(0.099)	(0.108)	(0.115)
Imbens-Manski 95% CI	[-0.072, 0.738]	[-0.269, 0.739]	[0.047, 0.814]
Trim share (*q*)	0.074	0.097	0.067
Group trimmed	Invited	Invited	Invited
Observations	2,633	2,282	2,368
Panel C: 2025			
Share of days attended	0.090***	0.098***	0.066**
	(0.026)	(0.029)	(0.027)
Lee Lower bound	0.089***	0.092***	0.064**
	(0.026)	(0.028)	(0.027)
Lee Upper bound	0.090***	0.105***	0.068**
	(0.026)	(0.028)	(0.027)
Imbens-Manski 95% CI	[0.038, 0.141]	[0.041, 0.155]	[0.012, 0.120]
Trim share (*q*)	0.000	0.002	0.001
Group trimmed	Comparison	Invited	Invited
Observations	7,999	7,903	7,941

Note: This table reports Lee (2009) sharp bounds on the average causal response (ACR) of attending the remedial camp on standardized endline test scores, addressing differential attrition between invited and comparison groups. Columns (1)–(3) report results for aggregate test scores, mathematics, and Spanish, respectively. Panel A reports pooled estimates across both cohorts, Panel B reports the 2024 cohort, and Panel C reports the 2025 cohort. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15. The instrument is the randomly assigned invitation to attend the camp. The row labeled *Share of days attended* reports the conventional 2SLS estimate using the full survivor sample. The *Lee Lower bound* and *Lee Upper bound* rows report estimates from re-running the 2SLS specification after trimming a fraction q of the higher-survival group's observed outcomes, where q equals the difference in survival rates divided by the higher survival rate. The row *Trim share (q)* reports the trimming fraction applied in each specification. Trimming is based on the rank of the residualized endline score, where residuals come from a regression of the outcome on baseline controls and randomization-strata fixed effects estimated on the full survivor sample. Imbens-Manski (2004) 95% confidence intervals are reported in brackets. Because Lee bounds partially identify the treatment effect under monotonicity, the intervals in brackets constitute the relevant inference for the identified set. Standard errors clustered at the school level are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE A60. Lee Bounds on the Effects on Learning of Attending the Remedial Camp (Raw Outcome Ranking)

	(1)	(2)	(3)
	Aggregate Ability	Math	Spanish
Panel A: Pooled years			
Share of days attended	0.103***	0.090***	0.086***
	(0.026)	(0.029)	(0.027)
Lee Lower bound	0.067**	0.035	0.052*
	(0.027)	(0.028)	(0.028)
Lee Upper bound	0.167***	0.161***	0.155***
	(0.025)	(0.027)	(0.026)
Imbens-Manski 95% CI	[0.023, 0.209]	[-0.011, 0.206]	[0.007, 0.197]
Trim share (q)	0.021	0.027	0.019
Group trimmed	Invited	Invited	Invited
Observations	10,632	10,185	10,309
Panel B: 2024			
Share of days attended	0.166*	0.016	0.222**
	(0.092)	(0.105)	(0.106)
Lee Lower bound	-0.065	-0.286**	-0.000
	(0.093)	(0.110)	(0.110)
Lee Upper bound	0.423***	0.409***	0.519***
	(0.098)	(0.109)	(0.111)
Imbens-Manski 95% CI	[-0.219, 0.585]	[-0.467, 0.589]	[-0.180, 0.701]
Trim share (q)	0.074	0.097	0.067
Group trimmed	Invited	Invited	Invited
Observations	2,633	2,282	2,368
Panel C: 2025			
Share of days attended	0.090***	0.098***	0.066**
	(0.026)	(0.029)	(0.027)
Lee Lower bound	0.089***	0.095***	0.064**
	(0.026)	(0.028)	(0.027)
Lee Upper bound	0.091***	0.101***	0.067**
	(0.026)	(0.029)	(0.027)
Imbens-Manski 95% CI	[0.038, 0.142]	[0.042, 0.154]	[0.012, 0.119]
Trim share (q)	0.000	0.002	0.001
Group trimmed	Comparison	Invited	Invited
Observations	7,999	7,903	7,941

Note: This table reports Lee (2009) sharp bounds on the average causal response (ACR) of attending the remedial camp on standardized endline test scores, addressing differential attrition between invited and comparison groups. Columns (1)–(3) report results for aggregate test scores, mathematics, and Spanish, respectively. Panel A reports pooled estimates across both cohorts, Panel B reports the 2024 cohort, and Panel C reports the 2025 cohort. The endogenous variable is the share of camp days attended, defined as total days attended divided by 15. The instrument is the randomly assigned invitation to attend the camp. The row labeled *Share of days attended* reports the conventional 2SLS estimate using the full survivor sample. The *Lee Lower bound* and *Lee Upper bound* rows report estimates from re-running the 2SLS specification after trimming a fraction q of the higher-survival group's observed outcomes, where q equals the difference in survival rates divided by the higher survival rate. The row *Trim share (q)* reports the trimming fraction applied in each specification. In this robustness specification, observations are ranked directly using the raw endline outcome in the trimming step, rather than the residualized outcome. Imbens-Manski (2004) 95% confidence intervals are reported in brackets. Because Lee bounds partially identify the treatment effect under monotonicity, the intervals in brackets constitute the relevant inference for the identified set. Standard errors clustered at the school level are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.