Research Institute for the Evaluation of Public Polices

http://irvapp.fbk.eu

FBK-IRVAPP Working Paper series

Research Institute for the Evaluation of Public Policies

Starting the school year on the right foot. Effects of a summer learning program targeting vulnerable students in Italy

Davide Azzolini

Martina Bazzoli

Sergiu C. Burlacu

Enrico Rettore

September 2023

FBK-IRVAPP Working Paper No. 2023-05

Starting the school year on the right foot. Effects of a summer learning program targeting vulnerable students in Italy

Davide Azzolini
FBK-IRVAPP, Urban Institute
azzolini@irvapp.it

Martina Bazzoli
FBK-IRVAPP
bazzoli@irvapp.it

Sergiu C. Burlacu
FBK-IRVAPP
sburlacu@irvapp.it

Enrico Rettore
University of Padua, FBK-IRVAPP, IZA
enrico.rettore@unipd.it

FBK-IRVAPP Working Paper No. 2023-05

September 2023

Research Institute for the Evaluation of Public Policies
Bruno Kessler Foundation
Vicolo dalla Piccola 2, 38122 Trento (Italy)

Phone: (+39) 0461.314209
Fax: (+39) 0461.314240

E-mail: info@irvapp.it
Website: http://irvapp.fbk.eu

The purpose of the IRVAPP Working Papers series is to promote the circulation of working papers prepared within the Institute or presented in IRVAPP seminars by outside researchers with the aim of stimulating comments and suggestions. Updated reviews of the papers are available in the Reprint Series, if published, or directly at the IRVAPP.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Institute.

Starting the school year on the right foot. Effects of a summer learning program targeting vulnerable students in Italy

Davide Azzolini

Martina Bazzoli

Sergiu C. Burlacu

Enrico Rettore

Abstract

We conducted a randomized controlled trial to evaluate the impact of a summer learning program for vulnerable students across ten cities in Italy (N=1,038). The program had two components: educational workshops in small groups (88 hours) and personalized tutoring (12 hours). Results indicate significant improvements in reading comprehension and marginally in grammar. Improvements in arithmetic and geometry are smaller albeit significant when aggregated into a single mathematics score. Effects were most pronounced among primary school students and among students with special needs or from vulnerable environments. The program compensated for summer learning loss, as treatment group students returned to school in September with higher learning levels than before the summer, while the control group experienced learning setbacks, predominantly in mathematics. While the study clearly shows that students start the new year with a higher level of competencies, it does not definitively establish the lasting impact of these effects. An explorative analysis of noncognitive skills provides conflicting insights: an increase in students' interest in acquiring new competencies suggests potential enduring effects, but the emergence of dissatisfaction with traditional school activities and heightened school-related stress raises concerns about reduced engagement with conventional schooling.

JEL-Code: I24

Key words: summer learning loss, summer learning program, Italy, achievement gap, RCT

Introduction

The phenomenon of students forgetting what they learn at school during summer breaks – also referred to as summer learning loss (or slide) – has been studied for decades (Cooper et al. 1996). Even if there is no full consensus on the size of this loss nor on which type of students would be more affected (von Hippel & Hamrock, 2019; Atteberry & McEachin, 2021; Workman et al., 2023), a widespread concern exists regarding the risk that summer learning loss contributes to widening educational inequalities, because children from lower social backgrounds may be exposed to less cognitively stimulating home environments (Chin and Philipps, 2004; Downey et al. 2004; Alexander et al., 2007).

Summer learning programs have emerged as an attempt to prevent or contain the loss of learning and the widening of learning gaps. It can be argued that allocating resources towards summer interventions is a valuable endeavor, because, while not all students may lose substantial knowledge over the summer break, the vast majority of studies find that learning progress stagnates during this period (Workman et al., 2023). Therefore, summer represents a prime time to invest in the education of economically disadvantaged students, aiding in bridging the gap in relation to their more affluent peers (McCombs et al., 2019). This is likely to be even more relevant in the present post-pandemic period, marked by significant learning setbacks and increased social disparities (Santibañez & Guarino, 2021; Betthäuser et al., 2023).

The evaluation literature suggests that not only summer learning programs hold the potential to alleviate summer learning loss, they also enable to increase children's learning (Kim & Quinn, 2013; McEachin et al., 2018; McCombs et al., 2019; Lynch et al., 2023; Cooper et al., 2000). The estimated effect size of these programs on reading and mathematics achievement are around +.10 standard deviations (Kim & Quinn 2013; Lynch et al., 2023). There is also evidence of summer learning programs' effects on non-cognitive outcomes such as study motivation and school engagement, including the reduction of unexcused absences, chronic absenteeism, and suspensions (Pyne et al., 2021; Lynch et al., 2023). However, this evidence is constrained by the limited number of studies exploring such outcomes. Furthermore, a significant limitation of this literature is its almost entire focus on North America, with the Swedish study by Fälth et al. (2019) serving as a notable exception. This inherently raises concerns about the broader applicability of the findings.

This study sets out to extend the evidence on the effectiveness of summer learning programs in the Italian context, where they have never been studied before. The study employs a randomized controlled trial to estimate the effects of a summer learning program targeted to

socially vulnerable students (i.e., *Arcipelago Educativo*) on student learning achievements in mathematics and reading. The program was implemented in 11 sites covering the main areas of the country. The study was pre-registered on the American Economic Association's registry for randomized controlled trials (ID number AEARCTR-0009958).[1] Any additional analyses beyond those pre-registered are considered exploratory.

The program consists of 100-hour educational activities organized in small groups (88 hours) and personalized tutoring (12 hours). The design of the program follows the state of the art in the evaluation literature in terms of the elements considered key in enhancing effectiveness: focused purpose, substantial duration, active involvement of students and families, small class sizes, and employing well-trained educators (McCombs et al., 2019; Lynch et al., 2023).[2] About seven out of ten invited students accepted to participate in the program and attended it for 71 hours, on average.

The experimental findings point to a statistically significant increase in students' reading comprehension (0.17 standard deviations, SD), a marginally significant increase in grammar (0.10 SD) and positive but insignificant improvements in arithmetic (0.10 SD) and geometry (0.05 SD). When the test scores are grouped by subject domain, the results indicate a statistically significant effect on reading literacy (0.17 SD) and marginally significant positive effects on mathematics (0.09 SD). The effects are primarily driven by primary school students and students with students with special needs or from vulnerable environments (in the Italian school system classified as having special educational needs - *Bisogni Educativi Speciali*).

The evaluation also included an exploratory analysis on several dimensions of noncognitive skills highlighted by the implementing organizations as being potentially impacted by the intervention. Results point to an increase in the desire to acquire new competencies, but also to an increase in dissatisfaction with normal school activities and school-related stress. While these results are only suggestive and should be taken with caution, potentially, they highlight why the students in the sample have learning difficulties in a regular school environment.

Considering the worldwide learning setbacks caused by the COVID-19 Pandemic (Betthäuser et al., 2023), the concept of summer learning loss has regained prominence in discussions surrounding education policy. As a result of this, the goal of devising new interventions — during the summer or throughout the school year in addition to regular

[1] Available at the following link: https://doi.org/10.1257/rct.9958-1.0

[2] Summer learning programs exhibit considerable diversity: certain programs might prioritize remediation, whereas others lean towards enrichment. Some programs may adopt a comprehensive approach, while others are more subject-specific. The existing evidence on programs' features and their implementation is not well established, however the elements listed in the paragraph appear to be among the most influential features.

schooling — to counteract learning loss and sustain learning achievements, especially among the most vulnerable students, has garnered unprecedented attention. By expanding the existing body of experimental literature on the effectiveness of summer learning programs in a context where such programs have yet to be evaluated, our study supports the notion that the summer period can be utilized to achieve this goal.

The context and the program evaluated

Italy is a country with lower than OECD average test scores, relatively high levels of social inequality in education (OECD 2019), and a strikingly low proportion of tertiary graduates (OECD 2023). Moreover, despite its 12 uninterrupted weeks of summer holidays (European Commission 2022), which are longer than most other European countries and the United States, summer programs with a high learning content instead of a recreational one, are a rarity. *Arcipelago Educativo* is one of the only systematic attempts to introduce this kind of programs in the country.

The 2022 edition of the program was implemented across 11 experimental sites located in 9 cities, in the three main areas of the countries: Milano, Torino, Venezia Marghera, in the North; Ancona, Aprilia, in the Center; Bari, Napoli, Rosarno and Palermo in the South. In 2022, the country continued to deal with the challenges posed by the COVID-19 pandemic and its protocols, e.g. in schools mask-wearing was obligatory and remote learning was activated in case of infections. Additionally, individuals who tested positive still had to undergo quarantine periods. The 100-hour activities spanned from morning to afternoon, encompassing lunch, and were offered free of cost to families.

The first component of the program consisted of 88 hours of educational and recreational workshops arranged in small groups (ten students each). The goal was to transfer educational content, mostly in mathematics and reading, through peer education, learning by playing and cooperative learning. Leisure activities, such as brief one-day excursions within the local community, were also arranged. The collaborative and playful approach of these activities also sought to cultivate learning through enjoyable means, thereby fostering mastery goal orientation, the inclination towards learning, and intrinsic motivation.

The second component was a 12-hour personalized tutoring intervention, which was delivered to groups of two or three students and whose major focus was helping students in specific areas where they had the greatest learning gaps. To maximize children's and their

families' engagement with the program, regular contact was maintained with parents via text messages.

The program was implemented by local organizations and educators across the 11 centers, following a common protocol which was devised, steered and monitored centrally by two non-profit organizations, *Save the Children Italia* and *Fondazione Agnelli*. The program was executed in an intensive format in eight centers (four weeks after the school year ended and the last week before the new school year started) or in an extensive one in the remaining three centers (every week but one or two in August). The fact that the program was centrally managed but implemented in a decentralized manner by local organizations, highly knowledgeable of the specific context, increases the potential of successfully scaling up or replicating such a program, a topic of intense debate in recent years in the evaluation literature (List, Suskind & Supplee, 2021; Attanasio et al., 2022).

Design

Recruitment & randomization

Shortly before the summer break, 17 primary and middle schools located in the 9 cities referred 1,634 students in need of learning support during summer, following the instructions of the implementing organizations. The implemented organizations contacted the parents or legal guardians of the students and invited them to enroll in the study. 1,038 students received the participation consent and were subsequently randomly allocated to either the treatment or the control group. The randomization was performed at the level of the family (i.e., siblings were allocated to the same treatment arm) and was stratified by school and educational level (primary or middle school). The initial target was to have 60 treated students in each experimental site. Due to the differences in the number of eligible students per site, the probability of being treated varies across sites and strata. 722 students were allocated to the treatment group.[3] The remaining 316 were allocated to the control condition and offered the possibility of participating in a shorter learning support program in the Fall, after the follow-up survey. All applicants were informed about the randomization study and were therefore aware of their status. Further details are provided in Appendix II.

[3] 94 of them were initially allocated to a reserve group, but all of them received the invitation to participate only a few days later.

Data Collection and measures

The baseline data collection was conducted between late May and early June 2022, before the end of the school year. The survey was composed of achievement tests and was implemented at school by the program staff in digital format. Each student took four grade-specific achievement tests: reading literacy (grammar and reading comprehension) and mathematics (arithmetic and geometry). The tests were created specifically for the study and piloted by several teachers in different classrooms. Simplified versions were administered to students labeled as having 'special education needs' by schools. Most of the items on the tests were multiple-choice questions, mainly with a single correct answer.[4] The total score on a given test is given by the share of items with a correct answer, rescaled to vary from 0 (no correct item) to 10 (all items were correct).[5]

The follow-up survey was identical to the baseline survey and was administered by the program staff at the end of the summer during the last week of the program or the first two weeks of the new school year. Further details on the achievement test can be found in Appendix I.[6]

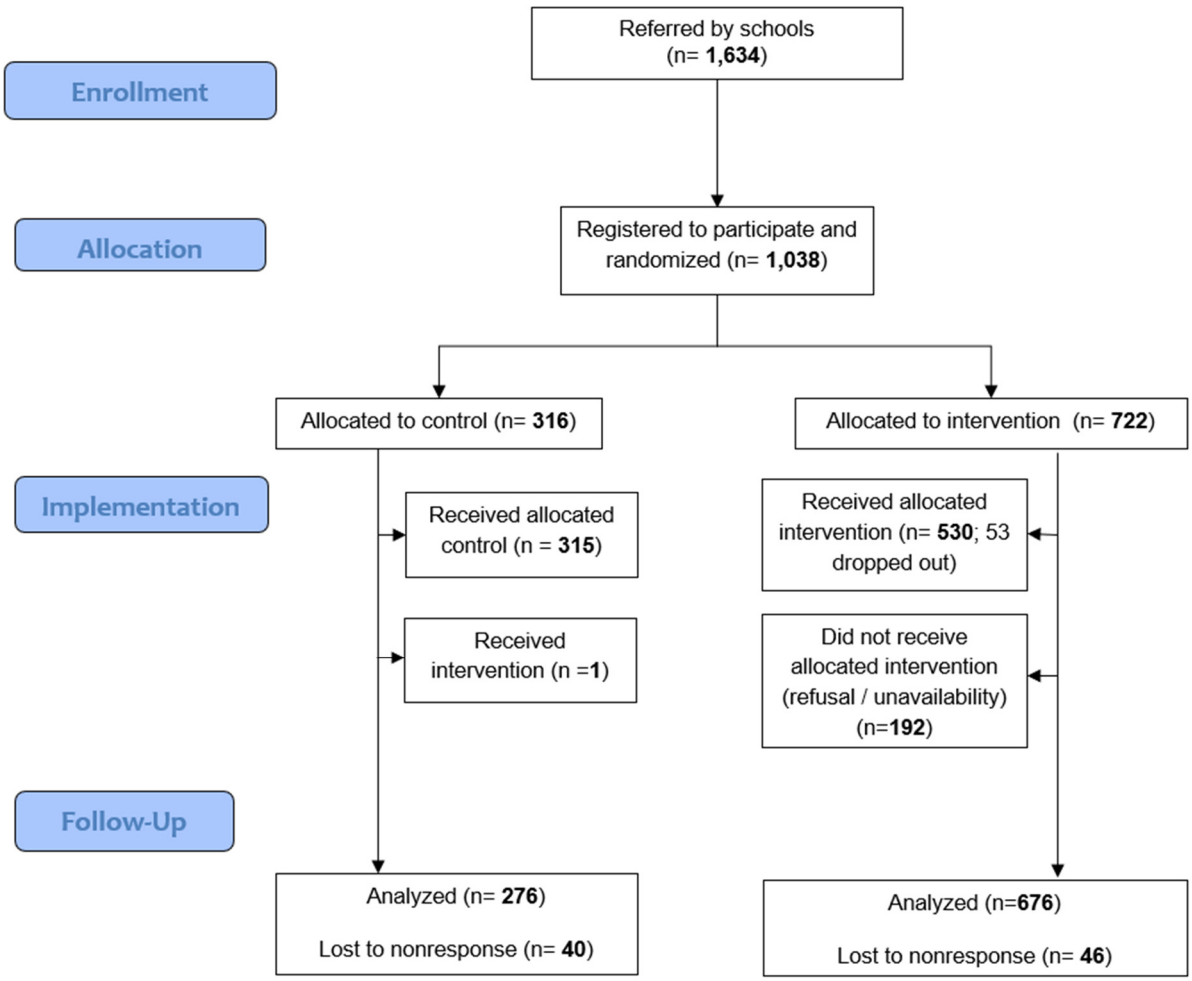
In addition to the tests, several psychometric measures of noncognitive skills were also collected at baseline and follow-up, further detailed in Appendix V. Finally, at baseline schools provided administrative data on the students, containing socio-demographic information, school grades, absenteeism, indicators of need for support various subjects and deprivation indicators (see Table 1).

[4] Few questions had multiple correct answers or true or false response options.

[5] In the case of multiple choice questions, they were scored based on the share of correct answers identified. For instance, if a given item had two correct responses but the student selected only one, the score on this item would be 1/2. If any incorrect response option was selected, the assigned score on the item was 0.

[6] All experimental sites, except for Palermo, Napoli P, and Torino A, administered the follow-up tests at two distinct time points for treated students and control group students. In fact, the former mostly completed the questionnaire in the last days of the program at the beginning of September (i.e. the week before the start of the new school year), while the latter filled it out in the initial days of the school year. The remaining three experimental sites conducted the endline survey simultaneously for both groups, with the tests being administered to both groups when school started. In this case, data collection predominantly occurred during the second week of classes. See Figure AI.2 in Appendix I.

Figure 1 CONSORT flow diagram



Findings

Sample description

The sample is relatively balanced in terms of gender with 45% of the sample being female students. About one-third of the applicants are of migration background, the share being larger in the northern cities. A similar share of students was reported to have special education needs—and thus entitled by law to receive personalized support (see Table A6).[7] In terms of grades, reported separately for primary and secondary school due to different scoring systems, on

[7] By BES, we refer to a condition, continuous or transitory, that hinders a pupil's learning and that requires appropriate attention from the school in order to succeed in their studies. BES are divided into three categories: Disability, Specific Learning Disorders and/or Specific Developmental Disorders and Socio-economic, Linguistic or Cultural Disadvantage

average grades are relatively low, close to the minimum passing grade (2 in primary school and 6 in secondary school). Schools provided three qualitative indicators of deprivation, varying from 1 to 5, with educational deprivation being the most prevalent in the sample. Finally, schools also indicated in which area each student required personalized support; the vast majority of students were labeled as needing support in literacy (reading and writing) and math/science. There are differences between primary and secondary school pupils, math being an area of more major need among secondary pupils, while literacy needs are more prevalent among primary school pupils. About one-fifth of pupils were reported as needing support to increase their base knowledge of the Italian language, which can be seen as a proxy for migrant background.

The implementation of the randomization protocol in the different experimental sites was monitored throughout the field operations and no deviations were detected. T-tests and F-tests were run to check the statistical equivalence of the groups at baseline and after attrition in the late summer test. These tests demonstrate the very high balance of the observable covariates across the two groups. More details about the randomization and all integrity checks are included in Appendix II.

Table 1 Descriptive Statistics

	Primary school			Secondary school			Total		
Variabile	Average	SD	N	Average	SD	N	Average	SD	N
Female	0.44	0.50	425	0.46	0.50	613	0.45	0.50	1,038
Native	0.59	0.49	425	0.68	0.47	613	0.64	0.48	1,038
Age	10.38	0.82	425	12.56	0.90	613	11.67	1.38	1,038
Special Edu. Needs (BES)	0.35	0.48	425	0.30	0.46	613	0.32	0.47	1,038
Average no. of absences a.s. 2021/22	18.90	15.45	421	26.27	31.69	604	23.24	26.50	1,025
School Achievements (first-semester s.y. 2021/22)									
Primary (1 min- 4 max), Secondary (1 min.10 max)									
Italian language	2.41	0.78	425	6.05	1.07	613			
History	2.38	0.82	425	5.97	1.15	613			
Geography	2.39	0.81	425	6.08	1.14	613			
Math	2.42	0.81	425	5.88	1.18	613			
Science	2.43	0.82	425	6.01	1.12	613			
English	2.49	0.82	425	6.12	1.25	613			
Average deprivation in (1min, 5max)[a]									
Material	2.00	1.13	417	1.92	1.09	607	1.95	1.11	1,024
Affective	2.18	1.24	417	2.31	1.17	609	2.25	1.20	1,026
Educational	2.46	1.22	415	2.50	1.25	609	2.48	1.24	1,024
Need of personalized support in[a]									
Literacy skills	88.04	32.49	976	77.78	41.61	976	82.17	38.29	976
Math and science skills	80.14	39.94	976	84.77	35.97	976	82.79	37.77	976
Italian language	23.44	42.42	976	17.92	38.39	976	20.29	40.23	976

Note: The information was provided by the schools before the end of the school year. "Native" indicates students with no reported migratory background. BES indicates students officially classified as having special education needs.

[a] the material deprivation indicators are qualitative assessments provided by the schools which indicate in which areas each student required personalized support.

Take up and implementation

530 out of the 722 students (73%) assigned to the treatment group accepted to join the program. On average, they participated in 61 hours of workshops and 10 hours of tutoring, for a total of 71 hours in relation to the scheduled 100.[8] Those assigned to the treatment who did not participate in the activities (no-shows)[9] either refused to participate when contacted by the program managers (the majority),[10] or never showed up for the activities.[11] The incidence of no-shows was slightly higher in the sites that implemented an intensive program format (77%) than those implementing the extensive one (72%). Similarly, among participants, the average number of hours attended was slightly lower in the intensive centers (69 hours) as opposed to the extensive ones (75 hours).[12]

Participants were generally satisfied with the program activities. Primary school students appeared to be slightly more satisfied than secondary students with the workshops (67% vs. 58%)[13], tutoring (55% vs. 42%), and local excursions (72% vs. 64%). Students perceived the program to be especially beneficial for 'making new friends' (76% for primary students and 69.5% for secondary students) and 'learning new things' (68% vs. 55%), while lower levels of agreement were expressed regarding the usefulness of the program for 'reviewing the content of the past school year' (47% vs 38%) and 'completing summer homework' (40% vs 35%).

Treatment effect estimates

The evaluation provides both Intent-to-Treat (ITT) and Treatment-on-Treated (TOT)[14] effects of the program. The estimates were obtained through linear regression models, which accounted for the randomization strata, the pre-treatment value of the outcome variable, and several

---

[8] The distribution is strongly skewed towards the maximum number of hours, the median being roughly 80 hours and almost 30% of students attending at least 90 hours. See Figure AIII.1 in the Appendix

[9] The study has primarily one-sided non-compliance due to no-shows. Only one student assigned to the control group ended up participating in the program.

[10] One of the main reasons for refusal was that the program conflicted with other family arrangements, during the summer. In the case of students from parents with a migratory background, many were traveling to their parents' home country during the summer. Part of these students were swiftly replaced with other students initially randomized to the waiting list. Those on the waiting list who refused to participate or never showed up are also included in the "no-show".

[11] No particular differences were found according to student characteristics with regard to non-participation or attendance rates. The only evidence found was that students with higher absence rates during the year were also those who attended fewer hours of the summer intervention. Furthermore, girls show fewer hours of attendance than boys.

[12] We attempted to investigate whether certain initial student characteristics were predictive of the modes and intensity of participation in the intervention activities. Overall, it's challenging to outline a well-defined profile of a regularly attending student. However, from these analyses, it becomes evident that students with higher absence rates throughout the year are also the ones who attended fewer hours of the summer intervention. Another factor negatively correlated with the attended hours is gender, with girls demonstrating fewer attendance hours compared to boys.

[13] The percentages are computed by dichotomising a Likert scale: 1 if the responses were "Very Satisfied" or "Extremely satisfied" and 0 if the responses were "Somewhat satisfied", "Little satisfied" or "Not satisfied at all"

[14] The TOT estimates were obtained through two-stage least-squares regressions (local average treatment effects), by instrumenting in the first stage actual treatment take-up with the offer to enroll in the program.

baseline characteristics. Covariates were included to enhance the precision of the estimates and do not qualitatively alter the results (see Appendix IV).

Table 1 shows the estimated effects of the program on the four achievement tests separately and on the aggregated scores by subject area (math and reading literacy).

The estimated ITTs and TOTs are positive and highly statistically significant for reading comprehension (ITT 0.332 or 0.169 SD; TOT 0.451 or 0.229 SD), while smaller and significant only at 10% for grammar (ITT 0.171 or 0.103 SD; TOT 0.233 or 0.140 SD). When considering the mathematical subjects, arithmetic and geometry, ITT and TOT estimates are positive but imprecisely estimated. ITTs and TOTs correspond to 0.195 (0.097 SD) and 0.267 (0.133 SD) for arithmetic, while in geometry the estimates are 0.104 (0.049 SD) and 0.142 (0.067 SD).

When the test scores are aggregated by subjective area, which reduces noise, the resulting effects for reading literacy are positive and statistically significant (ITT: 0.259 or 0.167 SD, TOT: 0.351 or 0.226 SD) and positive but only significant at 10% for mathematics (ITT: 0.166 or 0.095 SD, TOT: 0.226 or 0.13 SD).

Table 1 Effects of the program on achievement tests in mathematics and language

	Arithmetic	Geometry	Math	Comprehension	Grammar	Literacy
Control Mean (SD)	3.92 (2.01)	3.78 (2.13)	3.85 (1.74)	5.16 (1.97)	4.69 (1.66)	4.92 (1.55)
ITT (s.e)	0.195 (0.112)	0.104 (0.405)	0.166* (0.086)	0.332*** (0.007)	0.171* (0.098)	0.259*** (0.003)
TOT (s.e.)	0.267 (0.101)	0.142 (0.390)	0.226* (0.077)	0.451*** (0.005)	0.233* (0.088)	0.351*** (0.003)
N	939	946	947	934	936	947

Note: Estimates are obtained via OLS regressions. Normalized outcomes (0-10). Models control for randomization strata, pre-treatment outcomes and covariates (i.e., gender, age, migration background, grades, school days skipped, need for support in Italian, math or science, and deprivation indices). In cases of missing values, the median value of each stratum and the mean value per test type were imputed. Imputation was only applied to baseline values. Standard errors clustered at the family level in parentheses.
* p<.10, ** p<.05, *** p<.01

Robustness checks

In addition to incorporating the pre-treatment values of the outcomes and a rich set of covariates into the models, we perform a series of leave-one-out tests to ensure that the results are not driven by any specific experimental site. In these tests, we essentially replicate the main

estimates, but removing the 11 sites from the sample one at a time. We also include weights in the estimation models to account for the varying probability, within the strata, of being assigned to the treatment group. The results of these tests corroborate the obtained conclusions and are presented in Appendix II.

Heterogeneity

Exploratory analyses were also run to establish if the program's effects are stronger among particular subgroups. We focus on the following dimensions: school level (primary school students vs. secondary school students), having a migratory background, special educational needs (BES) condition, scores in the achievement tests at baseline (above or below the median), organizational model of the site (intensive vs. extensive), students' gender (male vs. female).

Results show that effects are largely driven by students with special education needs and by primary school students (Appendix IV). The fact that the program appears to be more effective among students identified as having special education needs may facilitate effective targeting if the program is scaled up or replicated, as schools throughout the country perform such evaluation of students. It is challenging to pin down why the program is less effective for secondary school students. While we can only speculate, it is possible that the activities are better suited for younger students or that learning gaps are more difficult to address in higher grades when the level of complexity of subjects increases. The former interpretation is in line with the lower levels of satisfaction with the program among secondary school students commented previously.

Mechanisms and interpretation

Assessing the program's effects relative to summer learning losses

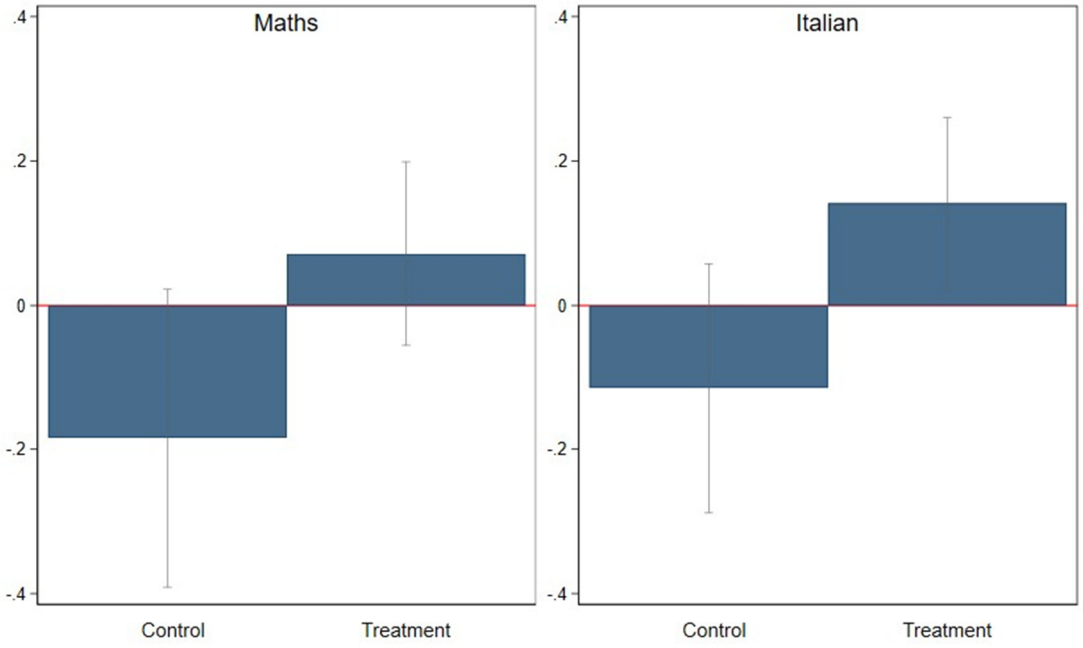
The fact that treated students have higher levels of competences in September than they would have had without the program, does not provide information about the actual capacity of the program to compensate for learning losses that may have occurred during the summer. A positive effect in September does not necessarily imply that the treatment group has higher levels of competence compared to before the summer break: treated students could have simply lost less ground compared to the control group but still fallen behind compared to the end of the previous school year.

To investigate this aspect, we leverage the availability of pre- and post-summer assessment tests. Figure 2 shows that control group students experienced learning losses in both

mathematics and reading literacy. In mathematics, the estimated loss is statistically significant at 10% and corresponds to 0.18 points, approximately 0.09 standard deviations. In reading literacy, the estimated losses are more modest and not statistically significant: 0.11 points, corresponding to about 0.06 SD.[15] We note that due to the same tests being administered at baseline and follow-up, the estimates of learning losses may be a lower bound due to memory effects.

In contrast to the control group, treated students did not experience any losses, displaying even actual gains. In mathematics, treated students gain 0.7 points (or 0.03 SD), imprecisely estimated, while in reading, the gain is larger and significant (+0.14 points or 0.08 SD). Therefore, the effects reported in Table 1 are the result of the losses experienced by controls and the gains benefiting treated students.

Figure 2 Summer Learning Growth, by experimental group



Note: The bars represent the unadjusted mean differences in normalized scores (0-10) measured in the post-summer and pre-summer assessments. * p<.10, ** p<.05, p<.01

Mechanisms behind the effects and the "non-cognitive" puzzle

The experimental design and data availability do not allow the disentangling of the separate effects of the two components of the programs, nor do they allow us to understand in detail which specific mechanisms explain the results and the heterogeneity in the effects. However,

[15] Losses estimated separately for the four subjects present a consistent picture but are not displayed here due to greater statistical uncertainty.

in addition to the achievement tests, students also responded to several psychometric scales, which allow to investigate — even if only in an explorative and suggestive fashion — three aspects that can provide a broader perspective on the effectiveness of summer learning programs, that are often neglected in the literature (Appendix V).

The first finding is that students in the treatment group reported a higher mastery goal orientation, which captures students' focus on developing and mastering new competencies. This can partly account for the estimated effects of the program on learning, suggesting that the effects were not solely due to a 'transfer of knowledge' that occurred during the summer activities. Instead, the program also increased students' interest in learning, possibly contributing to prolonged persistence of the learning effects over time.

The second aspect concerns students' attitudes toward the beginning of the new school year and the school environment in general. Students in the treatment group highlight that they feel more uncomfortable in class, they are more worried about classroom activities, and find classroom tasks less interesting. On one hand, this can be viewed as an indirect indication – in addition to what emerged from the satisfaction questions (see above) – of the quality of the experience they had during the archipelagos: students learned while having fun, in spaces that are less rigid than the "normal" school environment, to the extent that they feel more discouraged at the idea of returning to school. On the other hand, however, students must return to school in September; and returning with a negative attitude can create issues that affect well-being at school and future learning as well.[16]

The third aspect is related to the concern about school assignments and assessments. Treated students tend to be, on average, more worried about taking assignments and assessments. A moderate level of concern can be seen as an indication of students' consciousness and school commitment. However, the difference between the treated group and the control group is mainly concentrated in the category of "very concerned," suggesting that among the treated group, an excessive level of anxiety about assignments and assessments may have developed. This could also contribute to generating more negative attitudes in regards to going back to school.

[16] It should be noted that the majority of students interviewed in September had not yet returned to school, so their concern is related to the idea of returning to class, and it is not certain that this concern will persist at the actual resumption of school. The result theoretically may be influenced by the fact that the treatment group responds to the questionnaire the week before returning to school, while the control group typically does so during the first week. However, checks on subgroups of students who took the test concurrently confirm what is reported in the text.

Conclusions

The study contributes to the existing literature on summer learning programs by providing new experimental evidence from a previously unexplored national context, namely Italy. This research adds to the existing literature showing that summer learning programs have the potential to enhance students' learning outcomes and reduce learning gaps. Notably, the effects of the evaluated program are more pronounced and statistically significant for reading literacy (0.17 SD), while still positive though slightly weaker effects are observed for mathematics (0.09 SD).

It is worth highlighting that the program was targeted at students with low academic achievements, and even within this context, the effects were more substantial for those students who were identified by schools as having particularly severe educational needs. This finding is supportive of the notion that such interventions may yield greater effectiveness for the most vulnerable students.

Furthermore, the study also suggests that the program more than compensated for the (relatively) small summer learning losses estimated in the sample. Students in the treatment group exhibited zero to positive learning growths during summer, in contrast to students in the control group who experienced learning declines, particularly in mathematics.

Lastly, the study also explores often neglected non-cognitive aspects and shows that treated students accrue a higher interest in learning, which can be seen as a mechanism explaining the effects and as a driver for the persistence of the learning effects. At the same time, however, treated students report increased concern regarding going back to school in September and higher anxiety for school assignments and tests, which raises a question about the potential negative effects when students return to "normal" school.

The study is not without limitations. A first aspect pertains to the inability of the study to disentangle the distinct effects of individual program components. As a result, it does not allow for the identification of specific mechanisms driving the above-reported effects. This challenge is common in evaluation studies of summer learning programs (McCombs et al., 2019; Lynch et al., 2023). Our study utilized available non-cognitive survey data to shed light on potential mechanisms at play, but this evidence remains suggestive, requiring further research to advance our understanding.

A second consideration is the need to study further the heterogeneity of the effects. The study reveals that the program yielded significant, positive effects exclusively for the most vulnerable

students and primary students, with no such effects noted among secondary-level students. The reasons behind this discrepancy are unknown and call for the need for additional research.

Lastly, a potential critique is that the study only reports short-term effects, which could fade out after a few weeks of the new school year. It can be argued that the study demonstrates that treated students start the school year with a higher baseline of learning and that the non-cognitive data indicate heightened student interest in learning. However, students who participated in the program also showed an increased unease about returning to school and higher anxiety about assignments and tests. Thus, exploring the medium- and long-term effects of summer learning programs on both cognitive and non-cognitive student outcomes should be a priority of future studies.

Acknowledgements

The authors would like to thank Fondazione Agnelli (with Andrea Gavosto, Gianfranco De Simone, Barbara Romano, Giovanni Piumatti) and Save the Children Italia (with Raffaela Milano, Carlotta Bellomi, Francesca Bilotta, Alice Libianchi, Alessia Romeo) for allowing us to conduct this evaluation and for supporting us in the crucial steps of the study.

References

Alexander, K.L., Entwisle, D.R. and Olson, L.S., 2007. Lasting consequences of the summer learning gap. *American sociological review*, *72*(2), pp.167-180.

Anderman, E.M., Urdan, T. and Roeser, R., 2003, March. The patterns of adaptive learning survey: History, development, and psychometric properties. In *Indicators of Positive Development Conference, Washington, DC*.

Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., & Rubio-Codina, M. (2022). Early stimulation and nutrition: the impacts of a scalable intervention. *Journal of the European Economic Association*, 20(4), 1395-1432.

Cooper, H., Nye, B., Charlton, K., Lindsay, J. and Greathouse, S., 1996. The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of educational research*, *66*(3), pp.227-268.

Cooper, H., Charlton, K., Valentine, J.C., Muhlenbruck, L. and Borman, G.D., 2000. Making the most of summer school: A meta-analytic and narrative review. Monographs of the society for research in child development, pp.i-127.

Downey, D.B., Von Hippel, P.T. and Broh, B.A., 2004. Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*(5), pp.613-635.

Duckworth, A.L. and Quinn, P.D., 2009. Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment*, *91*(2), pp.166-174.

European Commission sep EACEA sep Eurydice, 2022. *The Organisation of school time in Europe. Primary and general secondary education – 2022sep2023*. Eurydice Facts and Figures. Luxembourg: Publications Office of the European Union.

Fälth, L., Nordström, T., Andersson, U.B. and Gustafson, S., 2019. An intervention study to prevent 'summer reading loss' in a socioeconomically disadvantaged area with second language learners. Nordic Journal of Literacy Research, 5(3), pp.10-23.

Kim, J.S. and Quinn, D.M., 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), pp.386-431.

Kover, D.J. and Worrell, F.C., 2010. The influence of instrumentality beliefs on intrinsic motivation: A study of high-achieving adolescents. *Journal of Advanced Academics*, *21*(3), pp.470-498.

List, J. A., Suskind, D., & Supplee, L. H. (Eds.). (2021). The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It. Routledge.

Lynch, K., An, L. and Mancenido, Z., 2023. The impact of summer programs on student mathematics achievement: A meta-analysis. *Review of Educational Research*, *93*(2), pp.275-315.

McEachin, A., Augustine, C.H. and McCombs, J., 2018. Effective Summer Programming: What Educators and Policymakers Should Know. American Educator, 42(1), p.10.

McCombs, J.S., Augustine, C.H., Unlu, F., Ziol-Guest, K.M., Naftel, S., Gomez, C.J., Marsh, T., Akinniranye, G. and Todd, I., 2019. Investing in successful summer programs. RAND Corporation. https:sepsepwww.rand.orgseppubssepresearch_reportssepRR2836.html.

OECD. 2018. PISA 2018 Technical Report, PISA, OECD, https://www.oecd.org/pisa/data/pisa2018technicalreport/

OECD. 2019. *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, *https:sepsepdoi.orgsep10.1787sep5f07c754-en*.

OECD. 2023. Population with tertiary education (indicator). doi: 10.1787sep0b8f90e9-en (Accessed on 03 April 2023)

Santibañez, L., & Guarino, C. M. (2021). The effects of absenteeism on academic and social-emotional outcomes: Lessons for COVID-19. Educational Researcher, 50(6), 392-400.

Skinner, E.A., Kindermann, T.A. and Furrer, C.J., 2009. A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and psychological measurement*, *69*(3), pp.493-525.

von Hippel, P.T. and Hamrock, C., 2019. Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them. *Sociological Science*, *6*, pp.43-80.

Workman, J., von Hippel, P.T. and Merry, J., 2023. Findings on summer learning loss often fail to replicate, even in recent data. *Sociological Science*, 10, pp.251-285.

Appendix I. Construction and distributions of learning outcomes

The achievement tests used in this study have been adapted by Fondazione Agnelli from material by Mondadori Education S.p.A. The tests are differentiated according to subject (grammar, comprehension, arithmetic and geometry) and grade. In addition, a simplified version of the test was prepared for students with Special Educational Needs (BES). Thus, 40 different tests were used (4 tests x 2 BES status x 5 grades).

The questionnaires were made accessible online, on the Google Forms platform, via the links provided to the schools in a document containing all the instructions for completion. The administration took place at school on computers or tablets.

In the first survey there were few cases of students filling in the same questionnaire twice or mistakenly filling in a questionnaire not directed at them. In these situations, it was decided to consider the first test filled in, even if it was not the one the student should have filled in. This is due to the risk of memory/learning effects or boredom.

The raw, non-normalized, test scores are based on a different number of questions depending on the test (from 6 to 23). To make the test scores comparable, the score of each test (for each individual grade, subject, and by special education needs status) was normalized between 0 and 10. A score of 0 indicates that none of the answers given in a given test are correct, while a value of 10 indicates that they are all correct.

We also calculated the standardized test scores from the normalized test scores subtracting from it the mean and dividing by standard deviation of the follow-up survey score of the control group.

Table AI.1 Response rates to the baseline survey on achievement tests

Experimental site	Control group	Treated group	Total
Ancona	92.6	79.7	83.3
Aprilia	100.0	97.6	98.6
Bari	94.1	88.4	90.3
Marghera	93.9	100.0	98.0
Milano	69.0	82.2	77.4
Napoli C	88.5	95.5	93.5
Napoli P	96.0	92.5	93.5
Palermo	91.7	92.4	92.2
Rosarno	100.0	98.5	98.7
Torino A	100.0	91.7	94.1
Torino V	95.8	94.9	95.2
Total	91.8	92.1	92.0

Figure AI.1 Distribution of learning outcomes in the baseline survey. Raw and normalized scores.

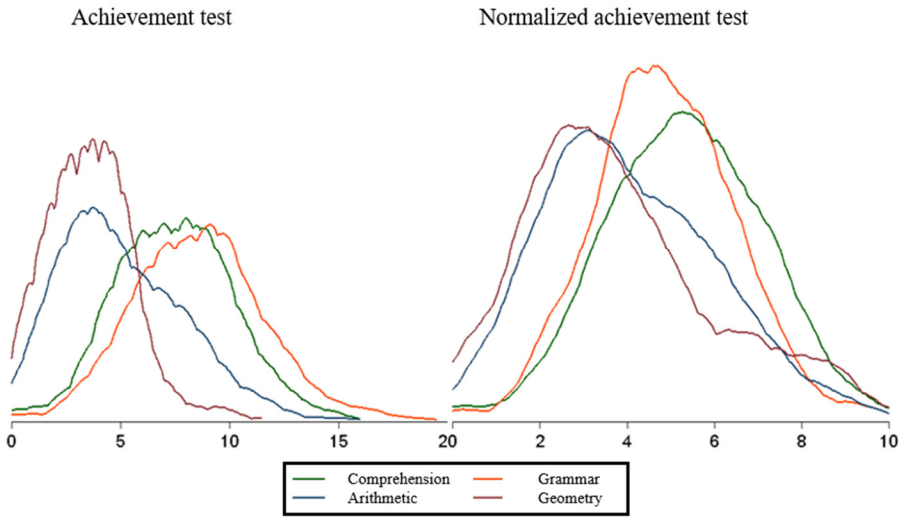


Table AI.2 Descriptive statistics of normalized achievement skills. Baseline survey.

Achievement test	School level	Test					Test BES				
		Average	SD	p25	p50	p75	Average	SD	p25	p50	p75
Arithmetic	P	4.3	1.9	2.9	3.9	5.4	5.4	2.4	4.3	5.7	6.9
	S	3.8	1.9	2.3	3.5	5.2	3.3	1.7	1.9	3.1	4.4
	Total	4.0	1.9	2.7	3.8	5.4	4.2	2.3	2.5	4.1	5.7
Comprehension	P	5.6	1.8	4.2	5.8	6.9	6.0	2.0	4.5	5.9	7.4
	S	5.1	1.7	3.9	5.0	6.1	4.9	2.0	3.7	4.8	6.2
	Total	5.3	1.7	3.9	5.3	6.6	5.4	2.1	4.1	5.4	6.7
Geometry	P	5.0	2.4	3.2	4.8	6.7	4.4	2.4	2.5	3.9	6.7
	S	3.2	1.6	1.9	3.1	4.2	2.9	2.0	1.3	2.5	4.4
	Total	3.9	2.2	2.3	3.5	5.0	3.6	2.4	2.0	3.1	5.0
Grammar	P	5.0	1.8	3.8	5.0	6.3	4.7	1.7	3.7	4.7	6.1
	S	4.7	1.4	3.8	4.7	5.6	5.0	1.8	3.8	5.0	6.2
	Total	4.8	1.6	3.8	4.7	5.9	4.9	1.8	3.8	4.8	6.2

Figure AI.2 Data collection timeline, by experimental site

Appendix II. Randomization and integrity checks

Table AII.1 shows the distribution of students in the randomized groups by experimental site. The randomization was performed by the evaluation team (i.e., the authors of this study) using the statistical software R. In practice the unit of randomization is the family, however there were less than 10% of pupils who had a sibling in the sample. The randomization was stratified by school and the educational level (primary or middle school).

Table AII.1 Distribution of enrolled students by experimental site and treatment group

Experimental site	Control group	Treatment group			Total
		Treated	Reserve group	Total Treated	
Ancona	27	59	10	69	96
Aprilia	54	70	15	85	139
Bari	34	60	9	69	103
Marghera	33	60	5	65	98
Milano	42	60	13	73	115
Napoli C	26	60	7	67	93
Napoli P	25	60	7	67	92
Palermo	24	60	6	66	90
Rosarno	12	60	6	66	78
Torino A	15	29	7	36	51
Torino V	24	50	9	59	83
Total	316	628	94	722	1038

We performed t and F tests to check group balance in the initial sample (tables AII.2) and in the final one (Table AII.3), after taking into account attrition in the follow-up survey (Table AII.4). The t test values are obtained from OLS regressions models using as outcomes each of the variables listed and as covariates the treatment status and strata fixed effects. The F test comes from an OLS model regressing the treatment status on all the variables listed in Table AII.2 in addition to strata fixed effects. The null hypothesis is that all the coefficients on the variables listed are 0.

Table AII.2 Equivalence tests conducted on the baseline sample

	N	Control Group Average	Treated Group Average	T-C	Sign.
Achievement tests					
Comprehension	936	0.08	0.06	-0.02	
Grammar	932	0.07	0.10	0.03	
Arithmetic	934	0.08	0.05	-0.03	
Geometry	936	0.07	-0.01	-0.08	
Non-cognitive tests					
Extrinsic motivation	965	0.24	0.23	-0.01	
Intrinsic motivation	965	0.18	0.16	-0.02	
Performance-approach goal orientation	965	0.00	0.03	0.03	
Performance-avoid goal orientation,	965	0.16	0.02	-0.14	*
Mastery goal orientation	965	0.31	0.32	0.01	
Behavioral engagement	965	0.01	0.00	-0.01	
Behavioral disaffection	965	-0.05	-0.02	0.03	
Emotional engagement	965	0.16	0.17	0.01	
Emotional disaffection	965	0.02	0.07	0.05	
Consistency of interest	965	0.09	0.02	-0.07	*
Perseverance of effort	965	0.02	0.12	0.10	
Resilience	965	0.30	0.32	0.02	
Characteristics					
Female	1038	0.44	0.46	0.02	
Native	1038	0.62	0.65	0.03	
Age	1038	11.67	11.66	-0.01	
Special Education Needs (BES)	1038	0.32	0.32	0.00	
Average no. of absences a.s. 2021/22	1025	22.14	23.73	1.59	
School Achievements (1min- 4 max)					
Italian language	1026	2.06	2.05	-0.01	
History	1024	2.05	2.01	-0.04	
Geography	1024	2.12	2.02	-0.1	*
Math	1024	2.07	1.99	-0.08	
Science	1022	2.08	2.03	-0.05	
English	1025	2.12	2.11	-0.01	
Deprivation (1min, 5max)					
Material	1024	1.96	1.95	-0.01	
Affective	1026	2.22	2.27	0.05	
Educational	1024	2.48	2.48	0.00	
It needs personalized support in					
Literacy skills	976	0.80	0.83	0.03	
Math and science skills	976	0.79	0.84	0.05	
Italian language	976	0.25	0.18	-0.07	*

Note: F(41,961)=1.26 Prob>F=0.13 calculated based on a regression model with imputed missing data. * p<.10, ** p<.05, p<.01

Table AII.3 Equivalence tests conducted on the follow-up sample

	N	Control Group Average	Treated Group Average	T-C	Sign.
Achievement tests					
Comprehension	867	0.09	0.07	-0.02	
Grammar	864	0.05	0.11	0.06	
Arithmetic	866	0.12	0.06	-0.06	
Geometry	867	0.11	-0.01	-0.12	
Non cognitive tests					
Extrinsic motivation	896	0.21	0.25	0.04	
Intrinsic motivation	896	0.16	0.18	0.02	
Performance-approach goal orientation	896	0.01	0.04	0.03	
Performance-avoid goal orientation,	896	0.16	0.04	-0.12	
Mastery goal orientation	896	0.29	0.33	0.04	
Behavioral engagement	896	-0.01	0.00	0.01	
Behavioral disaffection	896	-0.05	-0.03	0.02	
Emotional engagement	896	0.16	0.18	0.02	
Emotional disaffection	896	0.02	0.07	0.05	
Consistency of interest	896	0.07	0.02	-0.05	
Perseverance of effort	896	0.02	0.13	0.11	
Resilience	896	0.27	0.33	0.06	
Characteristics					
Female	954	0.43	0.46	0.03	
Native	954	0.63	0.66	0.03	
Age	954	11.61	11.65	0.04	
Special Education Needs (BES)	954	0.33	0.31	-0.02	
Average no. of absences a.s. 2021/22	941	21.82	23.28	1.46	
School Achievements (1 min- 4 max)					
Italian language	943	2.03	2.06	0.03	
History	942	2.01	2.01	0.00	
Geography	942	2.09	2.03	-0.06	
Math	943	2.03	2.00	-0.03	
Science	941	2.04	2.04	0.00	
English	943	2.07	2.11	0.04	
Deprivation (1min, 5max)					
Material	941	1.97	1.94	-0.03	
Affective	943	2.23	2.25	0.02	
Educational	941	2.48	2.47	-0.01	
It needs personalized support in					
Literacy skills	895	0.79	0.83	0.04	
Math and science skills	895	0.81	0.84	0.03	
Italian language	895	0.24	0.18	-0.06	

Note: F(41,892)=1.1 Prob>F=0.32 calculated based on a regression model with imputed missing data. * p<.10, ** p<.05, p<.01

Table AII.4 Endline survey's response rates and differential attrition

Experimental site	Control group	Treated group	Total	Diff. Att.
Ancona	85.2	92.8	90.6	7.6
Aprilia	81.5	85.9	84.2	4.4
Bari	100.0	91.3	94.2	-8.7
Marghera	97.0	100.0	99.0	3.0
Milano	78.6	94.5	88.7	15.9
Napoli C	92.3	91.0	91.4	-1.3
Napoli P	96.0	98.5	97.8	2.5
Palermo	83.3	92.4	90.0	9.1
Rosarno	66.7	95.5	91.0	28.8
Torino A	80.0	91.7	88.2	11.7
Torino V	91.7	94.9	94.0	3.2
Total	87.3	93.4	91.5	6.0

Appendix III. Implementation statistics

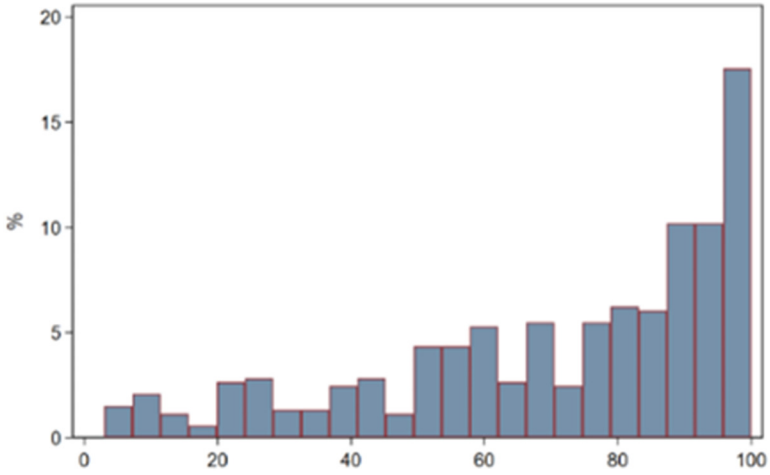
Table AIII.1 Participants, dropouts, and no-shows, for extensive and intensive formats

Format	Assigned to the treatment group	No-show	Treatment group
Intensive	515	145	370
Extensive	207	47	160
Total	722	192	530

Table AIII.2 Actual hours attended per type of activity and for extensive and intensive formats (average values).

Format	Workshop	Tutoring	N
Intensive	59.9	9.4	371
Extensive	64.2	10.7	159
Total	61.2	9.8	530

Figure AIII.1 Distribution of total actual hours attended



Appendix IV. Estimates and robustness checks

Table AIV.1 ITT estimates on students' learning

ITT	(1)	(2)	(3)	(4)	N
Arithmetic	0.165	0.205	0.195	0.179	939
	(0.229)	(0.105)	(0.112)	(0.129)	
Geometry	0.103	0.126	0.104	0.100	946
	(0.458)	(0.317)	(0.405)	(0.411)	
Mathematics	0.133	0.185*	0.166*	0.153	949
	(0.236)	(0.062)	(0.086)	(0.100)	
Comprehension	0.355***	0.335***	0.332***	0.328***	934
	(0.008)	(0.007)	(0.007)	(0.007)	
Grammar	0.277**	0.189*	0.171*	0.169*	936
	(0.019)	(0.073)	(0.098)	(0.095)	
Reading literacy	0.320***	0.272***	0.259***	0.253***	947
	(0.002)	(0.003)	(0.003)	(0.003)	
Randomiz. strata	Y	Y	Y	Y	
Clustered standard errors (family level)	Y	Y	Y	Y	
Outcome pre-treat.	N	Y	Y	Y	
Covariates	N	N	Y	Y	
IPW adjustment	N	N	N	Y	

Note: Normalized scores (0-10), robust standard errors in parentheses

Table AIV.2 Leave-1-out tests on achievement tests

Excluding the experimental site:	Arithmetic	Geometry	Math	Comprehension	Grammar	Literacy
Ancona	0.211*	0.099	0.178*	0.400***	0.207*	0.308***
	(0.128)	(0.130)	(0.101)	(0.127)	(0.110)	(0.093)
Aprilia	0.242*	0.089	0.173*	0.271**	0.067	0.176*
	(0.133)	(0.133)	(0.102)	(0.132)	(0.107)	(0.093)
Bari	0.182	0.081	0.152	0.287**	0.107	0.214**
	(0.125)	(0.133)	(0.101)	(0.127)	(0.106)	(0.091)
Marghera	0.181	0.172	0.196*	0.444***	0.298***	0.382***
	(0.131)	(0.132)	(0.103)	(0.133)	(0.108)	(0.095)
Milano	0.114	0.116	0.132	0.356***	0.131	0.238**
	(0.131)	(0.132)	(0.102)	(0.130)	(0.110)	(0.094)
Napoli C	0.195	0.124	0.169	0.329**	0.196*	0.265***
	(0.130)	(0.132)	(0.103)	(0.131)	(0.111)	(0.095)
Napoli P	0.235*	0.059	0.167	0.350***	0.200*	0.290***
	(0.130)	(0.131)	(0.102)	(0.127)	(0.108)	(0.092)
Palermo	0.175	0.127	0.175*	0.281**	0.166	0.238**
	(0.129)	(0.127)	(0.101)	(0.128)	(0.110)	(0.093)
Rosarno	0.201	0.099	0.167*	0.343***	0.179*	0.269***
	(0.125)	(0.128)	(0.099)	(0.125)	(0.105)	(0.091)
Torino A	0.240*	0.115	0.187*	0.345***	0.175	0.267***
	(0.124)	(0.126)	(0.098)	(0.123)	(0.107)	(0.090)
Torino V	0.157	0.057	0.118	0.248**	0.148	0.199**
	(0.126)	(0.131)	(0.101)	(0.126)	(0.108)	(0.093)

Note: Normalized scores (0-10), robust standard errors in parentheses

Table AIV.3 Heterogeneity of the effects (ITT) of *Arcipelago Educativo* on achievement tests

	Arithmetic	Geometry	Math	Comprehension	Grammar	Literacy
Baseline level >median vs. <median						
Treated	0.171	0.071	0.070	0.316*	0.179	0.253*
	(0.170)	(0.170)	(0.125)	(0.180)	(0.143)	(0.132)
Treated X Baseline lev.	0.061	0.030	0.143	0.028	0.005	0.056
	(0.253)	(0.249)	(0.200)	(0.250)	(0.213)	(0.188)
Secondary school vs. primary						
Treated	0.407*	0.453**	0.461***	0.504**	0.275*	0.382**
	(0.220)	(0.229)	(0.178)	(0.210)	(0.167)	(0.153)
Treated X Secondary	-0.353	-0.579**	-0.489**	-0.286	-0.174	-0.205
	(0.262)	(0.267)	(0.207)	(0.259)	(0.213)	(0.188)
Extensive format vs. intensive						
Treated	0.083	0.201	0.165	0.468***	0.187	0.332***
	(0.148)	(0.155)	(0.118)	(0.151)	(0.120)	(0.106)
Treated X Ex.format	0.322	-0.281	0.001	-0.396	-0.049	-0.215
	(0.273)	(0.256)	(0.210)	(0.255)	(0.224)	(0.190)
Female vs. male						
Treated	0.371**	0.060	0.221*	0.261	0.067	0.162
	(0.171)	(0.169)	(0.133)	(0.161)	(0.144)	(0.119)
Treated X Female	-0.402*	0.100	-0.125	0.161	0.238	0.222
	(0.234)	(0.241)	(0.183)	(0.239)	(0.197)	(0.168)
Native vs. children of immigrant						
Treated	0.070	-0.018	0.085	0.279	-0.063	0.159
	(0.221)	(0.214)	(0.171)	(0.214)	(0.173)	(0.148)
Treated X Native	0.195	0.192	0.127	0.083	0.365*	0.156
	(0.268)	(0.260)	(0.209)	(0.263)	(0.215)	(0.186)
Special needs students (BES) vs. non-BES						
Treated	-0.002	-0.047	0.001	0.182	0.046	0.123
	(0.139)	(0.148)	(0.115)	(0.143)	(0.117)	(0.105)
Treated X BES	0.596**	0.458	0.502**	0.459*	0.377	0.415**
	(0.279)	(0.279)	(0.217)	(0.278)	(0.240)	(0.203)

Note: normalized scores (0-10), robust standard errors in parentheses

Appendix V. Non-cognitive outcomes

The non-cognitive tests, unlike the achievement tests, were the same for all students. The tests were administered online on the REDcap platform.

There were in total twelve validated psychometric scales with varying number of items on the following dimensions: study motivation (Kover and Worrell 2010), engagement versus disaffection with learning (Skinner et al. 2009), adaptive learning (Anderman et al. 2003), consistency of interests, grit (Duckworth and Quinn 2009), and resilience (OECD, 2018).

The raw test score is calculated as the average of the items in any given single scale, and can range from 1 to 5. The score was then normalized between 0 and 10 so that it would be comparable with the achievement test score.

Table AV.1 Response rates to the baseline survey on non-cognitive tests

Experimental site	Control group	Treated group	Total
Ancona	88.9	82.6	84.4
Aprilia	98.1	96.5	97.1
Bari	94.1	89.9	91.3
Marghera	93.9	100.0	98.0
Milano	78.6	80.8	80.0
Napoli C	96.2	100.0	98.9
Napoli P	100.0	94.0	95.7
Palermo	95.8	93.9	94.4
Rosarno	100.0	97.0	97.4
Torino A	100.0	97.2	98.0
Torino V	95.8	89.8	91.6
Total	93.7	92.7	93.0

Figure AV.1 Distribution of non-cognitive outcomes in the baseline survey. Raw scores.

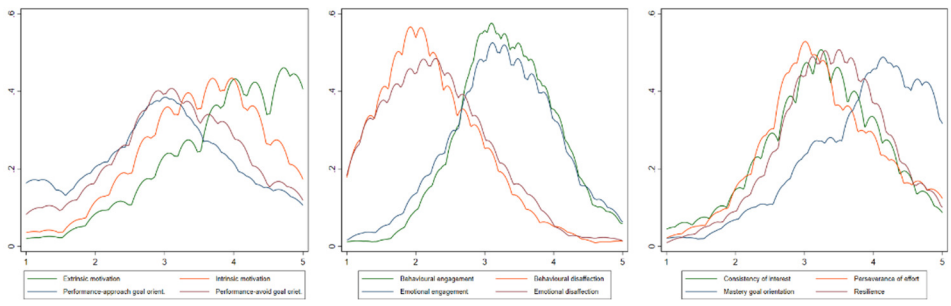


Table AV.2 Descriptive statistics of normalized non-cognitive skills. Baseline survey.

Non-cognitive test	level	average	sd	p25	p50	p75
Extrinsic motivation	P	8.2	2.1	7.5	9.3	10.0
	S	7.2	2.3	5.8	7.5	9.3
	Total	7.6	2.3	6.8	8.3	9.3
Intrinsic motivation	P	7.5	2.3	6.8	8.3	9.3
	S	6.1	2.2	5.0	6.8	7.5
	Total	6.7	2.3	5.0	6.8	8.3
Performance-approach goal orientation	P	5.5	2.8	3.3	5.8	7.5
	S	4.6	2.7	2.5	5.0	7.0
	Total	5.0	2.8	3.3	5.0	7.0
Performance-avoid goal orientation	P	5.8	2.6	3.8	5.8	7.5
	S	5.7	2.5	3.8	5.8	7.5
	Total	5.7	2.5	3.8	5.8	7.5
Mastery goal orientation	P	8.1	2.0	7.5	8.8	9.5
	S	7.3	2.1	6.3	7.5	8.8
	Total	7.6	2.1	6.3	8.3	9.5
Behavioral engagement	P	6.2	1.9	5.0	6.0	7.5
	S	5.4	1.8	4.0	5.5	6.5
	Total	5.7	1.8	4.5	5.5	7.0
Behavioral disaffection	P	2.7	1.9	1.5	2.5	3.5
	S	3.1	2.0	1.5	3.0	4.5
	Totale	2.9	1.9	1.5	2.5	4.0
Emotional engagement	P	6.7	2.0	5.5	7.0	8.0
	S	5.7	2.0	4.5	6.0	7.0
	Total	6.1	2.1	5.0	6.5	7.5
Emotional disaffection	P	2.8	1.9	1.5	2.5	4.0
	S	3.4	2.1	2.0	3.5	5.0
	Total	3.2	2.0	1.5	3.0	4.5
Consistency of interest	P	6.0	2.3	4.5	6.3	7.5
	S	5.8	2.1	4.5	5.8	7.0
	Total	5.9	2.2	4.5	5.8	7.5
Perseverance of effort	P	6.6	2.1	5.0	7.0	8.3
	S	5.1	2.0	3.8	5.0	6.3
	Total	5.7	2.2	4.5	5.8	7.5
Resilience	P	6.8	2.1	5.5	7.0	8.5
	S	6.1	2.0	4.8	6.5	7.5
	Total	6.4	2.1	5.0	6.5	8.0

Table AV.3 Effects of *Arcipelago Educativo* on non-cognitive skills

	Average control group	ITT	TOT	N
Motivation				
Extrinsic motivation	7.1	0.062	0.083	927
		(0.149)	(0.193)	
Intrinsic motivation	6.3	-0.094	-0.126	926
		(0.148)	(0.193)	
Adaptive learning				
Performance-approach goal orientation	4.9	-0.135	-0.180	926
		(0.175)	(0.227)	
Performance-avoid goal orientation	5.6	-0.087	-0.116	926
		(0.166)	(0.215)	
Mastery goal orientation	6.9	0.263*	0.350**	926
		(0.137)	(0.178)	
Engagement versus disaffection with learning				
Behavioral engagement	5.7	0.033	0.044	928
		(0.109)	(0.141)	
Behavioral disaffection	3.0	0.191	0.255	928
		(0.133)	(0.172)	
Emotional engagement	5.8	-0.203*	-0.270*	928
		(0.123)	(0.160)	
Emotional disaffection	3.0	0.360***	0.480***	928
		(0.131)	(0.171)	
Other personality traits				
Consistency of interest	5.8	-0.267*	-0.356*	926
		(0.147)	(0.191)	
Perseverance of effort	5.5	0.040	0.054	926
		(0.149)	(0.194)	
Resilience	5.8	0.220*	0.293*	926
		(0.125)	(0.162)	

Note: normalized scores (0-10), clustered standard errors in parentheses