I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17593

The Power of Language: Educational and Mental Health Impacts of Language Training for Refugee Children

Semih Tumen
Michael Vlassopoulos
Jackline Wahba

JANUARY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17593

The Power of Language: Educational and Mental Health Impacts of Language Training for Refugee Children

Semih Tumen
Amazon, TED University and IZA

Michael Vlassopoulos
University of Southampton, IFS and IZA

Jackline Wahba
University of Southampton and IZA

JANUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Power of Language: Educational and Mental Health Impacts of Language Training for Refugee Children*

This paper examines the causal impact of a large-scale Turkish language training program on the academic performance, school participation, and mental well-being of Syrian refugee children in T̈urkiye. Using rich administrative data and a staggered difference-in-differences design, we find that the program led to improvements in Turkish language and Math scores of refugee students, along with a significant reduction in their school absences. The language gains and associated improvements in school outcomes are more pronounced for younger refugee children with lower pre-program academic performance, which suggests that early language interventions are more effective for integration. In addition, we provide evidence that the language training program generated modest positive spillovers on native students. Finally, we identify mental health as a potential key channel through which the program enhanced educational outcomes, as alleviating language barriers improved concentration, reduced anxiety, and decreased bullying. Our findings underline the critical role of language training in improving refugee students' educational outcomes and mental well-being, and fostering social cohesion.

JEL Classification: I21, I31, J15, J24, Z13

Keywords: language, refugee integration, school outcomes, mental health

Corresponding author:
Semih Tumen
Department of Economics
TED University
Ziya Gökalp Caddesi No.48 06420
Kolej Çankaya
Ankara
Turkey

E-mail: semih.tumen@tedu.edu.tr

* We thank Catia Nicodemo, Giovanni Peri, Gaurav Khanna, Pinar Ozel, Serdar Polat, Metin Catar, Secil Balkir, the participants of the VII Workshop on Immigration, Health and Wellbeing in Verona, and seminar participants at the Ministry of National Education of the Republic of T̈urkiye and TED University for useful comments. The views expressed in this paper are of our own and do not necessarily reflect those of Amazon or other third parties.

1 Introduction

The integration of immigrants presents a significant challenge for many host countries, particularly in light of the recent increase in refugee populations.[1] This surge has intensified interest in evaluating the effectiveness of integration policies, as inadequate integration carries substantial costs for both immigrants and host societies. Among the various factors influencing integration, language proficiency is a key determinant with profound implications for economic, social, and cultural outcomes (Chiswick and Miller, 2015; Ginsburgh and Weber, 2020). Language proficiency serves as a fundamental enabler for refugees and immigrants, allowing them to navigate daily life, access essential services, and participate in a wide range of socioeconomic activities. It also empowers them to understand critical safety and security information, fostering both personal autonomy and community engagement. Beyond these benefits, language skills play a key role in facilitating immigrants' broader integration into society.

For children of migrant and refugee families, the acquisition of host-country language skills is especially critical. Proficiency in the local language significantly influences their educational attainment, social integration, and eventual participation in the labor market (Dustmann and Glitz, 2011). Limited language comprehension, on the other hand, might negatively affect their educational outcomes and overall well-being, with potential negative spillovers on native students. Recognizing these impacts, policymakers and key international stakeholders highlight the importance of language proficiency for the academic success and mental well-being of immigrant children (European Commission-EACEA-Eurydice, 2019; UNHCR, n.d.).[2]

In recent years, language training programs for refugee children have been introduced in many host countries, including Türkiye, Germany, and Sweden. However, evidence on the causal impact of these programs on the educational and broader outcomes of refugee children remains scarce. As Bahar et al. (2024) highlight in a recent survey: "We have no causal work documenting how these education language policies have affected refugee children's lives, but they are likely consequential." This gap in causal evidence is due to two main challenges. First, there is a shortage of micro-level data capturing educational outcomes of refugee students in a representative way. Second, in general, the implementation of these programs does not allow for the application of credible quasi-experimental designs needed for causal analysis. This paper addresses these challenges, providing the first causal evidence on the effectiveness of large-scale language training programs aiming to enhance the educational integration of refugee children.

We focus on Türkiye, which hosts one of the world's largest refugee populations due

[1] According to the UNHCR, the global forcibly displaced population reached a record high of 114 million by the end of 2023, representing a significant rise from previous years (UNHCR, 2024).

[2] The pivotal role that language plays extends beyond the case of refugees and has broader implications for integration of immigrants at all ages. Recognizing this, immigrant integration programs with a particular focus on language training are implemented with a large budget in many developed countries, including Canada, France, Germany, Norway, Israel, and Sweden (Pont-Grau et al., 2023).

to its cultural, religious, and geographical proximity to conflict-afflicted countries such as Syria, Iraq, Iran, and Afghanistan. Although the official figures suggest that the number of Syrian refugees in Türkiye is about 2.9 million by the end of 2024,[3] down from 3.8 million in 2022, the number of refugees from all nationalities (including the unregistered ones) is often argued to be about 5-6 million.[4] Given that almost 50 percent of all refugees in Türkiye are children of school age (Tumen, 2023), and considering that the root causes of these refugee waves are expected to persist for decades, it is not surprising that this massive influx of refugees has posed unparalleled challenges for the Turkish education system. This has necessitated the implementation of large-scale and multi-dimensional educational integration policies, rather than the smaller, more targeted approaches typically adopted in many other host countries. One such initiative is the language training program examined in this paper, which reached approximately half a million refugee children.

Using rich administrative micro-level data, we estimate the causal impact of this large-scale language training program on the academic performance and school attendance of Syrian refugee children. Our analysis focuses on a sample of about 50,000 refugee students in two Turkish provinces. The program was designed to improve the Turkish language reading, writing, speaking, and listening skills of refugee children, and, thereby, to promote both their educational integration and social cohesion in general. This initiative was part of a broader range of EU-funded refugee integration programs implemented in Türkiye (Tumen et al., 2023). Our identification strategy leverages the staggered introduction of the program across schools in a difference-in-differences framework. We test the robustness of our findings using the Callaway and Sant'Anna (2021) approach, which is particularly well-suited in our setting to address concerns related to the validity of the canonical two-way fixed effects (TWFE) approach in staggered rollout designs.

We find that the program significantly improved the Turkish language and Math scores of refugee students, with gains of 0.186 and 0.128 standard deviations, respectively, while reducing their school absences notably. Although there are no meaningful gender differences in the estimates, there is noteworthy heterogeneity along two important dimensions. First, the program effects are almost entirely driven by younger children, which suggests that early language interventions are more effective than later ones. Second, the improvements in test scores and school attachment mostly come from those with below-median pre-program academic performance. This means that the language training program was more effective for those who needed it the most. Furthermore, the impact persisted over the academic years following the intervention proving longer-term effectiveness of the program. We also document modest improvements in the academic performance of Turkish students, which provides suggestive evidence regarding positive spillovers on native students.

[3] For up to date figures, see https://data.unhcr.org/en/situations/syria/location/113.
[4] https://teyit.org/dosya/turkiyedeki-siginmaci-sayisi-veriler-ne-soyluyor.

After establishing the substantial impact of the program on the educational outcomes of refugee children, we turn attention to its potential effects on their mental well-being—a critical channel through which these educational gains may have been realized. Adapting to a new educational system in an unfamiliar language can be a source of significant stress and anxiety for these students, potentially hindering their ability to learn and fully engage in school activities. Therefore, by addressing language barriers, the program has the potential to improve students' mental well-being, which can in turn positively impact their academic performance.

Using data from a specialized survey conducted on a subsample of Syrian refugee students in Şanlıurfa, we estimate the impact of the language intervention on their mental health. The survey measured various mental health outcomes on a Likert scale, including concentration and focus, sleep disorders, eating disorders, conflicts with peers and siblings, feelings of unhappiness or depression, and willingness to attend school. We also construct an overall mental health score by aggregating all the sub-components.

We find that the overall mental health levels of refugee students improved significantly following the language training program. We further document that this improvement was mainly driven by gains in concentration/focus, sleep and eating disorders, and willingness to attend school. Importantly, we also find that the language training program reduced the refugee students' probability of involvement in a bullying incidence by 5.4 percentage points. These results suggest that alleviating language barriers not only directly enhances academic performance but also indirectly facilitates learning by reducing anxiety, improving self-confidence, and fostering a more supportive school environment. These findings highlight that mental health is an important channel through which the program contributed to the broader integration and educational success of refugee children.

Our paper contributes to several strands of the literature. First, several studies have documented the importance of language proficiency for adult immigrants' performance in the host-country labor market, and in particular the positive payoff in terms of labor market outcomes and earnings for adult immigrants and refugees (Chiswick and Miller, 1995; Dustmann and van Soest, 2001; Bleakley and Chin, 2004; Fasani et al., 2022; Schmid, 2023). More recently, studies have focused on establishing the causal impact of language training on labor market outcomes of adult immigrants (Lochmann et al., 2019; Arendt et al., 2024; Arendt and Bolvig, 2023; Foged et al., 2023; Heller and Mumma, 2023; Foged et al., 2024). We complement the findings of this literature by providing evidence on the impact of language training on children's educational outcomes and school integration, which are critical pre-cursors to labor market integration and success.

Secondly, a small but growing strand of the literature has focused on the assimilation of refugee children. For example, Boucher et al. (2021) examine the causal impact of interethnic exposure in an early childhood program setting on the social integration of refugee children in Türkiye. Alan et al. (2021) examine the impact of ethnic segregation of refugee children in Türkiye on peer violence and victimization on school grounds,

social integration and Turkish language ability. On the other hand, Hockel and Schilling (2022) investigate the academic effects of separate preparatory language classes for refugee children, finding that such segregation does not enhance their academic achievement. Although these studies provide insights about aspects of school integration of refugee children, none of them examine the causal impact of language training on educational outcomes of refugee and native children. Our paper is also related to the literature examining the effects of clustering immigrant and refugee children in classrooms on the educational outcomes of their native peers, with mixed findings—see, for example, Gould et al. (2004); Ohinata and van Ours (2013); Assaad et al. (2023). We contribute to this literature by presenting causal evidence on the positive spillover effects of language training for refugees on native children's educational outcomes.

Finally, our paper is also related to an emerging literature that examines the relationship between language proficiency and mental health. A few (non-economics) studies have explored this relationship. For example, Cavicchiolo et al. (2020) documented a positive correlation between language proficiency and mental health of immigrant children in Italy, controlling for individual, family, and group characteristics. Although broader benefits of language acquisition—such as enhanced consumption opportunities, social and civic inclusion, and human capital development—are well-recognized (e.g., Chiswick and Miller (2015)), systematic evidence on the causal impact of language training and acquisition on the psychological well-being of immigrant and refugee children remains limited. Addressing this gap, we provide evidence that the language training program, by mitigating language barriers, can improve both mental health and academic outcomes of refugee children, highlighting the multidimensional benefits of such interventions.

To the best of our knowledge, this is the first paper that estimates the causal impact of a large-scale language training program on three key outcomes: (*i*) the academic performance of refugee students, (*ii*) their mental health, and (*iii*) spillovers on the educational outcomes of native students. Our findings have important implications for educational policy targeting immigrant and refugee students. In particular, our results suggest that the impacts of language training programs not only lead to better educational outcomes and school attachment of refugee students, but also indirectly enhance academic performance by improving their mental well-being. By mitigating language barriers, these programs help reduce stress and foster a supportive environment that can positively influence learning. Our results highlight the critical role of early language education for immigrant and refugee children in fostering better integration, reducing bullying and conflict incidence, and improving concentration, eating and sleeping habits, and school engagement. Hence, investing in early language training of refugee and immigrant children should be a priority for host countries.

The rest of the paper is structured as follows. Section 2 outlines the language training program and provides an overview of the institutional setting. Section 3 provides a description of our data set and explains our empirical strategy. Section 4 presents our main

results and robustness checks. Section 5 explores mental health as a potential channel. Finally, Section 6 offers some concluding remarks.

2 Institutional Background

2.1 Education of Syrian refugees in Türkiye

The conflict in Syria began in early 2011 and rapidly escalated into a full-scale civil war, triggering a massive displacement and leading to one of today's most pressing humanitarian crises (Tumen, 2023; World Bank, 2023). While millions sought refuge, primarily in neighboring countries such as Türkiye, Lebanon, and Jordan, the displacement wave extended across a large geographical area, including regions in Europe and MENA. Türkiye, in particular, became the host to the largest population of Syrian refugees, reaching a peak of approximately 3.8 million by June 2022.[5] The prospect of a swift return to Syria has significantly diminished, owing to the persistent political instability, ongoing conflict, proxy wars, and the emergence of extremist groups. In this complex context, Türkiye has faced the formidable challenge of providing education to a vast and diverse refugee population. Below we elaborate on the educational initiatives, challenges, and interventions that have been undertaken to address the unique needs of Syrian refugee students in Türkiye.

The Syrian population in Türkiye is, on average, younger than the native population, with nearly half of Syrians under the age of 18 (Tumen, 2018). Syrian children have disproportionately borne the adverse effects stemming from violent conflict, forced displacement, and persistent instability. As of the beginning of the 2022-23 school year, approximately 1.4 million school-age (5-17) refugee children were present in Türkiye. These children face numerous challenges in their schooling, including lower test scores, reduced enrollment, and higher absenteeism compared to their native peers (Tumen et al., 2023). These disparities are frequently attributed to language barriers, identified as a primary challenge for the educational integration of refugee children (UNHCR, 2019). Therefore, a concerted focus on removing these barriers could substantially improve the long-term effectiveness of the overall refugee integration policy framework.

The approach to educational integration of Syrian children in Türkiye has gone through various stages. During the early stages of the crisis, e.g., between 2011-2016, the policy priorities were shaped around the humanitarian needs. Education of refugees was handled as a service separately provided outside of the Turkish education system.[6] After 2016, the EU Facility for Refugees in Türkiye (FRIT—a 6-billion EUR fund) was launched and a sizable portion of the FRIT funds were used to prioritize the full integration of refugees into the Turkish public education system. Accordingly, various school integration programs were implemented through the PIKTES (Promoting Integration of Syrian Kids into the Turkish Education System) project, which was administered by MoNE and financed

[5] See https://data.unhcr.org/en/situations/syria/location/113.

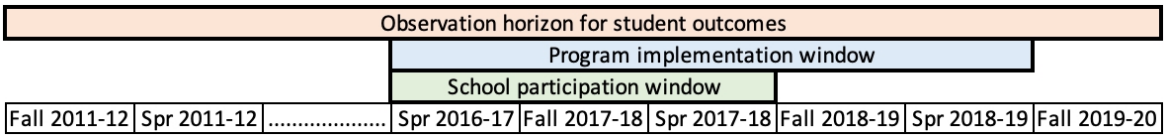
[6] See Boucher et al. (2021) for a more detailed chronology of these education services.

by the FRIT funds. The main ones were the back-up training, catch-up training, Turkish language training, and teacher training programs. The back-up and catch-up training programs aimed at providing academic support to enrolled and out-of-school refugee students, respectively. The language training program aimed at improving Turkish language skills of refugee students. This paper focuses on the language training program, which is described in the next subsection.

2.2 Language training program

The language training program aimed at improving the Turkish language skills of Syrian refugee students. Syrian children who were accustomed to the Arabic alphabet and language—especially the ones who were previously educated in Arabic—faced significant language barriers in Turkish schools.[7] The stark differences between the Arabic and Turkish languages in terms of linguistic structures, phonetics, and alphabets make Turkish language acquisition particularly difficult for Syrian children. The language training program was therefore designed to improve the Turkish reading, writing, speaking, and listening skills of refugee children.

Figure 1: Timing of program implementation and school participation



Notes: The figure describes the observation horizon for student data, program implementation window, and school participation window. All students in our sample were observed in both pre-treatment and post-treatment semesters so that we can implement our DiD estimation strategy with student fixed effects.

Figure 1 visualizes the data horizon and treatment timeline in our empirical design. The program was implemented between January 2017 and June 2019 (over 5 semesters) in 26 provinces. The target population was Syrian children aged 6-17, from grades 1 to 12. The program was officially titled the "Turkish Language Training Course for Foreigners," and was overseen by the Directorate General of Lifelong Learning (DGLL) under the Ministry of National Education of the Republic of Türkiye (MoNE). Teaching materials—consisting of a "Turkish for Foreign Students" book for in-class instruction and a set of short story books—were developed by the DGLL in coordination with UNICEF.

The language training program was delivered in 1,837 public schools and 207 Temporary Education Centers (TECs) in 26 provinces by 4,474 specially recruited Turkish language teachers.[8] Over its course, the program reached 449,634 Syrian students.

[7] In the Syrian culture, it is common to expose children to certain religion-related Arabic words such as "Allah," "Muhammad," and holy phrases such as Takbir, Salawat, and Basmala starting from early ages. Many children also attend Quran courses at pre-school age.

[8] These teachers were contract teachers employed within the PIKTES project—not the regular teachers in public schools. Before the program, the teachers received a short orientation training covering the details of the program and logistical arrangements.

Instruction was differentiated into two age groups: 6-12 and 13-17 years. For both groups, instruction was structured into three proficiency levels: beginner, intermediate, or advanced—assessed by teachers before the program. The duration of the program was 20 hours, complementing the standard Turkish language classes in the school curriculum. Depending on the physical capacity of schools, the number of enrolled refugee students in a given school, and the school's education mode (e.g., one shift versus two shifts), classes were conducted after the school hours either within the school building or an available TEC in the neighborhood.[9] The program was delivered once in a given semester, in the form of 2-hour blocks, held weekly over 10 weeks throughout the semester. Syrian students could attend multiple cycles, progressing through the levels (i.e., starting from a lower proficiency level and progressing toward more advanced levels), but could not repeat the same proficiency level.

The program was introduced to schools in a staggered way over several semesters.[10] The first cohort of schools began implementation in the Spring of 2016-17, and by the end of the Spring 2017-18 semester, all participating schools had delivered the program at least once. Participation and timing of program entry were decided by school administrators, who were given a three-semester window—between Spring 2016-17 and Spring 2017-18—to opt into the program.

3 Data and Research Design

3.1 Data description

We use micro-level administrative data regarding Syrian refugee students' outcomes from Gaziantep and Şanlıurfa—the largest Turkish provinces in the Syrian border region. These provinces host a total of about 870,000 Syrian refugees, which corresponds to an approximately 0.2 refugee-to-population ratio.[11] The location and demographic composition of these two provinces make them pivotal in the implementation of educational integration programs for refugee children.

Our dataset covers student outcomes between the Fall 2011-12 and Fall 2019-20 semesters. It includes 46,761 refugee and 804,654 Turkish students across 1,426 schools (Table 1). As mentioned above, the language training program was rolled-out gradually across schools during this period. Within our sample, the initial cohort of schools offering the training covered 32% of refugee students, while the second cohort, started in Fall 2017-18, covered an additional 36%, bringing the total coverage to 68% of refugee students by the end of the first year. By the end of the Spring 2017-18 semester, 88% of the refugee students were attending schools that had implemented the program. This means that 12% of the

[9] TECs served as venues to accommodate the overflow of refugee children from participating schools.

[10] Although participation was mandated for all schools, some were unable to implement the program due to resource constraints and planning challenges.

[11] The data source for province-level refugee concentration is the annual reports of the Ministry of Interior of the Republic of Türkiye, Presidency of Migration Management.

students in our sample attended schools that never participated in the program (they were never treated).

While our data set indicates which schools implemented the language training program and the timing of its introduction, it does not specify which refugee students participated in the program. Our analysis focuses on three key outcomes—school absences, academic achievement, and mental health—allowing us to evaluate the program's impact on both the educational and psychological well-being of refugee students.

School absences. The data set contains information on the days of absence from school at the semester level, which serves as our main outcome of interest. School absenteeism is an important student outcome that proxies various aspects of human capital development. First, the days of absence variable is directly used to measure learning losses due to reduced school attendance. For example, absenteeism has been widely used by many researchers and policy institutions to quantify learning losses that emerged during the Covid-19 pandemic. Second, absenteeism serves as a proxy for non-cognitive skills. Several studies in the literature document that absenteeism is negatively correlated with the "Big Five" personality traits that constitute the core of non-cognitive skills (Lounsbury et al., 2004). Finally, school absenteeism is strongly associated with various antisocial and/or risky behaviors that emerge later in life.[12]

Student achievement. To measure student achievement, we rely on student grades that we observe in the administrative records. Specifically, we focus on grades in two core subjects: Turkish language and Math. Note that grades are reported on a three-category scale (i.e., from 1 to 3) during the first three years of primary education, while a 0-100 grading scale is used from year 4 to year 12. For consistency, in our analysis of grades, we focus on grades from year 4 onward. These grades represent end-of-semester averages derived from a variety of assessment methods, such as written exams, oral exams, quizzes, projects, and homework. Moreover, the data set includes information on the country of origin of the student allowing us to differentiate between Turkish and refugee students.

Student mental health. We use a special survey designed to measure the mental health outcomes of refugee students. The survey was implemented in selected schools with high refugee concentration in Şanlıurfa. It followed the same families in three consecutive waves that coincide with the language training program's time frame. This overlap allows us to apply a similar empirical strategy to estimate the impact of the language training program on refugee students' mental health outcomes. The survey sample includes 3,246 students across 66 schools. Note that while the sample may not be fully representative of the refugee student population, tracking the same students over time enables us to use student fixed effects to account for unobserved heterogeneity. The survey collected information from parents as respondents, using a 5-point Likert scale (1: lowest score, 5: highest

[12] See Gubbels et al. (2019) for a comprehensive meta-analytic review of the related literature.

Table 1: Summary statistics

	Syrian	Turkish
Turkish language score (pre-treatment)	53.65	71.45
	(21.54)	(23.96)
Math score (pre-treatment)	49.33	69.16
	(23.65)	(26.05)
School absences (pre-treatment)	10.24	4.71
	(7.31)	(3.89)
Age	9.91	12.86
	(2.76)	(2.92)
Grade/year	5.45	7.11
	(2.35)	(2.71)
Number of siblings	3.28	2.34
	(2.67)	(1.99)
$\mathbb{P}(\text{Male} = 1)$	0.49	0.50
$\mathbb{P}(\text{Primary school} = 1)$	0.69	0.38
$\mathbb{P}(\text{Gaziantep} = 1)$	0.48	0.44
$\mathbb{P}(\text{Preschool attendance} = 1)$	0.11	0.59
$\mathbb{P}(\text{Origin is Aleppo} = 1)$	0.59	0
$\mathbb{P}(\text{Absenteeism} \geq 10 \text{ days})$ (pre-treatment)	0.41	0.09
$\mathbb{P}(\text{Absenteeism} \geq 20 \text{ days})$ (pre-treatment)	0.19	0.04
# of observations	46,761	804,654

Notes: The grading scale for the Turkish language and Math assessments is 0-100. Standard deviations are reported in brackets.

score) to evaluate seven key dimensions of mental health: (1) ability to concentrate, (2) sleep disorders, (3) eating disorders, (4) conflicts with friends/siblings, (5) feelings of unhappiness/depression, (6) willingness to attend school, and (7) involvement in bullying incidents in school.

Summary statistics. Table 1 presents summary statistics at the student level, highlighting key differences between refugee and Turkish children. Refugee children are younger, on average. This is partly due to higher dropout rates among refugee students, resulting in a smaller proportion of them advancing to higher grades. Refugee students (pre-treatment) also have lower academic achievement, and have approximately double the rate of absenteeism compared to Turkish children. About 70% of the Syrian sample is enrolled in primary school, compared to 40% for the Turkish sample. Syrian children are also less likely to have attended pre-school, which could contribute to differences in school readiness.

3.2 Empirical specification

To estimate the causal effect of language training on refugee students' outcomes, we exploit the staggered introduction of the language training program across schools. Accordingly,

we estimate difference-in-differences (DiD) specifications that take the following form:

$$y_{i,s,t,g,\tau} = \beta T_{s,t} + \theta_i + \gamma_g + (\sigma_t \times \delta_\tau) + \epsilon_{i,s,t,g,\tau}, \tag{1}$$

where i, s, g, τ, and t denote students, schools, grade/years, towns, and semesters, respectively; y is the outcome variable; $T_{s,t}$ is a dummy variable taking the value of 1 for semesters t following the introduction of the language training program in school s (treatment is an absorbing state); θ_i are student fixed effects, while γ_g are grade/year fixed effects; and ϵ is an error term. We also include semester-by-town fixed effects $(\sigma_t \times \delta_\tau)$ to capture time-varying town-level factors (e.g., changing local policies or conditions) that might affect the outcomes. Standard errors are clustered at school level throughout our analysis, as the language training program was implemented at the school level.

The coefficient of interest is β, which identifies the intent-to-treat (ITT) effect of offering the language training program at the school level. This is because, while we know the semester at which the program was offered in each school, we do not have individual attendance records. Consequently, our estimates capture the average effect of offering the program at the school level, rather than the likely larger treatment-on-the-treated effect, which measures the impact on students who directly participated in the program.

Our identification strategy relies on the assumptions that (i) in the absence of treatment, the outcomes of students in treated and untreated schools would have followed parallel trends and (ii) there is no heterogeneity in treatment effects across students and time. In the following sections, we provide empirical evidence in support of these assumptions. In addition, we perform robustness checks and employ alternative estimation methods to address potential threats that could challenge the validity of our identification strategy.

3.3 Identification

We next discuss various identification issues relevant to staggered DiD design and the steps taken to address them.

Balance tests. The validity of our staggered DiD design hinges on the assumption that the timing of the language training rollout across schools is exogenous to the outcomes of interest. This assumption could be violated if, for instance, language training was adopted earlier by schools with refugee children that were less proficient in the Turkish language. To address this concern, we perform a series of balance tests, the results of which are reported in Tables A1, A2 and A3.[13] First, Table A1 presents student-level balance

[13]The balance tests are based on two types of regressions: student level (Table A1) and school level (Tables A2 and A3). In these tables, each row corresponds to a different regression, with the dependent variable being the pre-treatment characteristic indicated. The samples are restricted to the pre-program period. In student-level analysis, controls include a dummy variable indicating the treatment status, school fixed effects, town-semester fixed effects, and grade/year fixed effects. In school-level analysis, controls include treatment status, town fixed

tests, comparing the pre-treatment characteristics of the treated versus never-treated students in a cross-sectional analysis. The results indicate no systematic differences in pre-treatment outcomes or observable characteristics. Second, at the school level, Table A2 shows no evidence that schools that received earlier treatment differed systematically from other schools in terms of pre-treatment characteristics. Finally, Table A3 examines the distribution of students across schools based on the timing of program adoption. These results collectively provide reassurance that our empirical design is not confounded by non-random program implementation.

Parallel trends. To test for parallel trends, we estimate dynamic event study models that account for the staggered timing of treatment adoption across schools. The analysis window spans six semesters, with the semesters before treatment adoption labeled as -3, -2, and -1, and those after as 0, 1, and 2. Then, for each semester k, we define a treatment variable $1[\text{Semester Since Treatment} = k]$, which takes the value of 1 for treated students and 0 for never-treated students. We then estimate the following event study specification:

$$y_{i,s,t,g,\tau} = \sum_{k \neq -1} \delta_k \cdot 1[\text{Semester Since Treatment} = k] + \theta_i + \gamma_g + (\sigma_t \times \delta_\tau) + \epsilon_{i,s,t,g,\tau}, \quad (2)$$

where the coefficients δ_k capture the dynamic treatment effects relative to the normalized pre-treatment period ($t = -1$). The other terms in the equation are defined as in the previous specification. The results are reported separately for students ever-treated, treated once, and treated twice in Figures 2 and A1-A2 in the Appendix. These figures provide evidence supporting the parallel trends assumption for all three outcomes.

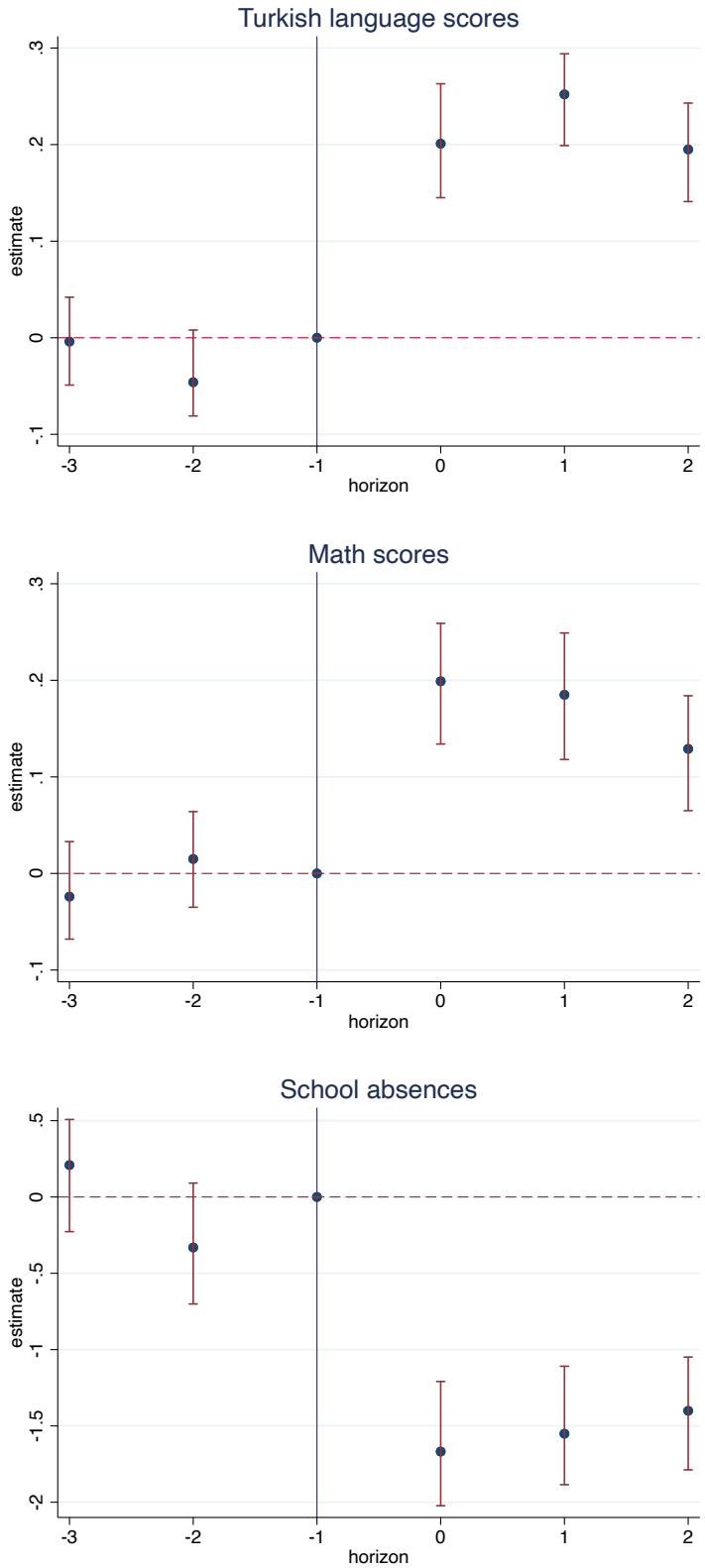
Heterogeneous treatment effects. Beyond the concern of random assignment of treatment, recent literature suggests that staggered DiD designs may not provide valid estimates of the causal parameters of interest due to treatment effect heterogeneity (e.g., de Chaisemartin and D'Haultfoeuille (2023); Roth et al. (2023); Wing et al. (2023)). To address this issue, in Section 4.5, we present results employing a heterogeneity-robust estimator (Callaway and Sant'Anna, 2021). Our results are robust and generally indicate larger treatment effects compared to those obtained using OLS.

4 Impact of language training on test scores and school absences.

We use the staggered DiD specification in Equation 1 to estimate the impact of the language training program on our outcomes of interest. In this section, we focus on our three education outcomes that we obtained from administrative records: Turkish language

effects, and town-semester fixed effects. Note that the main difference between Tables A2 and A3 is that, in the former, the treatment variable indicates treated and control schools, while in the latter it indicates treatment timing.

Figure 2: Event study results: Ever-treated & Never-treated



Notes: Timing of treatment is centered around 0 for different enrollment cohorts. Event study analysis described in Section 3.3 compares students in the ever-treated and never-treated schools.

scores, Math scores, and days of school absences per semester. We begin by presenting the baseline estimates for the entire sample, followed by a heterogeneity analysis by gender, age, and previous academic success. We also test the persistence of the program's effects by examining the one- and two-period-ahead outcomes. In addition, we estimate the impact of the program using a categorical treatment specification, which focuses on treatment intensity rather than defining the treatment as a binary variable. Finally, in Section 4.5, we conduct robustness checks to assess the validity of our staggered DiD design using recent testing procedures developed in the literature.

4.1 Main results

In Table 2, we present estimates of the effect of the language training program on the three outcomes of interest. We begin by presenting a simple specification that includes school and semester fixed effects in Column 1. In Column 2, we add grade/year fixed effects and account more flexibly for time effects by including town-specific semester fixed effects. In Columns 3-5, we replace school by student fixed effects and add gradually grade/year and town-semester fixed effects. The estimated effects are consistently positive for academic performance and negative for absenteeism across the different specifications. The point estimates tend to decrease in magnitude as we add student fixed effects and town/semester fixed effects, but remain statistically significant at the 1% level for Turkish language scores and absenteeism, and at the 10% level for Math scores.

The estimated impacts on academic outcomes and school attachment of refugee children are sizable. For example, considering the estimates in Column 5, we find that language training improves Turkish language scores by 0.186 standard deviations, and Math scores by 0.128 standard deviations. These results suggest that the language training program not only enhances language proficiency but also facilitates learning in other subjects, such as Math.

The effect on school absenteeism is equally significant, amounting to about 1.9 fewer days missed per semester, which constitutes about 18.5% of the average absenteeism rate among refugee children. To put this into perspective, when considering the average absenteeism gap of 5.5 days between refugee and Turkish children, the training succeeds in closing about one-third of this gap.

4.2 Heterogeneity

We next examine heterogeneity of treatment effects based on gender, age, and previous academic achievement.

Gender. Table 2 presents results separately by gender. Overall, we find no substantial gender differences in the impact of the training. For the academic outcomes, girls seem to be benefiting slightly more in both Turkish language and in Math. On the other hand, boys seem to experience a larger reduction in absenteeism.

Table 2: Overall effect: Impact on outcomes of Syrian refugee children

	[1]	[2]	[3]	[4]	[5]
Turkish language scores					
All sample	0.286***	0.245***	0.222***	0.219***	0.186***
	(0.037)	(0.044)	(0.041)	(0.040)	(0.046)
Males	0.279***	0.237***	0.211***	0.213***	0.179***
	(0.044)	(0.051)	(0.047)	(0.047)	(0.051)
Females	0.294***	0.251***	0.226***	0.222***	0.190***
	(0.046)	(0.052)	(0.048)	(0.047)	(0.053)
Math scores					
All sample	0.216***	0.165**	0.208***	0.176***	0.128*
	(0.048)	(0.062)	(0.051)	(0.052)	(0.063)
Males	0.209***	0.159**	0.201***	0.170**	0.109
	(0.061)	(0.066)	(0.063)	(0.065)	(0.070)
Females	0.229***	0.176**	0.224***	0.184**	0.144*
	(0.063)	(0.068)	(0.065)	(0.066)	(0.071)
School absences					
All sample	-3.065***	-2.358***	-2.196***	-2.155***	-1.916***
	(0.316)	(0.319)	(0.342)	(0.344)	(0.366)
Males	-3.148***	-2.512***	-2.201***	-2.176***	-2.007***
	(0.348)	(0.354)	(0.378)	(0.372)	(0.391)
Females	-2.912***	-2.211***	-2.099***	-2.085***	-1.748***
	(0.356)	(0.362)	(0.384)	(0.381)	(0.401)
School fixed effects	Yes	Yes	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes
Town-semester fixed effects	No	Yes	No	No	Yes
Grade/year fixed effects	No	Yes	No	Yes	Yes
Semester fixed effects	Yes	No	Yes	Yes	No
# of observations			46,761		
(# of males)			22,945		
(# of females)			23,816		

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3: Heterogeneity with respect to starting age

	[9-11]	[12-14]	[15-17]
Turkish language scores			
Program effect	0.226***	0.125*	0.025
	(0.048)	(0.055)	(0.066)
Math scores			
Program effect	0.185*	0.121	0.019
	(0.071)	(0.073)	(0.095)
School absences			
Program effect	-2.218***	-1.676**	-1.089*
	(0.408)	(0.479)	(0.539)
Student fixed effects	Yes	Yes	Yes
Town-semester fixed effects	Yes	Yes	Yes
Grade/year fixed effects	Yes	Yes	Yes
# of observations	24,943	12,041	9,777

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Age. Table 3 presents results separately for 3 age groups: 9-11, 12-14, and 15-17. The impact of language training is larger at younger ages, especially for academic outcomes. Specifically, the effect is highest for the 9-11 group (0.225 standard deviations for Turkish language and 0.185 standard deviations for Math), is about half the size for the 12-14 group, and is very small and statistically insignificant for older children. For absenteeism, the effect in the 9-11 age groups is double that in the 15-17 age group. These findings highlight the importance of early interventions in improving both academic performance and school attendance.

Previous achievement. Table 4 presents results separately by pre-program achievement level for the respective outcome—defined as below and above median. The impact of the language training appears to be driven largely by students who were below the median in terms of academic performance before the program. This suggests that the program has been effective in closing the pre-existing achievement gaps among refugee students.

4.3 Turkish children

We next turn attention to the educational outcomes of Turkish children. Note that the program did not offer additional instruction to these children. However, we might expect some spillover benefits given the improvement observed in the academic outcomes of Syrian children.

Results are presented in Table 5 and suggest a modest improvement in academic outcomes of Turkish children, though the effects are small (0.041 sd) and statistically insignificant for Turkish language in our preferred specification in Column 5. The impact

Table 4: Heterogeneity with respect to previous academic performance

	[1]	[2]	[3]	[4]	[5]
Turkish language scores					
Program effect (below median)	0.385***	0.366***	0.354***	0.341***	0.275***
	(0.054)	(0.062)	(0.058)	(0.057)	(0.066)
Program effect (above median)	0.153**	0.136*	0.116	0.125*	0.095
	(0.057)	(0.064)	(0.061)	(0.060)	(0.070)
Math scores					
Program effect (below median)	0.244***	0.191*	0.232***	0.201**	0.171*
	(0.067)	(0.080)	(0.069)	(0.070)	(0.081)
Program effect (above median)	0.154*	0.114	0.141*	0.124	0.089
	(0.068)	(0.082)	(0.071)	(0.071)	(0.080)
School absences					
Program effect (below median)	-3.942***	-2.944***	-2.845***	-2.751***	-2.591***
	(0.601)	(0.604)	(0.652)	(0.658)	(0.688)
Program effect (above median)	-2.234**	-1.599*	-1.412	-1.376	-1.254
	(0.703)	(0.721)	(0.766)	(0.774)	(0.808)
School fixed effects	Yes	Yes	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes
Town-semester fixed effects	No	Yes	No	No	Yes
Grade/year fixed effects	No	Yes	No	Yes	Yes
Semester fixed effects	Yes	No	Yes	Yes	No
# of observations			46,761		

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

on Math is slightly more substantial at 0.066 sd and is statistically significant at the 10% level. The impact on absenteeism is also small (-0.3 days) and not statistically significant.

These results suggest that addressing the language barriers faced by refugee students can also have positive effects on the academic performance of native students, particularly in quantitative subjects. Although we are not able to pin down the exact mechanisms driving this result, two plausible explanations include (1) improved in-classroom interactions (including knowledge transfer) among peers, and (2) increased teacher effectiveness due to a more inclusive and integrated classroom environment.

4.4 Persistence

We next examine the persistence of the program's impacts over time. As we describe in Section 3, participating schools implemented the program in a staggered manner between Spring 2016-17 and Spring 2017-18, with the program ending by June 2019 (Figure 1). Our dataset covers student outcomes between Fall 2011-12 and Fall 2019-20. Note that we cannot observe student-level program participation. In our persistence analysis, we assume that students who were present in the school for three consecutive semesters after the start of the program in that school "complete" the program. Accordingly, we define the period t as the last semester a student received the training, while $t+1$ and $t+2$ are defined as one and two semesters after, respectively.[14] The results presented in Table 6 indicate that the effects remain substantial up to three semesters after its conclusion. For Math scores, we even find a tendency for improvement over time. For absenteeism, we observe a gradual decline on the impact. While the impact on absenteeism gradually diminishes, it remains sizable (1.5 fewer days of absence) and statistically significant up to three semesters after undertaking the training. These findings suggest that the effects of the language training are not only immediate but also endure over time, particularly in terms of academic performance.

4.5 Robustness

To test the robustness of our baseline estimates, we use the estimator proposed by Callaway and Sant'Anna (2021), as it is particularly suited for settings featuring absorbing treatments, a mix of not-yet-treated and never-treated controls, and potentially serially correlated outcomes over time due to learning-curve type of effects.[15]

[14] Note that this is a conservative assumption. It is likely that many refugee students attended fewer than three sessions and many of them did not attend at all. Therefore, our results in this section likely represent lower-bound estimates of persistence effects.

[15] Several approaches have been developed to address potential biases due to heterogeneous treatment effects in staggered DiD settings, with the choice of method depending on the nature of the treatment and data structure. For instance, Sun and Abraham (2021), de Chaisemartin and D'Haultfoeuille (2020), and Borusyak et al. (2024) are alternative methods to be used when the comparison group is the never-treated ones, when the treatment is transient and not-yet-switchers are used as controls, and when the time horizon is long and outcome variables are not serially correlated over time, respectively. See de Chaisemartin and D'Haultfoeuille (2023) for a detailed survey and comparison of these methods.

Table 5: Impact on Turkish students

	[1]	[2]	[3]	[4]	[5]
Turkish language scores					
All sample	0.116***	0.098***	0.107***	0.093***	0.041
	(0.019)	(0.021)	(0.021)	(0.022)	(0.025)
Males	0.122***	0.104***	0.111***	0.099**	0.045
	(0.024)	(0.027)	(0.026)	(0.028)	(0.031)
Females	0.109***	0.091**	0.098**	0.087*	0.039
	(0.026)	(0.029)	(0.028)	(0.034)	(0.035)
Math scores					
All sample	0.136***	0.119***	0.128***	0.117***	0.066*
	(0.021)	(0.026)	(0.022)	(0.026)	(0.032)
Males	0.129***	0.115***	0.125***	0.110***	0.061
	(0.022)	(0.026)	(0.023)	(0.025)	(0.035)
Females	0.141***	0.121***	0.135***	0.129***	0.078*
	(0.024)	(0.031)	(0.025)	(0.028)	(0.037)
School absences					
All sample	-0.875***	-0.713**	-0.842***	-0.711**	-0.301
	(0.119)	(0.271)	(0.211)	(0.278)	(0.309)
Males	-0.946***	-0.758**	-0.921***	-0.883**	-0.396
	(0.145)	(0.278)	(0.234)	(0.288)	(0.345)
Females	-0.766***	-0.643*	-0.732**	-0.639*	-0.245
	(0.151)	(0.292)	(0.266)	(0.299)	(0.387)
School fixed effects	Yes	Yes	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes
Town-semester fixed effects	No	Yes	No	No	Yes
Grade/year fixed effects	No	Yes	No	Yes	Yes
Semester fixed effects	Yes	No	Yes	Yes	No
# of observations			804,654		
(# of males)			401,961		
(# of females)			402,693		

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 6: Dynamic effects

	$[t]$	$[t+1]$	$[t+2]$
Turkish language scores			
Program effect	0.186***	0.191***	0.176**
	(0.046)	(0.048)	(0.051)
Math scores			
Program effect	0.128*	0.145*	0.161*
	(0.063)	(0.068)	(0.071)
School absences			
Program effect	-1.916***	-1.699***	-1.421**
	(0.366)	(0.394)	(0.407)
Student fixed effects	Yes	Yes	Yes
Town-semester fixed effects	Yes	Yes	Yes
Grade/year fixed effects	Yes	Yes	Yes
#of observations	46,761	44,962	42,355

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Figure A3 presents two sets of estimates where the control group consists of both not-yet-treated and never-treated students. The results are similar to our baseline estimates with two nuances: (i) the magnitudes of the estimates are higher for both Turkish language and Math scores, and (ii) standard errors are larger by 15-20%, though this does not affect the statistical significance of the results.

As a complementary exercise, we follow the approach adopted by studies such as Dobkin et al. (2018) and Fadlon and Nielsen (2019), and the best practices suggested by Miller (2023), and repeat our analysis for an alternative restricted sample. Specifically, we focus only on refugee students that have received the language training at some point in our sample period, and identify the treatment effect from the timing of treatment. In other words, we construct counterfactuals for treated students using students that received the same training in subsequent semesters, excluding the never-treated students from our analysis. This approach addresses the possibility that since the treatment is assigned at school level, the expectations of the students in ever-treated and never-treated schools might have evolved differently, which could affect our results. We present these additional estimates in Table A4. The results are almost identical to our baseline estimates, which allays concerns about differential expectations and anticipation. This result also implies that the never-treated sample is not systematically different from the ever-treated one.

5 Improved mental health as a channel

In this section, we examine the role of mental health as a potential key channel through which the language training program influences the academic outcomes of refugee chil-

dren. Our hypothesis is that language barriers are a major source of mental stress for refugee students, and that the language training program, by alleviating these barriers, can improve both mental health and academic outcomes.

To investigate this channel, we leverage a unique survey that was designed to capture the levels of mental tension experienced by refugee children in educational environments and the behavioral implications of such stress. The survey sample includes 3,246 refugee students across 66 schools in Şanlıurfa. We refer to this sample as the "mental health sample." Although participation in the survey was voluntary, and we do not claim that the sample is fully representative, this data offers rare and valuable insights for both academic research and policy-making in this context.

The survey was conducted with the parents of refugee students, who responded to a set of questions across three waves, creating a panel dataset. The time frame of these survey waves spans both the time period before and after the language training program, which allows us to implement a DiD strategy, similar to our baseline specification, to estimate the impact of language training program on the mental health outcomes of refugee students. The survey included 6 mental health questions with Likert-scale responses (1: lowest - 5: highest) assessing the following mental health dimensions: (1) concentration in educational and daily activities, (2) sleep disorders, (3) eating disorders, (4) conflict with friends and siblings, (5) feelings of unhappiness/depression, and (6) willingness to attend school.[16] We construct an "overall mental health index," by summing the responses to these 6 mental health questions. In addition, the survey includes another question asking whether the student has been involved in bullying, without specifying whether the child was a bully or a victim.

We first verify whether our baseline results are consistent within the mental health sample, specifically examining the impact of the language training program on the academic outcomes of these students. Table A6 in the Appendix shows that the program did improve the Turkish language grades by 0.208 sd and reduced absences by 1.7 days. Although these effects are somewhat smaller than those observed in the full sample, and are estimated with less precision (significant at the 10% level), they nonetheless suggest that the program had a positive impact on academic performance for students in the mental health sample.

We next turn attention to the analysis of mental health outcomes. Table 7 presents results from our baseline DiD specification (Equation 1), using the overall mental health index. We also examine the impact on the various sub-components of the index, as well as on bullying incidence. The results reveal that there are significant mental health benefits associated with the language training program. Specifically, we find that the mental health index increases by 0.311—or 0.067 standard deviations. The analysis of the sub-components reveals that this improvement was mainly driven by improvements

[16] Willingness to attend school may not directly correspond to actual school attendance, but it can be directly observed by parents.

Table 7: Mental health effects

	Program effect	Mean (baseline)	St.dev. (baseline)
Concentration	0.149*	3.231	0.488
	(0.073)		
Sleep disorder	0.186*	3.457	0.512
	(0.081)		
Eating disorder	0.201**	3.675	0.526
	(0.076)		
Conflict w/ friends/siblings	0.112	3.112	0.495
	(0.081)		
Unhappy/depressed	0.094	3.956	0.582
	(0.079)		
Willingness to attend school	0.245***	2.956	0.465
	(0.069)		
Overall mental health	0.311***	22.543	4.623
	(0.091)		
Bullying incidence	-0.054**	0.194	0.185
	(0.021)		
#of observations		3,246	

Notes: The mental health questionnaire was completed by parents based on their observations of their children's behavior and well-being. Responses are measured on a 5-category Likert scale (1:lowest score-5:highest score). The overall mental health score is obtained by summing the scores of the six questions, resulting in a 6–30 scale. The bullying incidence is described by a dummy variable. The regressions control for individual fixed effects, town-semester fixed effects, and grade/year fixed effects. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

in concentration/focus, sleep disorders, eating disorders, and willingness to attend school. These dimensions are critical, as they directly influence students' ability to engage in learning and adapt to the school environment. We also find a significant reduction in bullying incidence of 5.4 percentage points, representing a 28% decline from the average baseline rate. This suggests that the program not only improved individual well-being but also contributed to fostering a more cohesive and supportive school environment.

Our interpretation of these findings is that the language training program had both direct and indirect effects on the educational outcomes of refugee students, with mental well-being serving as a key mediating factor. The direct impact stems from the improvement in Turkish language skills, which enabled students to more effectively communicate with their peers and teachers, thereby enhancing their ability to learn. Indirectly, better language skills enhanced students' confidence and overall mental well-being, promoting a sense of comfort in the mixed educational and social environment. This in turn contributed to improved academic performance and higher school attendance. Note that although a feedback loop may exist, where improved academic success could further improve mental health, we cannot fully disentangle the dynamics of this relationship. Overall, these results highlight the broader positive impact of language training on the well-being

of refugee students, extending beyond its direct impact on academic achievements.

6 Conclusion

Refugee and immigrant children face many hurdles in accessing, continuing, and excelling in education, which are typically exacerbated by limited proficiency in the host-country's language. This paper presents the first causal evidence on the effectiveness of language training programs aiming to enhance school integration of refugee children. Using rich administrative micro-level data on Syrian refugee students in Türkiye, we evaluate the impact of a large-scale language training program designed to improve their educational integration. Our analysis employs a staggered DiD design, leveraging the phased rollout of the program across schools.

The results show that language training substantially improves Turkish language proficiency and supports learning in technical subjects such as Math. The impact on school absenteeism is also positive, amounting to about 1.9 fewer days missed, which is equivalent to closing about one-third of the gap in absenteeism between refugee and native children. Importantly, there are no substantial gender differences in the impact of the training, but there are important age differences, suggesting that language interventions targeting refugee students are much more effective when targeting younger refugee students. Moreover, although native students did not receive any additional instruction, we document positive spillover effects for native students, particularly modest improvements in Math scores.

A key contribution of this study is providing evidence for the potential role of mental well-being as a mechanism linking language proficiency to improved educational outcomes. More specifically, we find that refugee students' mental well-being improved significantly in response to the language intervention and the improvement was mainly driven by better concentration/focus, reduction in sleep and eating disorders, and increased willingness to attend school. These findings suggest that better host-country language skills reduce refugee students' anxiety/stress, improve their self-confidence, and foster social cohesion. Moreover, we document a 5.4 percentage point reduction in bullying involvement—a 28% decline compared to the baseline incidence rate—highlighting the intervention's broader benefits for the school environment.

The paper highlights the positive role played by language training programs and its wide benefits for refugee children and positive spillover for native students. These findings underscore the need for host countries to invest in early language training of refugee and immigrant children to ensure better integration and social cohesion.

References

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021): "Building social cohesion in ethnically mixed schools: An intervention on perspective taking," *Quarterly Journal of Economics*, 136, 2147–2194.

Arendt, J. N. and I. Bolvig (2023): "Trade-offs between work-first and language-first strategies for refugees," *Economics of Education Review*, 92, 102353.

Arendt, J. N., I. Bolvig, M. Foged, L. Hasager, and G. Peri (2024): "Language training and refugees' integration," *Review of Economics and Statistics*, 106, 1157–1166.

Assaad, R., T. Ginn, and M. Saleh (2023): "Refugees and the education of host populations: Evidence from the Syrian inflow to Jordan," *Journal of Development Economics*, 164, 103131.

Bahar, D., R. J. Brough, and G. Peri (2024): "Forced migration and refugees: Policies for successful economic and social integration," NBER WP#32266.

Bleakley, H. and A. Chin (2004): "Language skills and earnings: Evidence from childhood immigrants," *Review of Economics and Statistics*, 86, 481–496.

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting event-study designs: Robust and efficient estimation," *Review of Economic Studies*, 91, 3253–3285.

Boucher, V., S. Tumen, M. Vlassopoulos, J. Wahba, and Y. Zenou (2021): "Ethnic mixing in early childhood: Evidence from a randomized field experiment and a structural model," CEPR Discussion Paper #15528.

Callaway, B. and P. H. C. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

Cavicchiolo, E., S. Manganelli, L. Girelli, A. Chirico, F. Lucidi, and F. Alivernini (2020): "Immigrant children's proficiency in the host country language is more important than individual, family and peer characteristics in predicting their psychological well-being," *Journal of Immigrant Minority Health*, 22, 1225—-1231.

Chiswick, B. R. and P. W. Miller (1995): "The endogeneity between language and earnings: International analyses," *Journal of Labor Economics*, 13, 246–288.

——— (2015): "International migration and the economics of language," in *Handbook of the Economics of International Migration*, ed. by B. R. Chiswick and P. W. Miller, Elsevier, vol. 1, chap. 5, 211–269.

de Chaisemartin, C. and X. D'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–2996.

——— (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey," *Econometrics Journal*, 26, C1–C30.

Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo (2018): "The economic consequences of hospital admissions," *American Economic Review*, 108, 308–352.

Dustmann, C. and A. Glitz (2011): "Migration and education," in *Handbook of the Economics of Education*, Elsevier, vol. 4, 327–439.

Dustmann, C. and A. van Soest (2001): "Language fluency and earnings: Estimation with misclassified language indicators," *Review of Economics and Statistics*, 83, 663–674.

European Commission-EACEA-Eurydice (2019): "Integrating Students from Migrant Backgrounds into Schools in Europe: National Policies and Measures," Tech. rep., Publications Office of the European Union, Luxembourg.

Fadlon, I. and T. H. Nielsen (2019): "Family health behaviors," *American Economic Review*, 109, 3162–3191.

Fasani, F., T. Frattini, and L. Minale (2022): "(The struggle for) Refugee integration into the labour market: Evidence from Europe," *Journal of Economic Geography*, 22, 351–393.

Foged, M., L. Hasager, and G. Peri (2024): "Comparing the effects of policies for the labor market integration of refugees," *Journal of Labor Economics*, 42, S335–S377.

Foged, M., L. Hasager, G. Peri, J. N. Arendt, and I. Bolvig (2023): "Intergenerational spillover effects of language training for refugees," *Journal of Public Economics*, 220, 104840.

Ginsburgh, V. and S. Weber (2020): "The economics of language," *Journal of Economic Literature*, 58, 348–404.

Gould, E. D., V. Lavy, and M. D. Paserman (2004): "Immigrating to opportunity: Estimating the effect of school quality using a natural experiment on Ethiopians in Israel," *Quarterly Journal of Economics*, 119, 489–526.

Gubbels, J., C. E. van der Put, and M. Assink (2019): "Risk factors for school absenteeism and dropout: A meta-analytic review," *Journal of Youth and Adolescence*, 48, 1637–1667.

Heller, B. H. and K. S. Mumma (2023): "Immigrant integration in the United States: The role of adult English language training," *American Economic Journal: Economic Policy*, 15, 407–437.

Hockel, L. S. and P. Schilling (2022): "Starting off on the right foot: Language learning classes and the educational success of immigrant children," Ruhr Economic Papers #983.

Lochmann, A., H. Rapoport, and B. Speciale (2019): "The effect of language training on immigrants' economic integration: Empirical evidence from France," *European Economic Review*, 113, 265–296.

Lounsbury, J. W., R. P. Steel, J. M. Loveland, and L. W. Gibson (2004): "An investigation of personality traits in relation to adolescent school absenteeism," *Journal of Youth and Adolescence*, 33, 457–466.

Miller, D. L. (2023): "An introductory guide to event study models," *Journal of Economic Perspectives*, 37, 203–230.

Ohinata, A. and J. C. van Ours (2013): "How immigrant children affect the academic achievement of native Dutch children," *Economic Journal*, 123, F308–F331.

Pont-Grau, A., Y.-H. Lei, J. Z. E. Lim, and X. Xia (2023): "The effect of language training on immigrants' integration: Does the duration of training matter?" *Journal of Economic Behavior & Organization*, 212, 160–198.

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 235, 2218–2244.

Schmid, L. (2023): "The impact of host language proficiency on migrants' employment outcomes," *American Economic Review: Insights*, 5, 477–493.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225, 175–199.

Tumen, S. (2018): "The impact of low-skill refugees on youth education," IZA Discussion Paper #11869.

——— (2023): "The case of Syrian refugees in Türkiye: Successes, challenges, and lessons learned," Background Paper for the *World Development Report 2023: Migrants, Refugees, and Societies*. The World Bank.

Tumen, S., M. Vlassopoulos, and J. Wahba (2023): "Training teachers for diversity awareness: Impact on school outcomes of refugee students," *Journal of Human Resources*, forthcoming.

UNHCR (2019): "Stepping up: Refugee education in crisis," The UN Refugee Agency.

——— (2024): "Global trends: Forced displacement in 2023," The UN Refugee Agency.

——— (n.d.): "Language training programmes," The UN Refugee Agency, Integration Handbook, https://www.unhcr.org/handbooks/ih/language/language-training-programmes.

Wing, C., M. Yozwiak, A. Hollingsworth, S. Freedman, and K. Simon (2023): "Designing difference-in-difference studies with staggered treatment adoption: Key concepts and practical guidelines," *Annual Review of Public Health*, 45, 485–505.

World Bank (2023): "World Development Report 2023: Migrants, Refugees, and Societies," The World Bank. Washington, DC.

The power of language: Educational and mental health impacts of language training for refugee children

Appendix

By **Semih Tumen**, **Michael Vlassopoulos**, and **Jackline Wahba**

A Additional tables

Table A1: Balance test (student level)

	Treated vs untreated
Turkish language scores (pre-treatment)	0.031
	(0.056)
Math scores (pre-treatment)	0.008
	(0.061)
School absences (pre-treatment)	-0.054
	(0.076)
Gender	0.012
	(0.024)
Age	0.613*
	(0.342)
Grade/year	-0.366
	(0.699)
Number of siblings	0.371
	(0.266)
Preschool attendance	0.009
	(0.048)
Origin is Aleppo	-0.031
	(0.082)
School fixed effects	Yes
Town-semester fixed effects	Yes
Grade/year fixed effects	Yes
#of observations	46,761

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Each row corresponds to a different regression, where the referred variable is the outcome variable and controls include a dummy variable indicating the treatment group, school fixed effects, town-semester fixed effects, and grade/year fixed effects. The sample only includes pre-program observations. Grade/year fixed effects are excluded in the grade/year balance test. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A2: Balance test (school level)

	Treated vs untreated
Turkish language scores (pre-treatment)	0.066
	(0.109)
Math scores (pre-treatment)	-0.108
	(0.116)
School absences (pre-treatment)	-0.324
	(0.245)
Gender	-0.014
	(0.044)
Age	0.544
	(1.311)
Grade/year	-0.612
	(1.065)
Number of siblings	0.206
	(0.412)
Preschool attendance	0.014
	(0.054)
Origin is Aleppo	-0.062
	(0.096)
# of teachers	0.766
	(1.312)
School type	0.032
	(0.061)
Town fixed effects	Yes
Town-semester fixed effects	Yes
#of observations	2,869

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Each row corresponds to a different regression, where the referred variable is the outcome variable and controls include a dummy variable indicating the treatment group, town fixed effects, and town-semester fixed effects. The sample only includes pre-program observations. Standard errors clustered at town level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A3: Balance test (school level—treatment timing)

	First vs Never	First vs Second	First vs Third
	[1]	[2]	[3]
Turkish language scores (pre-treatment)	0.041	0.009	-0.011
	(0.114)	(0.099)	(0.107)
Math scores (pre-treatment)	-0.121	-0.054	0.014
	(0.116)	(0.091)	(0.118)
School absences (pre-treatment)	-0.324	0.046	-0.226
	(0.294)	(0.283)	(0.301)
Gender	0.021	-0.008	-0.011
	(0.051)	(0.044)	(0.049)
Age	0.211	0.755	-0.111
	(1.406)	(1.003)	(1.285)
Grade/year	-0.108	-0.754	0.211
	(1.241)	(0.981)	(1.072)
Number of siblings	0.111	0.004	-0.053
	(0.518)	(0.397)	(0.404)
Preschool attendance	0.001	-0.028	0.031
	(0.079)	(0.054)	(0.066)
Origin is Aleppo	-0.111	0.041	0.028
	(0.108)	(0.079)	(0.091)
# of teachers	0.244	0.965	-0.362
	(1.511)	(1.045)	(1.298)
School type	-0.021	0.066	0.041
	(0.086)	(0.054)	(0.071)
Town fixed effects	Yes	Yes	Yes
Town-semester fixed effects	Yes	Yes	Yes
# of observations	1,262	1,951	1,492

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Each row corresponds to a different regression, where the referred variable is the outcome variable and controls include a dummy variable indicating the treatment timing, town fixed effects, and town-semester fixed effects. In each column, the treatment definition is different—comparing first- vs never-treated, first- vs second-treated, and first- vs third-treated groups in columns 1, 2, and 3, respectively. The sample only includes pre-program observations. Standard errors clustered at town level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A4: Overall effect: Excluding the never-treated group

	[1]	[2]	[3]	[4]	[5]
Turkish language scores					
All sample	0.274***	0.234***	0.214***	0.208***	0.179***
	(0.040)	(0.049)	(0.046)	(0.047)	(0.050)
Males	0.268***	0.229***	0.201***	0.205***	0.171***
	(0.049)	(0.055)	(0.052)	(0.053)	(0.056)
Females	0.290***	0.242***	0.221***	0.215***	0.184***
	(0.050)	0.057)	(0.053)	(0.051)	(0.058)
Math scores					
All sample	0.224***	0.171**	0.218***	0.185***	0.132*
	(0.051)	(0.065)	(0.054)	(0.055)	(0.066)
Males	0.214***	0.165**	0.208***	0.177**	0.114
	(0.062)	(0.069)	(0.064)	(0.068)	(0.072)
Females	0.240***	0.182**	0.231***	0.189**	0.151*
	(0.066)	(0.070)	(0.067)	(0.069)	(0.073)
School absences					
All sample	-2.848***	-2.267***	-2.024***	-2.044***	-1.844***
	(0.325)	(0.331)	(0.353)	(0.360)	(0.378)
Males	-2.998***	-2.379***	-2.094***	-2.065***	-1.891***
	(0.366)	(0.372)	(0.391)	(0.389)	(0.407)
Females	-2.744***	-2.054***	-1.991***	-1.975***	-1.699***
	(0.371)	(0.386)	(0.396)	(0.395)	(0.422)
School fixed effects	Yes	Yes	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes
Town-semester fixed effects	No	Yes	No	No	Yes
Grade/year fixed effects	No	Yes	No	Yes	Yes
Semester fixed effects	Yes	No	Yes	Yes	No
# of observations			41,316		
(# of males)			20,351		
(# of females)			20,965		

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A5: Summary statistics:
(Mental health sample)

Turkish language score (pre-treatment)	51.44
	(23.62)
Math score (pre-treatment)	47.16
	(25.11)
School absences (pre-treatment)	10.98
	(7.63)
Age	10.14
	(2.91)
Grade/year	5.65
	(2.44)
Number of siblings	3.56
	(2.94)
$\mathbb{P}(\text{Male} = 1)$	0.50
$\mathbb{P}(\text{Primary school} = 1)$	0.67
$\mathbb{P}(\text{Gaziantep} = 1)$	0
$\mathbb{P}(\text{Preschool attendance} = 1)$	0.09
$\mathbb{P}(\text{Origin is Aleppo} = 1)$	0.49
$\mathbb{P}(\text{Absenteeism} \geq 10 \text{ days})$ (pre-treatment)	0.44
$\mathbb{P}(\text{Absenteeism} \geq 20 \text{ days})$ (pre-treatment)	0.21
# of observations	3,246

Notes: The grading scale for the Turkish language and Math assessments is 0-100. The table reports the averages and standard deviations (in brackets). The mental health sample includes only the Syrian students attending school in Şanlıurfa.

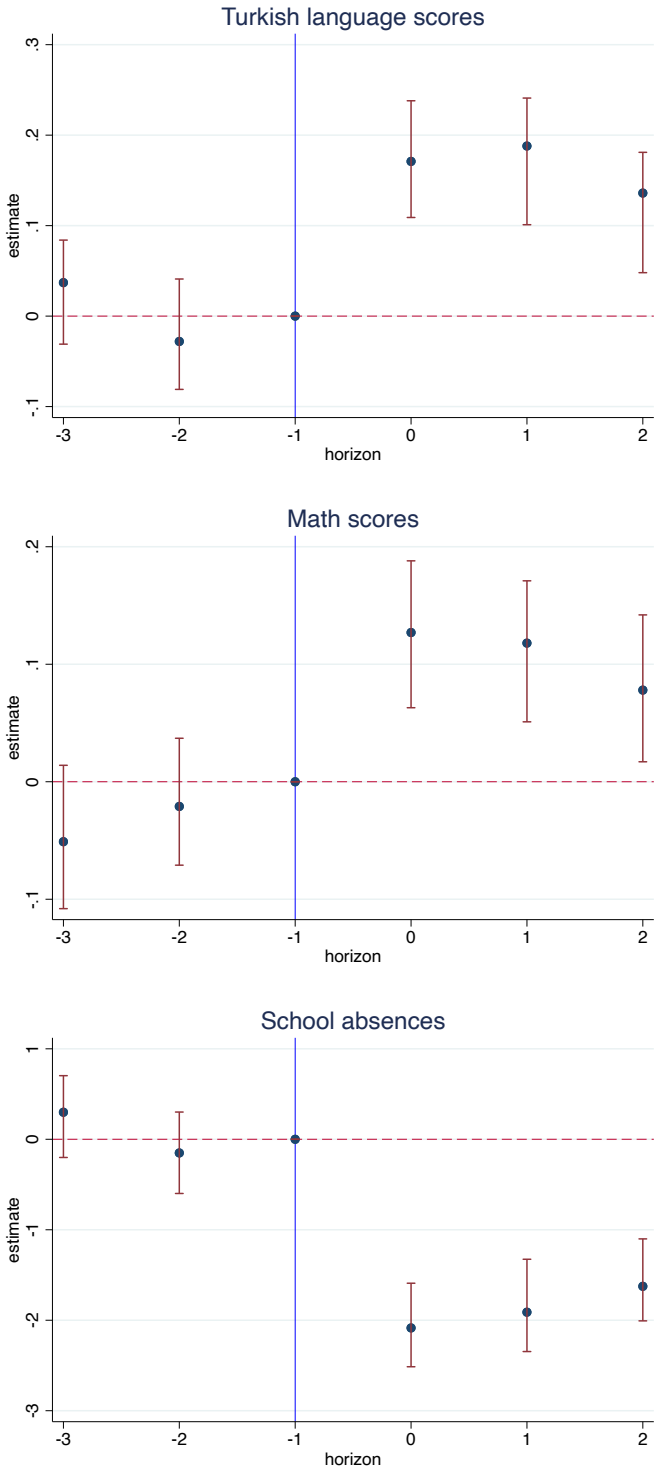
Table A6: Language training impact on the mental health sample

	[1]	[2]	[3]	[4]	[5]
Turkish language scores					
Program effect	0.326***	0.274**	0.264**	0.235**	0.208*
	(0.099)	(0.103)	(0.108)	(0.101)	(0.102)
Math scores					
Program effect	0.245**	0.188	0.222*	0.201	0.145
	(0.101)	(0.107)	(0.104)	(0.106)	(0.111)
School absences					
Program effect	-2.745***	-1.965**	-1.845**	-1.811**	-1.746*
	(0.728)	(0.845)	(0.866)	(0.868)	(0.877)
School fixed effects	Yes	Yes	No	No	No
Student fixed effects	No	No	Yes	Yes	Yes
Town-semester fixed effects	No	Yes	No	No	Yes
Grade/year fixed effects	No	Yes	No	Yes	Yes
Semester fixed effects	Yes	No	Yes	Yes	No
#of observations			3,246		

Notes: Turkish language and Math scores are standardized (around mean zero and unit standard deviation) within the same school, semester, and grade level. Standard errors clustered at school level are reported in brackets. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

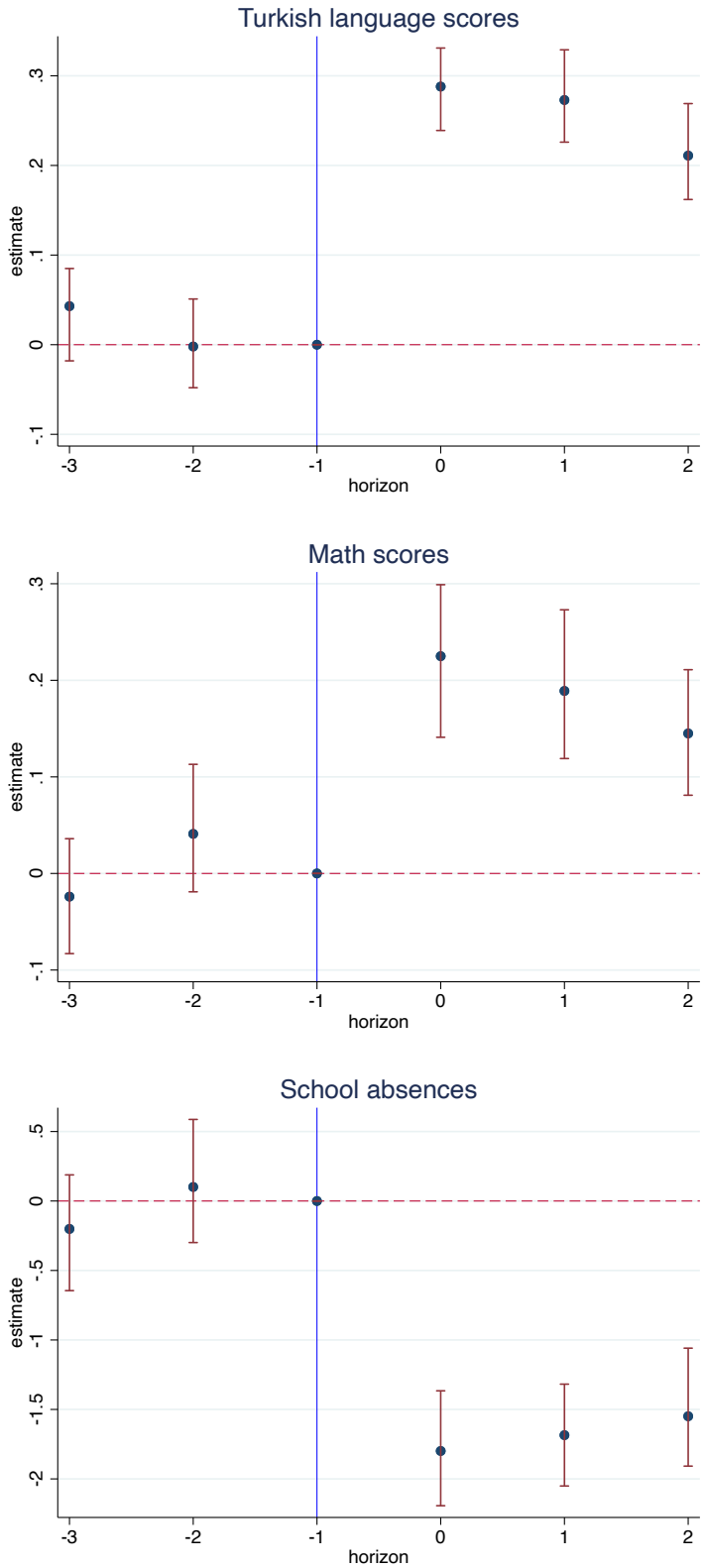
B Additional Figures

Figure A1: Event study results: Never-treated vs treated once



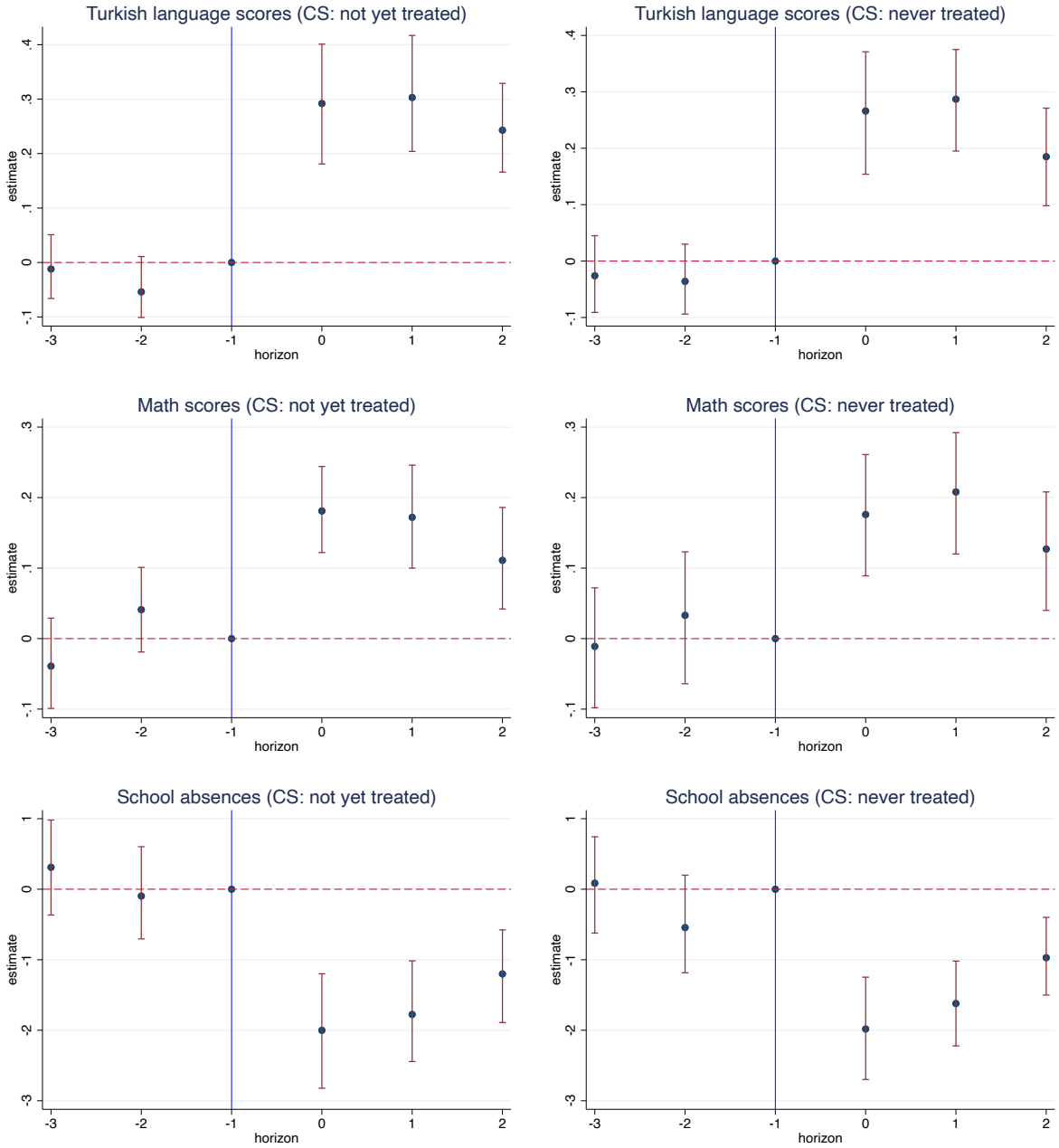
Notes: Timing of treatment is centered around 0 for different enrollment cohorts. Event study analysis described in Section 3.3 compares students in the ever-treated and treated-once schools.

Figure A2: Event study results: Never-treated vs treated twice



Notes: Timing of treatment is centered around 0 for different enrollment cohorts. Event study analysis described in Section 3.3 compares students in the ever-treated and treated-twice schools.

Figure A3: Event study results: Callaway and Sant'Anna (2021) (CS) estimator



Notes: The event study analysis in this plot is based on the Callaway and Sant'Anna (2021) (CS) estimator.