WORKING PAPER 2026:1

Supporting integration in schools: Evidence from a randomized targeted intervention

Hanna Mühlrad
Kristina Sibbmark
Anna Sjögren
Malin Tallås Ahlzén

IFAU
INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Supporting integration in schools: Evidence from a randomized targeted intervention [a]

Hanna Mühlrad[b], Kristina Sibbmark[c], Anna Sjögren[d] and Malin Tallås Ahlzén [e]

December 12, 2025

Abstract: We evaluate a two year school development program aimed at enhancing the quality of education for recent migrant students and Swedish language learners through customized packages of professional development for teachers and support of school management. We exploit the pairwise randomized roll out to 63 municipalities between 2016 and 2019 to examine effects on student achievement and explore underlying mechanisms. Over a 7 year follow-up period, core subject test scores improved by 0.021 sd, driven by a 0.032 sd improvement in mathematics performance. Test score gains in mathematics were present for students, regardless their background, during and post-implementation. Swedish test score gains generally materialized post-implementation, while second generation immigrant students gained already during implementation. Test score gains are more pronounced for foreign background students, for boys and at the low end of the test score distribution. The support program passes a cost-benefit test.

Keywords: refugee migration; dialogue based school improvement, host country effects, RCT
JEL-codes: I21 I22 I24 I28 H52 J45, F22

---

[a] We thank Adrian Adermon, Nihad Bunar, Aina Bigestans, Susan Dynarski, David Figlio, Eric Taylor, Petra Thiemann, and Björn Öckert for valuable comments and suggestions. We are thankful for generous help and access to information about the intervention from the National Agency for Education and especially Anneli Melén's patience and support. We are grateful to Helena Holmlund for contributions to the project early on. We also thank participants at the FEW Workshop in Malaga, Nordic Summer Institute in Labor Economics, and seminar participants at IFAU. We gratefully acknowledge funding from Forte, grant STY-2021-00736. A pre-analysis plan was registered as AEARCTR-0010109.

[b] Global Public Health, Karolinska Institutet, hanna.muhlrad@ki.se

[c] Institute for Evaluation of Labour Market and Education Policy (IFAU), kristina.sibbmark@ifau.uu.se

[d] Institute for Evaluation of Labour Market and Education Policy (IFAU) and UCLS, anna.sjogren@ifau.uu.se

[e] Institute for Evaluation of Labour Market and Education Policy (IFAU), malin.tallas.ahlzen@ifau.uu.se

1 Introduction

The complex interplay between globalization, climate change, and geopolitical conflicts has resulted in increased immigration rates in many countries, increasing the need to identify effective integration policies in general, and to develop the ability of schools to accommodate growing numbers of migrant and multilingual students (Dustmann, Schönberg, and Stuhler, 2016; Pinson, Bunar, and Devine, 2023).[1] The OECD recommended already in 2015 that schools "Integrate language and subject learning from the earliest grades" (OECD, 2015), based on the growing evidence from research on multilingualism and majority language acquisition supporting such approaches (see e.g Cummins, 2019; Ruiz De Zarobe and Jiménez Catalán, 2009). Yet, an important question is if and how central school agencies can promote improved immigrant student accommodation and make integration a reality in the classroom, especially in view of challenges in promoting local ownership and buy-in among school professionals (Kraft, Blazar, and Hogan, 2018). Also, there may be fears that focusing resources to improve education for migrant students, may encourage native flight (Farre, Ortega, and Tanaka, 2018). Furthermore, most of the evidence relating to multilingual learners, majority language acquisition and refugee or migrant school integration, is to date qualitative or based on rather small scale studies (Kalinowski, Gronostaj, and Vock, 2019; Palik and Østby, 2023; Stolk, Kaplan, and Szwarc, 2023). Evidence on how to effectively promote immigrant student integration and learning based on national policy initiatives, is lacking.

In this paper we study the effects on student performance and school responses of a targeted national 2-year support program aiming to improve the reception and integration of refugee and immigrant students, and to improve the quality of education for recent immigrant students, Swedish language learners and multilingual students (Skolverket, 2021). The program was launched in response to the large influx of migrants and asylum seekers in the years leading up to the 2015-2016 Syrian refugee crisis (Bunar, 2017). During these years, the overall share of recent immigrant students doubled and reached a high of 8 percent in 2018, with even larger impact in the rural areas where refugees were placed (Getik, Sjögren, and Sundberg, 2024; Mörtlund, 2020).[2]

The support program, which targeted municipalities with a high influx of migrants and limited previous experience in integrating migrant students, was rolled out using pairwise randomization between 2016 and 2019. In promotion of local engagement and relevance, the program was based on dialogue and customized to the needs of each treated municipality following a needs assessment analysis. The analysis was conducted in collaboration between a local team and consultants from the Swedish National Agency for Education (SNAE) during the first six months of the program. An agreed upon support package

[1] In Sweden the share of foreign-born residents has doubled from just over 10% in the 2000 to some 20% two decades later (SCB, 2025).

[2] The group of recent migrant students is comprised of asylum seekers and foreign born students with at most 4 years of residency. Getik, Sjögren, and Sundberg (2024) show that test scores of incumbent students actually increased and schools responded by reducing class sizes.

was then implemented with financial and managerial support from the SNAE.[3] A central component of the implemented support packages was professional development (PD) and coaching for teachers and school professionals to develop language awareness and how to integrate language development and subject learning for Swedish learners into teaching practices.[4] The program also included training modules to improve school management practices, assessment of migrant students' skills, and many other modules relating to improving integration of new students e.g training related to tutoring in mother tongue and home language classes.

We exploit the pairwise randomization of municipalities, which were ranked according to a needs-index in 7 rounds, with 4-13 treated municipalities per round. The randomization resulted in 63 treated and 63 control municipalities out of Sweden's total of 290 municipalities. In total, 796 compulsory schools and some 15,000 teachers and 160,000 students in the 63 treated municipalities were affected by the program, making this a large scale experimental program, and unique in research of multilingual students and promotion of migrant student integration.[5]

We evaluate the effects of the program by comparing student mean performance on standardized national tests in core subjects (mathematics and Swedish) taken in grades 3, 6, and 9. To this end, we make use of universal administrative data and information extracted from the agreements between the SNAE and the participating municipalities. We measure effects on test scores up to 7 years after entering the program, focusing on the overall effects as well as on what happens during and after the implementation period to assess if the program had any initial crowding-out effects and if there was lasting impacts.[6]

We find that in the 7-year follow-up since program start, the average core subject test scores improved by 0.021 sd. The positive effect is driven by a 0.032 gain in mathematics, whereas the average gain in Swedish was small and insignificant. Test score gains in mathematics were present for all students, regardless of their background, and driven by improvements both during program implementation and post-implementation period. Swedish test scores, while improving only for second generation immigrants during the implementation, instead improved for all groups in the years just after the implementation. When accounting for compositional changes, the full 7 year follow-up gains in Swedish become marginally significant for recent immigrant students.

[3] In a very different context of medium sized enterprises in Mexico, Bruhn, Karlan, and Schoar (2018) evaluate the impact of an RCT to provide consultancy services to improve management, which were customized to the participating enterprises.

[4] These modules build on insights from research on Content and Language integrated learning (CLIL) and Language scaffolding (see e.g., Breidbach and Viebrock, 2013; Cummins, 2019; Gibbons, 2015; Ruiz De Zarobe and Jiménez Catalán, 2009). The Swedish term used is "Språk och Kunskapsutvecklande Arbetssätt" (SKUA).

[5] In a review of coaching programs Kraft, Blazar, and Hogan (2018) categorize studies with more than 100 teachers as large.

[6] Teaching activity could be crowded out by PD activities, substitute teachers are likely less effective. Moreover organizational change could initially be disruptive. Fade-out of intervention effects can result if teachers or schools more broadly over time revert to pre-intervention practices, see e.g. Sims et al. (2025).

The targeted support program thus had positive effects on learning in both mathematics and Swedish, and there is evidence of sustained gains of an order of magnitude enough to pass a cost benefit test, giving thirteen euros in return for every euro spent. Despite the program's focus on immigrant students, the gains emerged earlier among second-generation immigrants and students with a Swedish background, whereas positive effects for foreign-born students emerged only after the implementation period. We further find that the positive effects among Swedish born students are present mostly for boys, while there are gains for both boys and girls among immigrant students. The test score gains were also larger for poorly performing students.

Although we cannot directly observe how the program might have changed practices in the classroom, we are able to analyze the program's impact on access to language related pedagogical resources, class room organization and teaching resources.[7] We find that treated municipalities, while expanding student access to language related pedagogical resources, did so with a small delay for immigrant students relative to control municipalities. This delayed response may reflect an overall increased focus on recent immigrants–even in control municipalities–while the intervention targeted all multilingual students and produced more sustained effects by changing teaching practices. We further find that class sizes increased and that there was some increase in the teacher–student ratio in the post-implementation period, suggesting that class room organization changed. We also find that treated municipalities were more likely to place recent migrants in reception classes during the program's implementation phase, but over time they became less likely to segregate students. [8] It is possible that late access to language enhancing resources and initial segregating practices contributed to delayed program effects among immigrant students, at the same time there is suggestive evidence that early test score gains among second generation students might be a consequence of the programs' focus on language awareness and support of language and learning enhancing teaching practices.

This paper makes contributions at the intersection of several disciplines e.g. economics of education, sociology of education and education science. First, we provide unique experimental evidence of the effectiveness of a large scale national policy initiative attempting to improve the quality of education of migrant and multilingual students through professional development and managerial support in ordinary schools, thus contributing to the growing field of migration and education, recently summarized in Pinson, Bunar, and Devine (2023).

Second, our paper adds to the literature on teacher professional development, reviewed in e.g. Taylor (2023) and Kraft, Blazar, and Hogan (2018). We contribute with experimental evidence from a policy that by design aims to increase local ownership and buy-in

[7] Language resources include instruction in the home language, second language instruction in Swedish, and tutoring in the home language. Note, that students speaking a language other than the majority language Swedish at home are entitled by law, but not obliged, to participate instruction to learn and develop their home language.

[8] We define a classroom as a reception class if more than 90 percent of the students are recently immigrated. See Bunar and Juvonen (2022) for a critical discussion of segregating practices in Swedish schools.

among school professionals, something which have been identified as key to successful professional and school development programs (Kraft, Blazar, and Hogan, 2018). Because of the targeted program's immigrant and language focus, our paper is particularly informative about the effects of professional development in language awareness and Content and Language integrated learning (CLIL), an area where experimental evidence is rare (Cummins, 2019; Kalinowski, Gronostaj, and Vock, 2019; Stolk, Kaplan, and Szwarc, 2023).

Furthermore, our results are informative of the effectiveness of top-down school development programs and thus relate to the school turnaround literature, in particular because of the intervention's multifaceted nature, and the fact that participating municipalities had rather poorly performing schools initially (Redding and Nguyen, 2020; Schueler et al., 2022). Interestingly, in spite of the programs focus on language, we find positive significant effects effects in mathematics in the post-implementation period, in line with the meta-study results of Schueler et al. (2022).[9] More specifically, we contribute by evaluating a large scale collaborative, customized support program in a school context (Bruhn, Karlan, and Schoar, 2018). To the best of our knowledge, there are no previous evaluations of such school improvement programs, let alone, based on experimental designs.

More generally, our paper also relates to the literature studying education effects of migration (e.g. Brandén, Birkelund, and Szulkin, 2019; Figlio et al., 2023; Figlio and Özek, 2019; Getik, Sjögren, and Sundberg, 2024; Gould, Lavy, and Daniele Paserman, 2009), and more broadly to studies of host country effects of migration in general (Dustmann, Schönberg, and Stuhler, 2016).

The rest of the paper is organized as follows. Section 2 describes the Swedish school system and the refugee crisis, followed by a description of the intervention, experimental design and data in Section 3. Section 4 presents the results, and we conclude in Section 5.

2 Background: The Swedish scshool system and the refugee crisis

This section provides a brief description of some institutional factors of the Swedish school system, and some background regarding the integration of immigrant students and the impact of the 2015 refugee crisis on schools.

2.1 The Swedish school system and immigrant integration

School is compulsory and free of tuition for resident children aged 6 to 16. Also non-resident, asylum seeking school age children, are offered free schooling during the asylum

[9]Moreover, Holmlund, Häggblom, and Lindahl, 2024 find that teacher in-service training in how to promote reading and writing proficiency had effects on test scores in civics and science, in addition to Swedish.

process. The Swedish school system is decentralized. While there is a national curriculum, national guidelines, and mandatory national tests developed and administered by the National Agency for Education, municipalities are responsible for funding schools and they are required to fund both municipal-run schools and licensed independent schools (Skollag, 2010).[10] In addition to municipal funding, school heads (municipal and independent) and schools can obtain - often after an application process - government funding in the form of earmarked grants, administered by the SNAE. Schools and/or school heads (municipalities or independent school organizations) can also be offered various forms of nationally funded support programs, such as the program studied in this paper. Participation in the support program is then voluntary.

Compulsory school comprises three stages, the lower stage from the preschool year (grade 0) to 3rd grade, the middle stage from 4th to 6th grade, and the upper stage from 7th to 9th grade.[11] As part of a national accountability system, students take mandatory national tests in the core subjects: Mathematics and Swedish at the end of each stage, and English in grades 6 and 9.[12] These national tests are locally graded at the school using national guidelines and in 6th and 9th grade, test results serve as guidance when teachers set the end-of-year grades.[13] The 9th grade national tests are high-stakes for students since they influence the final compulsory school grades, determining high school eligibility and competitiveness in admission to schools/high school programs. In 3rd grade the national tests are only used to screen if students are at risk of falling behind.

Municipalities are responsible for providing school placements for all school-age children in the municipality. This implies that some municipal schools need to maintain slack. Independent schools are not required to take more students than they have room for, they can operate at their chosen capacity and apply their own admission process. Thus, although families are free to apply to the school of their choice,[14] children moving to a municipality during their school years (including recent migrants) are mostly received in municipal schools with free slots or where new slots are created.

There has been rising school segregation since the 1990's, as a result of increased res-

[10] Independent schools are entitled to municipal funding based on their number of students, provided they follow the national regulations, which include not charging fees to students (see e.g., Holmlund, Sjögren, and Öckert, 2019).

[11] Grade configuration varies across schools, but a majority of schools have all three stages.

[12] In 9th grade, there are national tests also in one of the social science subjects (Geography, History, Religion, Social science) and one of the natural science subjects (Biology, Chemistry, Physics), which subject is randomized at the school level.

[13] Vlachos (2019) shows that although the test grades are subject to teacher subjectivity, they are more objective measures of student performance than the teacher set end-of-year grades.

[14] Families can choose schools for their children in the sense that they can wish for, or apply to, a specific school–either municipal or independent. While admission systems are coordinated between the municipal schools, most independent schools run their own admission processes. Municipal schools admit students based on parental preferences and residential proximity, conditional on providing school placements within a reasonable distance from the home (Björklund et al., 2004). Independent schools can instead choose to admit students based on either residential proximity or queue time (in queues they administer themselves) and they can give siblings preferential treatment (Skollag, 2010).

idential segregation, school choice and independent school expansion (Böhlmark, Holmlund, and Lindahl, 2016).[15] There is also a large performance gap between immigrant and native students, which Grönqvist and Niknami (2017) find is largely accounted for by socioeconomic background and neighborhood effects. However, Holmlund, Sjögren, and Öckert (2019) show that school segregation in the immigrant/native dimension actually declined as a result of the large influx of migrants during the years around 2015 refugee crisis. The fraction all-native schools–mostly located in rural areas–declined, as they received a large share of the refugees due to their capacity to provide housing.

Since the mid-1970's, students speaking a language other than Swedish at home have had the right to instruction aiming to support the development of that language, their Swedish language development and other subject learning (Regeringen, 1975; Utbildningsdepartementet, 1977). In the 1980 Curriculum, second language learner Swedish instruction was formalized in the subject Swedish as a second language, and students have since been assigned either to (ordinary) Swedish or Swedish as second language, depending on teacher or headmaster assessments (Skolöverstyrelsen, 1980).

Until the 2015 refugee crisis, there were few formal regulations and recommended practices regarding the reception of immigrants in schools. Moreover, local capacity and demand–rather than mandates and formal regulation–have governed student access to various forms of language support, resulting in varying practices across schools and municipalities (Bunar, 2010). Municipalities with more immigration organized reception classes, but it was up to head masters to place immigrant students in appropriate classrooms, depending on local conditions. With increasing migration influx in the 2010's, the absence of policies and guidelines became a concern and led to more formalized national guidelines, both regarding placement of students and access to language support (Regeringen, 2014; Skolverket, 2015, 2016; Utbildningsdepartementet, 2013).

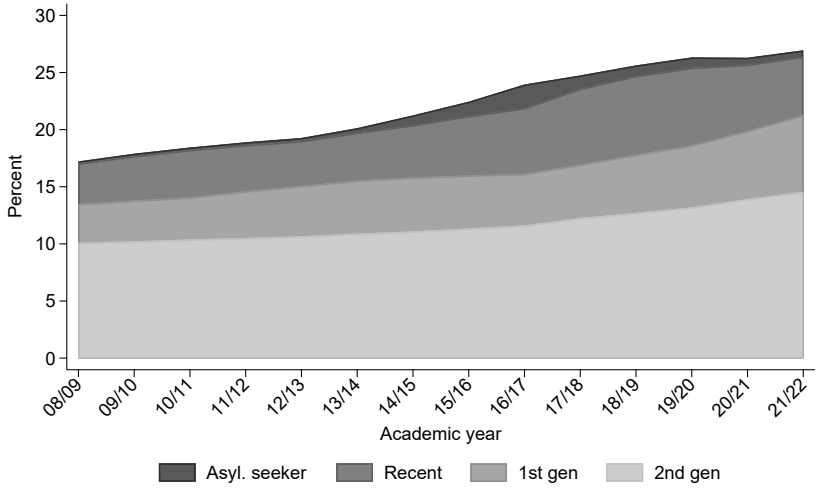
2.2 The refugee crisis and Swedish schools

Leading up to the refugee crisis of 2015, the fraction of recent immigrant students rose rapidly in Swedish schools. Until 2016 when migration policy radically tightened, Sweden had the highest per capita refugee inflow in Europe. This development is shown in Figure 1. The fraction of second-generation immigrant students rose from about 10 percent in 2008/2009 to 13 percent in 2021/2022. At the same time, the total group of first-generation immigrant students (comprised of foreign-born students with more than four years of residency, recently arrived with at most four years of residency, and asylum seekers) doubled from about 6 percent to over 12 percent. Although asylum seekers make up a small share of the overall student population, there was a clear peak of 2.5 percent during the crisis years around 2015–2017. As these students become residents, the group of recent migrants grows. Over time, a growing fraction of first-generation immigrant

[15] While some 70 percent of the increase in the between-school variation in a composite measure of student background was due to rising residential segregation, 30 percent is accounted for by independent school expansion and school choice (Holmlund, Sjögren, and Öckert, 2019).

students also accumulate more than four years of residency.

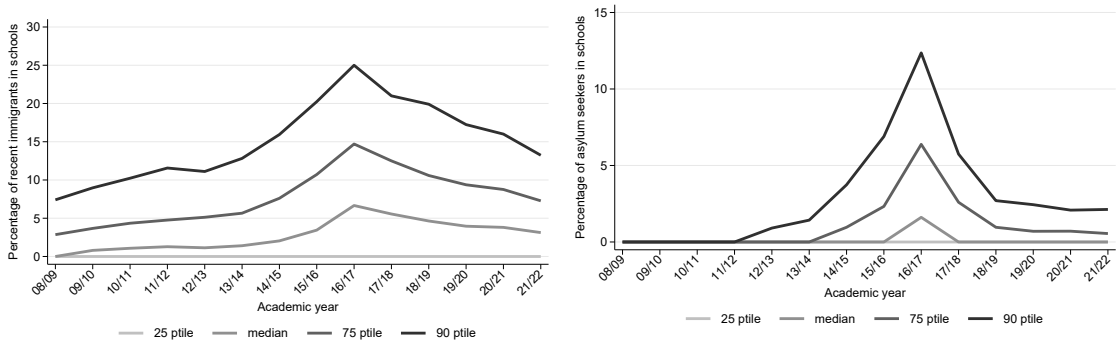
Figure 1: Stock of compulsory school students by migration background



Note: The figure shows the share of foreign-born and foreign background (two foreign born parents) students for the 2008–2022 period, by migrant status. "Asylum seekers" are non-resident students with asylum seeking status, "Recent" includes foreign-born students with at most four years of residency, "1st gen" are foreign-born students with more than four years of Swedish residency, and "2nd gen" are students born in Sweden to two foreign-born parents.

These average numbers hide significant heterogeneity in the fraction of recent migrant students across schools. Figure 2 therefore takes a closer look at how the share of recent migrants and asylum seekers developed across the distribution.

Figure 2: The distribution of Recent migrant and asylum seeking students across schools



(a) Share of recent migrant students by percentile

(b) Share of asylum seeking students by percentile

Note: The figures (a) and (b) show trends in the distribution of School by grade level share of resent migrant and asylum seeking students at different percentiles of the distribution in Sweden between 2008/2009 and 2021/2022.

The patterns visualized in Figure 2 are evidence of the clustering of immigrant background students in certain schools. In particular, many refugees were received in small rural municipalities because there was accommodation available and because temporary accommodation could rapidly be arranged in camping sites, hostels and vacation cabin

villages. However, theses municipalities had little previous experience with immigration,[16] which is why they launched the targeted support program studied in this paper.

Many of the municipalities receiving refugees had only one or a few schools, and only a minority actively tried to counteract school segregation when assigning newly immigrated children to schools (Mörtlund, 2020). To improve refugee reception in schools, and increase equity across schools, the government introduced general support to all schools, and they also provided direct per student support for schools receiving asylum seeking students (see e.g., Bunar, 2017; Mörtlund, 2020). A major challenge to municipalities and schools was the lack of experience in integrating students with a refugee background– some of whom had little prior schooling, had experienced trauma, and no knowledge of the majority language, Swedish. Yet, Getik, Sjögren, and Sundberg (2024) find a weak positive effect on incumbent student test scores of being exposed to recent migrants for the years 2008-2022, and in particular the 2015 refugee crisis. Increased resources in the form of reductions in class size and increased take up of home language classes were plausible explanations.[17] The targeted support program in focus in this paper may also have contributed to the positive effects.

3 The intervention, experimental design and data

This section describes the targeted support program studied and the design of its randomized roll-out. We also describe our data and present the identification strategy.

3.1 The targeted support program

Our aim is to evaluate the effects on student outcomes of the two-year targeted dialogue-based school support program that was initiated by the Swedish government and administered by the SNAE, in response to the 2015 refugee crisis.

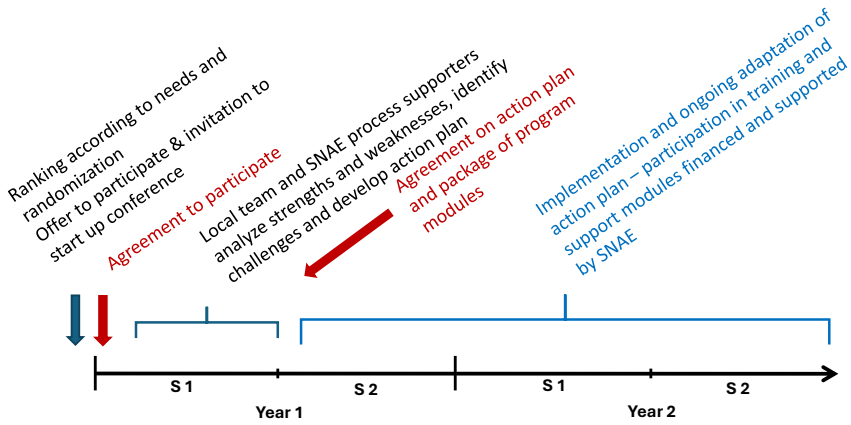
Figure 3 presents the timeline of the intervention. First, municipalities selected into treatment received an invitation to participate in the program. Once they accepted, a local team consisting of representatives from the municipality's school board and schools initiated a 6 month analysis period, in collaboration with and assisted by consultants from SNAE. During this initial stage of the program, strengths and weaknesses of the municipality's schools, local needs and suitable support measures were identified. A customized support package was put together and agreed upon in a contract between the municipality and the SNAE. The agreed-upon package of support measures was then implemented by the municipality during the program's second stage. This stage was initially planned to

[16] The change in the geographical distribution of recent migrant student shares is illustrated in Appendix Figure B1.

[17] In other European contexts, only a few papers (e.g., Green and Iversen, 2022; Hassan et al., 2023; Tumen, 2021) analyze effects of refugee migration on student outcomes and only Tumen (2021) uses data covering the acute refugee crisis, also finding positive effects.

last 18 months but was sometimes extended. Both stages took place with financial and managerial support from the SNAE.

Figure 3: Time line of the targeted intervention



Note: The figure shows an idealized timeline of the targeted intervention from selection into treatment through randomization and through the implementation phase.

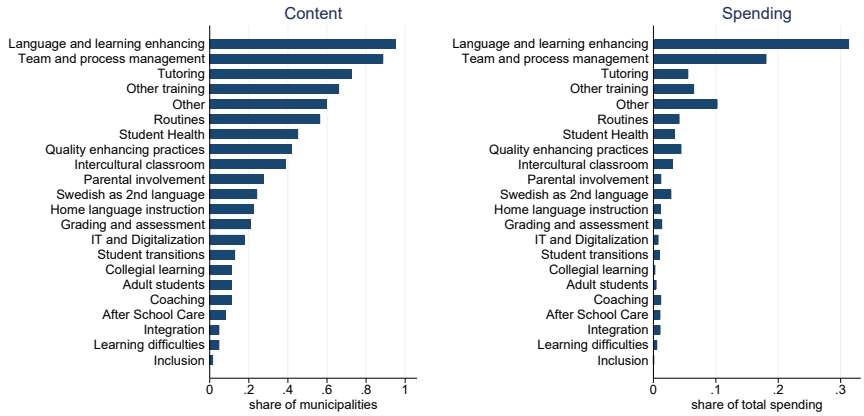
Our evaluation of program effects takes into account possible dynamic effects of the program.[18] It is plausible that the program's impact varies between the initial 24-month phase, consisting of the analysis (6 months) and implementation (18 months) periods, and the post-treatment period, during which teachers and other school professionals have participated in training modules and adopted new practices. Initially, effects might even be negative if resources are absorbed in the analysis process, reorganizations are disruptive in the short run or if teacher participation in in-service training crowds out regular teaching activities or exposes students to less efficient substitute teachers. Follow-ups of professional development programs, e.g. US School Improvement Grants (SIG) school turnaround programs, have indeed shown that potential benefits can take time to materialize (Grönqvist, Öckert, and Rosenqvist, 2025; Sun, Kennedy, and Loeb, 2021; Sun, Penner, and Loeb, 2017). At the same time, program effects may fade after program participation if teachers and organization revert back to pre-treatment practices (Sims et al., 2025).[19]

[18] Ideally, we would like to trace the year-to-year dynamics of the program's impact. Limited power and lacking data during the pandemic, however, restrict us to aggregating into two-year periods.

[19] It cannot be ruled out that treatment effects vary across waves. Such variation could reflect differences in the SNAE's analytical and consulting capacity affecting the quality of the support packages, or differences in municipalities' ranking in terms of need or in the time elapsed since the crisis. Unfortunately we do not have sufficient power to explore this.

Based on the agreed action plans between SNAE and the participating municipalities, we are able to characterize the content of the support packages and allocation of spending (see Appendix Section C for more details on the different modules).[20] Figure 4 summarizes this information showing the share of treated municipalities' action plans that contain various modules by order of magnitude (left) and the share of total budgeted spending allocated to each module (right). Almost all support packages involved teacher training in language and learning enhancing teaching strategies to support language development and learning of migrant students. More than 30 percent of the total budget was also allocated to this type of professional development.[21] Most packages also included team and process management components connected to the intervention itself. In addition, many packages included training of teachers and tutors involved in tutoring subject learning in the student's mother tongue, as well as training of home language teachers. Moreover, several packages involved developing routines and providing managerial and administrative support for organizing refugee student reception and integration, including the assessment of newly arrived students' initial skills. Packages further involved appointment of a local refugee reception coordinator and training of other personnel groups involved in the schools receiving refugee children. Combining the shares of spending allocated to all forms language oriented modules, i.e., tutoring in mother tongue, Swedish as a second language and home language instruction and Language and learning enhancing professional development, some 40 percent of the budget was allocated to language oriented professional development.

Figure 4: The support program modules: Content and spending



Note: The lhs figure shows share of treated municipalities receiving a given support module. The rhs figure shows the share of spending spent on the various modules based on information extracted from agreements between SNAE and municipalities. Note that budget information is not available for all municipalities.

[20] Many agreements contain detailed budget information on budgets for each planned module. This is used to infer what share of spending goes to the various modules.

[21] This PD-module is developed based on an amalgam of ideas from research on teaching multilingual learners, i.e. Scaffolding language and Content and language integrated learning (CLIL) (Cummins, 2019; Gibbons, 2015; Ruiz De Zarobe and Jiménez Catalán, 2009).

3.2 Randomization and roll-out

The support program was rolled out in seven waves during 2016-2019, one per semester, starting in the spring of 2016.[22] In each wave, municipalities were first ranked by their need of support according to a needs-index. The index was based on measures of previous experience in receiving refugees, presence of school age refugees and newly arrived migrants in the municipality, and measures of predicted school performance of the present group of recent migrants. In each wave, the five top ranked municipalities were guaranteed participation in the program (referred to as guarantee municipalities). The reason was to adhere to the government's intention that the program, while allowing for effect evaluation, also reached the municipalities in most need of support. Randomization into treatment and control then took place among subsequent municipal pairs, i.e. 6-7, 8-9... each round. Control municipalities could re-enter the selection process in the second to next round.[23] Also this was required by the SNAE to increase the chances of high-need municipalities receiving treatment. However, re-sampling of control municipalities came at the cost of losing some randomization pairs in the evaluation. To ensure that treated municipalities are only compared to their yet untreated control municipalities, the initial randomization pair is excluded from further follow up once a control municipality receives treatment.

As mentioned, municipalities randomized into treatment were contacted by the SNAE, informed about the program and offered to participate. Those agreeing to do so signed a first agreement with the SNAE. The time until an invitation was accepted varies between municipalities within a sampling round. Some municipalities declined, others wanted to postpone participation. Since acceptance and time to contract may be endogenous to characteristics of the municipality, outcomes measured after randomization are considered potentially affected by the program. Also the municipalities randomized into the control group were contacted by the SNAE. They were informed about the program and that they later could re-enter the selection process.

The initial ambition of the SNAE was to include 10-15 treatment municipalities per round, but capacity constraints allowed for fewer. Hence 5-12 municipal pairs were randomized to treatment and control each round, resulting in a total of N=63 in the treatment arm and control arm respectively. Out of the 63 control municipalities, 24 eventually received treatment, either as a guarantee municipality or as part of the evaluation due to re-entry into randomization. The sampling process is described in Tables A1 – A3. In total, the support program thus affected some 160,000 students, 15,000 teachers, and almost 800 schools at the compulsory school level, excluding the guarantee municipalities.

[22] The model was developed in collaboration between the SNAE and IFAU.

[23] The needs index was re-computed each round, implying that an initially highly ranked control municipality would not necessarily enter as a guarantee municipality in later waves, should conditions have changed.

3.3 Data and measurement

We use data from Swedish administrative student registers covering the universe of compulsory school students in grades 0-9 for the years 2013-2023. These data are linked to population, school, and teacher registers allowing us to link students to their parents and schools. Furthermore, we use information on the needs-index rankings provided by the SNAE, the agreements between the SNAE and the participating municipalities, and other program documentation. The rich information about the program allows us to characterize the intervention packages and, to the extent possible, assess the relative importance of the modules and the costs of the program.

We are interested in the program effects on all students in the treated municipalities. Hence, we include students irrespective of their migration background. Since the aim of the program was to improve the quality of education not only for recent migrant students but also for other groups of multilingual students, we might expect also incumbent students with foreign background to benefit from language enhancing investments, in particular since the treated municipalities had little experience with foreign background students. Furthermore, any impact on teacher qualifications, school- and classroom organization may affect also Swedish background students. Second, we include students irrespective of the type of school they attend. Even though the support was targeted at the municipal schools, where most recent immigrants were also enrolled, we include students in independent schools in order to avoid concerns with endogenous school placements and selection. We provide a robustness check where independent schools are excluded from the analysis (see Appendix Table A4), showing similar results as the main analysis and we show in Table 1 that participating municipalities have a lower share of students in independent schools than the country on average (about 5% compared to 12% in Sweden overall).

Given the program's aim, it is central to analyze effects depending on student background. We distinguish four groups: (i) Swedish-background students, who have at least one Swedish-born parent,[24] and three groups of students with a foreign background: (ii) second-generation immigrants, whose parents were both born outside Sweden; and first-generation immigrants, who are further divided according to their time in the country when the schooling outcome is measured–(iii) non-recent immigrants (residence permit obtained more than four years earlier) and (iv) recent immigrants (residing in Sweden for at most four years). The latter group, recent immigrants, also includes asylum seekers. Analyzing effects on asylum seeking students is challenging. First, non-resident students cannot be followed across years in the data because they lack traceable personal identifiers. Second, they cannot be liked to parents, thus there is very little background information for this group, except sex and age. As a result we analyze asylum seekers together with recent immigrants when possible, but provide the main results also for asylum seekers separately in the Appendix.

[24] In our sample, 11.2 percent of the Swedish-background students have one parent born outside of Sweden.

The primary outcome of the study is students' school performance, measured by their test scores in the core subjects Swedish and Mathematics in grades 3, 6 and 9.[25] Test scores are normalized within grade and test year in the full population of students, excluding asylum seekers and immigrants.[26] In order to achieve greater statistical power, test scores are pooled across grades and averaged across subjects, but we also provide separate analyses for the main outcomes.

A challenge is that test score data is not available for the pandemic school years 2019/20 and 2020/21 when tests results were not collected. Moreover, in the school year 2017/18, 9th grade mathematics tests leaked beforehand on a large scale. In affected schools a spare-test was taken and results were not collected. To handle missing test score data during the pandemic and improve the coverage of treatment waves in each follow-up horizon, we aggregate the follow-up period into year 1-2, 3-4 and 5-7. An alternative approach is to use teacher assessment grades as an outcome, but these do not exist in 3rd grade and are arguably less objective than test scores (Vlachos, 2019). We thus analyze effects on test scores in the main analysis, but as a robustness test we also impute missing test scores with teacher assessment grades to investigate the sensitivity of results to missing data. This exercise reassuringly yields similar results as the main analysis (see Appendix Table A5 and commentary).

Links to tax and population registers, including family links and background information on parental income and education, migration status, country of origin and municipality of residence, allow us to predict the 9th grade test scores. The prediction, which serves as a composite measure of student family background, is based on student sex, age in months, birth order, measures of parental earnings and education, and immigration background of parents and students. The measure of predicted test scores is used to assess balance between treatment and control municipalities, but also to assess to what extent the program may have affected student composition in the treated municipalities, because of selectivity in test taking or student mobility.

We measure student in- and out-mobility explicitly at the municipal level since the program may impact the attractiveness of the municipality. We define a student as new to a municipality, i.e. inward mobility=1 in time t, if the student was enrolled in another municipality in the previous year (t-1). Similarly, we define outward mobility=1 in time t, if the student was enrolled in the municipality in the previous (t-1), but not in year t.

We construct several measures of school resources in order to assess the program's effects on organization and resource allocation. From the student register we construct indicators of access to tutoring in the home language, and an indicator for if the student has any other kind of extra support (unrelated to multi-lingual learning), such as adapted

[25] English is a core subject in grade 6 and 9, but we restrict our attention to the two subjects available for grades, 3, 6, and 9.

[26] Note that the pre-analysis plan states that asylum seekers and newly arrived immigrants should be excluded, but the purpose of excluding these groups is that test score should not trend upward for the native population due to changing student composition. Given that the 2015-immigrants are no longer newly arrived towards the end of our evaluation period, a stable composition was not achieved when only newly arrived migrants were excluded.

curriculum, special aid, etc. From grade and test score registers, we obtain indicators for if the student has a teacher assessed grade in a home language class, indicating that the student has participated in instruction to acquire a language, other than Swedish, spoken in the student's home environment. We also measure if the student had a grade in the subject "Swedish as a second language" rather than the subject Swedish, which is the regular Swedish course most Swedish background students take.

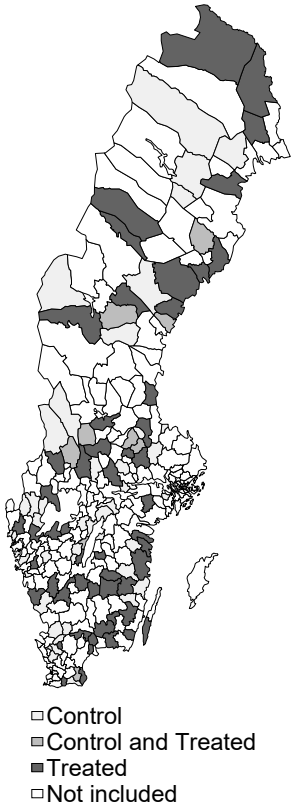
We measure school organization, grade and class assignment in several ways. A measure of class size is based on students home room class assignment in the student register, alternatively based on assignments in core subject based on test results registers. We measure individual student exposure to recent immigrants at the classroom level to assess any impact on classroom organization. We also measure student's grade for age to assess program effects on retainment, or placement of recent immigrants in classrooms with younger students. We also create school level measures of the presence of reception classes, defined as class rooms with more than 90 percent recent immigrants and indicators as to whether Swedish as a second language is taught in separate classrooms or integrated with ordinary Swedish instruction.

Based on the teacher register we construct measures of teaching resources and teacher qualifications at the school and municipal level. Unfortunately, data constraints prevent us from connecting teachers to grades or classes, but we construct measures at the school and municipality level. First we construct a measure of teacher-student ratio, which is measured a the number of full-time teachers per 100 students. Next, we measure the quality of teachers at the school and municipal level as the fraction of certified teachers and the mean years of experience.

3.4 Description of treatment and control municipalities

Figure 5 visualizes the geographic dispersion of municipalities randomized into treatment, control municipalities, and those controls that eventually become treated. As can be seen, the different types are evenly dispersed across the country.

Figure 5: Treatment and control municipalities



Note: The map shows the treatment status of municipalities for the period 2016-2019.

Table 1 presents pre-treatment characteristics of the treatment and control municipalities and a comparison to other municipalities for 2015, i.e. prior to the initiation of the targeted support program. Treatment and control municipalities, which were largely rural, are less populated than other municipalities. Consistent with the criteria for the need-based index, sampled municipalities are characterized by a high fraction of asylum seekers but comparably few second generation immigrants. While the targeted municipalities were disadvantaged in terms of parental background and test scores, their schools initially had higher fractions of certified teachers, more experienced teachers and also a higher teacher to student ratio. Test participation was similar across municipal groups, and as one would expect given the lower share of students with foreign background, fewer students have access to various forms of language oriented resources. The percent of students that go to schools with a special reception class is somewhat lower in targeted municipalities, and the share of students attending an independent school is much lower than the country average, reflecting that many targeted municipalities are rural.

.

Table 1: Pre-program characteristics of treatment and control municipalities, 2015

	Treated	Control	Never treated	All
Students (#)	3,835	2,566	21,074	17,633
Schools (#)	22.3	12.8	70.6	60.4
Teachers (#)	334.5	216.3	1618.1	1360.0
Boys (%)	49.3	49.3	50.0	49.9
Asylum seekers (%)	2.2	2.1	1.2	1.3
Recent immigrants (%)	7.9	7.0	6.1	6.3
Foreign born (%)	2.5	2.3	3.4	3.2
Second gen. (%)	5.4	5.3	12.7	11.3
Mother earnings	222,053	233,415	257,918	252,080
Father earnings	326,965	341,111	377,263	369,044
Years of schooling, mother	12.7	12.7	12.9	12.9
Years if schooling, father	12.1	12.1	12.5	12.4
Teachers, certified (%)	87.7	88.4	85.4	85.9
Teachers, years of experience	14.0	14.0	13.1	13.2
Teachers per 100 students	8.5	8.3	7.8	7.9
Test scores	-0.087	-0.106	-0.032	-0.043
Predicted test score	-0.112	-0.109	-0.054	-0.064
Test participation (%)	94.7	95.2	95.8	95.7
Swedish 2nd language (%)	8.8	7.9	11.9	11.3
Home language (%)	3.1	2.4	3.5	3.4
Tutoring (%)	1.9	1.8	1.8	1.8
Special aid (%)	6.8	6.7	6.0	6.1
Reception class in school (%)	19	20	18	18
Independent school (%)	5.4	5.2	15.2	13.3
Observations	135,774	92,851	894,241	1,110,647

Notes: The descriptive statistics are based on students in grades 0-9 in the student registry in year 2015.

In Table 2 we present individual characteristics for students in grades 3, 6 and 9, i.e. when we measure test scores, in 2015. Characteristics of the students in sampled municipalities are presented in column 2, and full population averages for students in the corresponding cohorts and year, in column 1. A comparison of the sample relative to the full population confirms the findings in Table 1. Importantly, comparing the Treated (column 3) and Controls (column 4) in Table 2 we see that students are similar in terms of measurable characteristics, as we would expect with randomization. In section 3.5, we will formally test that the treated and control municipalities are comparable the years before the intervention was implemented.

Table 2: Descriptive statistics of students in grades 3, 6, and 9 in 2015.

	(1) All	(2) Sample	(3) Treated	(4) Control
Girl	0.452	0.443	0.445	0.440
	(0.498)	(0.497)	(0.497)	(0.496)
Swedish background	0.743	0.787	0.784	0.791
	(0.437)	(0.409)	(0.411)	(0.406)
Second generation immigrant	0.100	0.046	0.046	0.046
	(0.300)	(0.209)	(0.209)	(0.209)
Non-recent immigrant	0.039	0.030	0.031	0.029
	(0.193)	(0.170)	(0.172)	(0.168)
Recent immigrant	0.106	0.117	0.120	0.114
	(0.307)	(0.321)	(0.324)	(0.317)
Asylum seeker	0.023	0.031	0.031	0.031
	(0.149)	(0.173)	(0.173)	(0.173)
Age at immigration	0.773	0.785	0.810	0.750
	(2.692)	(2.778)	(2.812)	(2.731)
Earnings, mother	247,844	221,800	218,684	2260,36
	(199,631)	(162,915)	(160,203)	(166,441)
Earnings, father	356,901	318,724	314,890	323,934
	(333,531)	(234,986)	(228,762)	(243,098)
Years of education, mother	12.793	12.643	12.653	12.629
	(2.319)	(2.191)	(2.208)	(2.168)
Years if education, father	12.365	12.033	12.042	12.021
	(2.309)	(2.091)	(2.093)	(2.090)
Test score	-0.043	-0.128	-0.113	-0.149
	(0.906)	(0.914)	(0.910)	(0.919)
Predicted test score	-0.064	-0.456	-0.456	-0.456
	(0.369)	(0.764)	(0.764)	(0.764)
Observations	342082	73841	42545	31296

Notes: The descriptive statistics (sd in parentheses) are based on the students present in test score registers in grade 3, 6 or 9, in year 2015. Swedish background students have at least one Swedish-born parent, Second generation immigrant students' both parents are foreign born. Non-recent immigrants have immigrated more that 4 years ago, recent immigrants at most 4 years ago (includes asylum seekers). Age at immigration is reported for immigrants students. Earnings include zeros, years of education pertain to parents with non-missing data.

3.5 Identification

Identification of program effects relies on the randomized staggered roll out of the program. We compare the outcomes of students in treated and control municipalities, controlling for randomization pair (and consequently wave). Treatment is assigned based on the timing of the randomization for each wave. Outcomes measures after this date are potentially affected by the program.

Because the program was rolled out just before the Covid19-pandemic, the follow up period is characterized by missing data for certain years. We therefore aggregate the follow up period in two year intervals and exclude the pandemic years from the analysis.

3.5.1 Empirical specification

Pair-wise randomized treatment allows us to evaluate effects from the intervention using a standard regression model, within each pair. We estimate the following main specification:

$$Y_{imgt} = \alpha + \beta Treated_m + \gamma Pair_m + \lambda Grade_g + \theta X'_m + \delta_t + \varepsilon_{imgt} \quad (1)$$

Y_{imgt} is the outcome for individual i, residing in municipality m, attending grade g, in year t. $Treated_m$ is an indicator variable, taking the value 1 if the municipality was randomized to treatment, thus β reflects the treatment effect of interest. $Pair_m$ capture randomization-pair fixed effects and $Grade_g$ are grade fixed effects. X_m is a vector of municipal controls, and δ_t denotes year fixed effects. Standard errors are robust and clustered at the municipal level.

The vector of municipal controls X_m includes the needs-rank, measured in the randomization round of each pair, p, and municipality controls averaged in the two years before entering the program: averages of test scores in mathematics and Swedish respectively, teacher-student ratio, teacher experience and share certified teachers, as presented in Tables 3, A6 and A7. The municipal controls also include the shares of students with Swedish background, second generation immigrants and recent migrant students, as well as predicted test scores. To ensure that the estimated effects are not driven by pre-existing differences across student groups, all specifications include also predicted test scores and average pre-intervention test scores specifically for second generation immigrants, non-recent and recent immigrants. We also include indicators of belonging to either of these student groups. Since treatment is at the municipal level, we weight the regression to give equal weight to each randomization pair. We also provide results giving equal weight to each student, which gives very similar results (see Appendix Tables A8, A9 and A10 for average test score, Swedish and mathematics respectively).

3.5.2 Balance

Identification of causal effects of the targeted support program relies on successful randomization of adjacently ranked municipalities. We examine this using individual data in

Table 3, and for municipality averages in Appendix Tables A6 and A7.

In Table 3, we test if our main outcome, test scores, is well balanced the two years before the intervention. We assess the average test score in Panel A, as well as mathematics and Swedish separately in Panels B and C. In Panel D, we evaluate the predicted test score, the composite measure of student- and parental characteristics (see description of variables in Table 2).

Overall, test scores are well balanced. However, evaluating the differences between treated and controls for different groups of students, we find some imbalance for students with foreign background. In treated municipalities before the intervention, recent immigrants had lower test scores both on average and across subjects, and second-generation immigrant students had lower test scores in mathematics. This imbalance is also reflected in the predicted test scores. The imbalance detected in Table 3 suggests that the students with foreign background are somewhat negatively selected in the treated municipalities, which might lead us to underestimate the effects for these groups of students. As a remedy, we include these pre-determined municipality and group-specific test scores and predicted test scores in all specifications.

Table 3: Estimation of pre-intervention outcomes

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: Test score					
Treated	-0.002	0.010	-0.050	-0.015	-0.098**
	(0.018)	(0.017)	(0.035)	(0.043)	(0.038)
Obs	129,107	113,063	6,301	3,980	5,763
Control mean	-0.112	-0.074	-0.248	-0.301	-0.607
Panel B: Math score					
Treated	-0.016	-0.004	-0.104**	0.002	-0.130***
	(0.020)	(0.019)	(0.043)	(0.052)	(0.041)
Obs	127,022	111,273	6,210	3,881	5,658
Control mean	-0.086	-0.040	-0.322	-0.405	-0.561
Panel C: Swedish score					
Treated	0.010	0.024	-0.006	-0.025	-0.108**
	(0.018)	(0.017)	(0.043)	(0.046)	(0.044)
Obs	124,826	110,087	6,075	3,815	4,849
Control mean	-0.116	-0.093	-0.153	-0.163	-0.598
Panel D: Predicted test score					
Treated	-0.016*	-0.004	-0.038**	-0.062***	-0.002
	(0.009)	(0.008)	(0.017)	(0.019)	(0.010)
Obs	476,320	395,668	26,227	11,196	31,982
Control mean	-0.892	-0.834	-1.134	-0.807	-1.204

Notes: Measured in the 2 years before sampling in the wave. Estimations include the full set of controls as specified in Equation 1.

In Appendix Tables A6 and A7, we further test if municipality level averages of predetermined characteristics differ between treated and control municipalities. The mea-

sures of municipality characteristics suggests that the randomization was indeed successful overall. The small and insignificant difference in Ranking score in column 1 of Table A6 means that there is no difference in the composite measure of need of support, based on inflow of migrants, previous experience characteristics of the recently arrived and asylum seeking students, as assessed by the SNAE. Nor are there any significant differences in teacher characteristics (columns 2-4 of Table A6). Treated and control municipalities are also similar with respect to composition of students with different immigrant background (Appendix Table A7). The one significant exception at the municipal level is the number of students in column 5 of Table A6, indicating that municipalities randomized into treatment are larger–about 500 students more–than the municipalities in the control group.

The presence of pre-treatment differences detected in this section, albeit generally small, motivates us to include controls for predetermined characteristics at the municipal level in our estimations, in addition to the measures of test scores and predicted test scores specific to different student groups. Testing the joint significance of the control variables, we find the overall differences to be significant once number of students is included, but not otherwise. We verify that our model successfully accounts for pre-treatment imbalances by estimating the full dynamics of test score effects pre- and post treatment in Section 4.1.1. We also assess to what extent our main findings can be interpreted as causal effects on learning, as opposed to driven by post-treatment compositional changes, by including student individual characteristics as controls in a sensitivity analysis (Appendix Tables A11- A13).

4 Results

We evaluate the effects of the targeted support program on student performance, using test scores in the core subjects Swedish and mathematics for students of different migration background.[27] We present effects for the full seven-year follow-up period, as well as separately for the implementation phase and the post-program years. Evaluating effects for different time horizons enables us to assess if it takes time for effects to materialize–as has been shown in other school improvement programs (Sun, Kennedy, and Loeb, 2021)–and if there is any evidence that the program has long lasting effects (Sims et al., 2025). We investigate program impact on pedagogical resources, class room organization and student mobility in search for mechanisms through which student outcomes were affected. Finally we describe the program content in different types of municipalities, and how it relates to program effects on test scores and resources.

[27] Test scores are pooled across grades 3, 6 and 9.

4.1 Main results: Test scores

Table 4 presents the effects of the support program on the average of core subject test scores for different follow-up horizons.[28] Over the full 7-year follow-up period, there is a significant positive overall effect of 0.021 sd (Panel A). We find significant improvements for students with Swedish background and second generation immigrant students of 0.019 sd and 0.066 sd respectively. It is worth noting that the point estimates for the immigrant student groups are relatively large, 0.023 and 0.036 sd, exceeding that for students with a Swedish background, although imprecisely estimated.

Evaluating different time periods separately in Panels B-D, we find that the test score improvements arise at different times across student groups. Already during the implementation period, test scores improved for Swedish background (albeit marginally significant) and second generation immigrant students. In the two post-intervention periods, all point estimates are positive and students with foreign background exhibit especially large gains. Second generation immigrants' test scores improved substantially, 0.128 sd in years 3-4 and 0.073 sd in years 5-7, while non-recent immigrants test scores improved even more in the early post implementation period, 0.315 sd, and 0.073 sd in years 5-7. The improvements for recent immigrants are only marginally significant in years 3-4, but 0.179 sd in the late post-intervention period. The estimates for Swedish background students are smaller and insignificant or marginally significant.[29]

[28] The estimates are visualized in Appendix Figure B2.

[29] Appendix Table A14 shows that the overall test score gains for the full follow-up period are present only in grade 6 and in grade 9. There is also heterogeneity across student groups and grade. Differences may reflect length of exposure to the intervention or variations in its impact–either because school stage itself matters or because the content of the intervention differed. It could also be due to differences in the time elapsed since the intervention.

Table 4: Effects of targeted support on average test scores (Swedish and mathematics) in grades 3, 6 and 9

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.021**	0.019**	0.066**	0.023	0.036
	(0.008)	(0.008)	(0.027)	(0.028)	(0.024)
Obs	209,081	171,022	13,605	11,511	12,943
Control mean	-0.135	-0.071	-0.299	-0.340	-0.656
Panel B: Implementation period					
Years 1-2	0.012	0.018*	0.110***	0.013	-0.019
	(0.008)	(0.010)	(0.040)	(0.034)	(0.031)
Obs	91,833	76,079	5,054	3,210	7,490
Control mean	-0.158	-0.090	-0.320	-0.354	-0.686
Panel C: Early post intervention period					
Years 3-4	0.052*	0.032	0.128***	0.315***	0.084*
	(0.026)	(0.025)	(0.041)	(0.064)	(0.046)
Obs	39,880	32,633	2,852	2,110	2,285
Control mean	-0.096	-0.027	-0.297	-0.318	-0.676
Panel D: Late post intervention period					
Years 5-7	0.035***	0.018*	0.073**	0.073**	0.179***
	(0.011)	(0.011)	(0.032)	(0.029)	(0.030)
Obs	77,368	62,310	5,699	6,191	3,168
Control mean	-0.127	-0.071	-0.282	-0.341	-0.565

Notes: Average test scores is the average of Math and Swedish in grades 3, 6 and 9. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Next, we analyze the results for Swedish and mathematics separately (see the estimates visualized in Appendix Figure B3). This is motivated by the fact that the program aimed to improve schooling for immigrant students and Swedish learners, emphasizing professional development in language awareness and integrating language development with subject learning in all subjects.[30]

The results for test scores in Swedish show no significant improvement for any group of students in the full 1-7 follow-up period (Table 5, Panel A). However, both non-recent and recent immigrants show substantial improvements in the early post-implementation period, 0.326 sd and 0.214 sd respectively. In the late post-period, the effects fade, and only recent immigrants make significant improvements in Swedish test scores.

Turning to Mathematics in Table 6, we find that the intervention was rather successful in improving mathematics performance, especially in the late post-period. Panel A shows that the program led to overall improved mathematics test scores of 0.032 sd. For students with Swedish background, the increase was 0.029 sd. The overall gains for second generation and non-recent immigrant students are larger, 0.103 sd and 0.061 sd respectively. Recent immigrant students, however, did not gain significantly overall. Panels B-D show that mathematics test scores improved already during the implementation period, except

[30]This split is further motivated by evidence of spill-overs across subjects, see for instance Machin and McNally (2008) who find positive spill overs on mathematics results from the "literacy hour".

for recent immigrants for whom there is a negative, but not significant point estimate. In year 3-4, there is a positive effect for second generation immigrants (0.184 sd) and an even larger impact for non-recent immigrants (0.308 sd). All student groups show substantial improvements in mathematics in years 5-7.

Table 5: Effects of targeted support on test scores in Swedish in grades 3, 6 and 9

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.011	0.011	0.040	-0.012	0.041
	(0.008)	(0.008)	(0.031)	(0.034)	(0.032)
Obs	202,372	166,388	13,315	11,198	11,471
Control mean	-0.122	-0.079	-0.184	-0.200	-0.634
Panel B: Implementation period					
Years 1-2	0.007	0.010	0.109**	-0.014	0.029
	(0.010)	(0.010)	(0.045)	(0.047)	(0.045)
Obs	88,752	74,108	4,938	3,119	6,587
Control mean	-0.148	-0.093	-0.214	-0.225	-0.722
Panel C: Early post intervention period					
Years 3-4	0.075***	0.049***	0.071	0.326***	0.214***
	(0.019)	(0.016)	(0.064)	(0.097)	(0.051)
Obs	38,658	31,764	2,789	2,059	2,046
Control mean	-0.098	-0.051	-0.188	-0.217	-0.635
Panel D: Late post intervention period					
Years 5-7	-0.004	-0.012	0.001	0.015	0.094***
	(0.009)	(0.009)	(0.034)	(0.030)	(0.034)
Obs	74,962	60,516	5,588	6,020	2,838
Control mean	-0.103	-0.077	-0.156	-0.182	-0.420

Notes: Effects of targeted support on test scores in Swedish in grades 3, 6 and 9. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table 6: Effects of targeted support on test scores in mathematics in grades 3,6 and 9

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.032***	0.029***	0.103***	0.061*	0.024
	(0.011)	(0.010)	(0.033)	(0.032)	(0.028)
Obs	199,602	163,463	13,091	10,937	12,111
Control mean	-0.115	-0.036	-0.395	-0.451	-0.619
Panel B: Implementation period					
Years 1-2	0.021*	0.027**	0.120**	0.047	-0.039
	(0.012)	(0.012)	(0.046)	(0.041)	(0.033)
Obs	85,728	71,216	4,755	2,903	6,854
Control mean	-0.137	-0.062	-0.408	-0.488	-0.610
Panel C: Early post intervention period					
Years 3-4	0.035	0.022	0.184***	0.308***	-0.085
	(0.036)	(0.035)	(0.051)	(0.056)	(0.065)
Obs	38,523	31,525	2,776	2,023	2,199
Control mean	-0.064	0.021	-0.380	-0.387	-0.649
Panel D: Late post intervention period					
Years 5-7	0.070***	0.048***	0.171***	0.101***	0.226***
	(0.018)	(0.016)	(0.036)	(0.038)	(0.034)
Obs	75,351	60,722	5,560	6,011	3,058
Control mean	-0.118	-0.037	-0.393	-0.456	-0.618

Notes: Effects of targeted support on test scores in mathematics in grades 3, 6 and 9. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

4.1.1 Interpretation of results

So far, we have presented evidence that participation in the targeted support program improved test scores in core subjects. While the effects on Swedish test scores were strongest in the early post-implementation period and then faded, the positive effects in mathematics emerged during implementation and were particularly pronounced in years 5–7. The program has positive effects, especially for students with an immigrant background, with earlier and more persistent improvements among second-generation and longer-established immigrant students, while recently arrived immigrants benefit mainly in the longer run. Swedish background students experience more modest gains, but there is no sign the focus on integration and multilingual learners hurt them. Taken together, the results show a clear pattern: the program's effects appear early, during the implementation, for Swedish background and second-generation students, but later—and then more strongly—for immigrant students.

The positive program effects overall, and in mathematics, are robust to multiple-hypothesis correction for all students. The mathematics gains are significant for Swedish background and second generation immigrants, while the effects on the average score is significant only for second generation immigrants (see Appendix Table A15). Nevertheless, when estimating an interaction model instead of split-sample specifications, we find

a significant overall effect on test scores of 0.026 sd, and 0.037 sd on the mathematics test score, but heterogeneous effects by migration background are not statistically significant (see Appendix Table A16). Thus, we conclude that the targeted program was successful overall–and results suggest especially so for second-generation students.

Although the the program's aim was to improve education for recent immigrant students, Swedish language learners and multilingual students, we find some smaller positive effects also for Swedish background students, especially in mathematics. The estimated effects for foreign born students are less precise, but the evidence suggests that they gained from the intervention in the longer term in both mathematics and Swedish.[31] Trying to understand these patterns is important and we will come back to this when we analyze program impacts on teaching resources and organization (Section 4.3.)

In order to interpret the results as causal effects of the intervention on student learning, we first need to investigate the extent to which the program might have affected student composition. Student composition might have been altered if the propensity to participate in testing changed in treated municipalities, or as a result of changes in inward and outward mobility. In Appendix Table A18 we explore differences in test taking and student composition between treated and control municipalities during the intervention, and post-intervention in years 3-4, and years 5-7.

We find no overall effect on test taking in treated municipalities during the intervention, but some indications of reduced test taking among Swedish born students and increased test taking among foreign born students in the early post-implementation period. The magnitudes are however very small. There are some small changes in test taking also in the longer run. In Panel B, we explore if the composition of students changes in treated municipalities by estimating the model on student's predicted test score, which is a summary measure of student background characteristics. While there is some indication that non-recent immigrants become slightly favorably selected during the implementation period, also reflected in the predicted test scores among test takers (panel C), the opposite is true for second generation immigrant students. In the post-implementation period, there are no signs of changes in the student composition.

We also verify that the model successfully achieves balance between treated and control municipalities, not only in the years just prior to the intervention, but also before that. We present the full dynamics of core subject test scores, pre- and post-intervention in Appendix Figure B4. The model is successful, although there is some evidence, albeit not significant, of worse test scores for recent migrant students in treated municipalities in the early pre-period.

We take a further step to assess effects on learning, by accounting also for the detected minor changes in student composition. Tables A11, A12 and A13 present the effects on test scores when including the full set of individual student controls. The resulting estimates are very similar to the main results presented in Tables 4, 5 and 6, although the

[31] Recent immigrants include asylum seekers. Since this group is small and cannot be linked across registers we have chosen not to provide results separately. Evaluating the effects on asylum seekers separately in a tentative analysis, suggests overall positive effects on math score (Appendix Table A17).

positive effects on recent immigrants in the full seven year follow up now turn marginally significant in both Swedish and mathematics.

We can thus conclude that compositional changes are not driving the positive impact on test scores, yet changing migration flows or mobility patterns are possible consequences of the intervention. When we investigate this in more detail, we see a reduction in the number of students in the treated municipalities, and an increase in the number of asylum seeking students during the implementation period. Moreover, there is an indication that recent migrants were less likely to move from treated municipalities during the implementation and also less likely to move to the treated municipalities in the post-implementation period, while second generation immigrants were more likely to move away post-implementation (See Appendix Tables A19 and A20). Both the overall fraction of students moving and their characteristics are essentially unaffected, hence declining number of students likely reflects time-varying differences in cohort size, consistent with the imbalance in detected Table 3, rather than effects of the intervention.

To conclude this section, the estimated positive effects on test scores are consistent with real improvements in the school performance in treated municipalities. Moreover, the estimated model successfully shows that there is balance between treated and control municipalities also 3–4 years prior to the intervention, supporting a causal interpretation of results. Although some compositional changes are observed post-treatment, the analysis provides no evidence that these drive the estimated effects on student outcomes. On the contrary, accounting for these changes, the positive effects on recent migrant students are strengthened.

4.1.2 Costs and benefits

We provide a back of the envelope calculation of whether the intervention can be motivated from a cost and benefit perspective. Participation in the program generated a significant positive effect on core subject test scores over the 7-year follow up of 0.021 sd, and test score gains in mathematics of 0.032 sd. Making use of the estimations of returns to mathematics skills and life-time earnings based on Swedish data presented in Grönqvist, Öckert, and Rosenqvist (2025), we compare benefits and costs of the intervention.

Beginning with benefits, first, Grönqvist, Öckert, and Rosenqvist (2025) estimate that a one standard deviation improvement in mathematics skills increases life time earnings by 9 percent. Second, they estimate the net present value of average real gross life-cycle earnings,[32] to be about SEK 9,700,000 (or €940,000) in 2020 prices. Hence, a standard deviation increase in mathematics skills raises lifetime earnings by SEK 870,000. Consequently, our estimate of 0.032 sd, implies a gain of SEK 27 840 per student.

Next, we estimate the costs of the program. We do this based on the amount of resources transferred to the participating municipalities during the implementation years, according to the municipalities' final reports to the SNEA. These figures are presented in

[32]The net present value includes employer contributions and is discounted at 3 percent to age 16, for men born 1952–53.

Table A21. The cost per unique student enrolled in the municipalities during the implementation years was SEK 2,152. Although, there were some costs also at the SNAE that are not included in these calculations, it is safe to say that the program easily passes a cost-benefit test, giving almost 13 in return for every krona spent.[33]

4.2 Heterogeneous effects

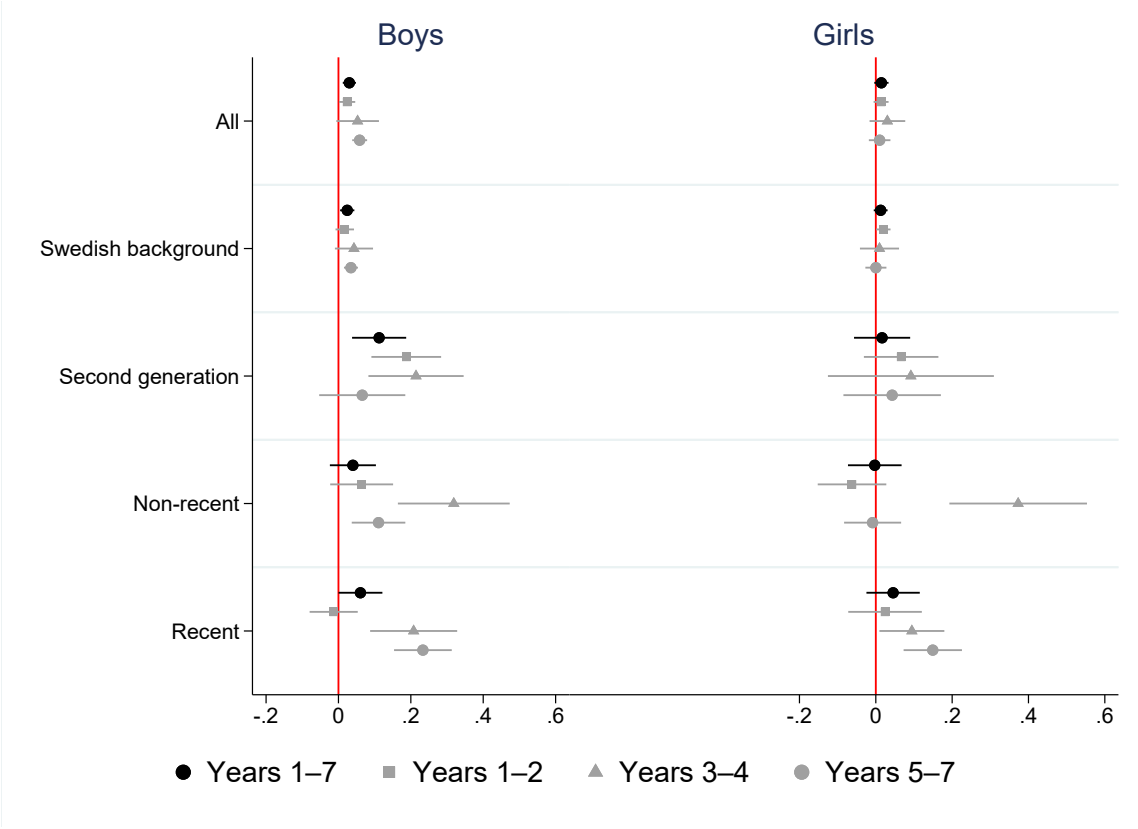
While we have demonstrated that effects of the targeted support program differ by migration background of the students, it is also of interest to explore differences by gender. Research shows that children may respond differently to human capital investments and school interventions depending on gender. Boys are often found to more sensitive to stressful environments and school quality, and they generally perform worse in school (Autor et al., 2019; Bertrand and Pan, 2013) while girls have an advantage in language development (Fort, Ichino, and Zanella, 2020).

The black estimates in Figure 6 indicates that boys benefited on average when considering the full 7-year period, whereas girls did not. While the difference is insignificant in an interaction model (see Appendix Table A22), we find that boys, regardless of migration background, benefited from the program. Most boys gained both during and post-implementation, while girls typically did not benefit. The overall effect of 0.030 sd for boys reduces the test score gap between boys and girls by some 10 percent. The effect is larger, 0.112 sd, among second generation immigrants, substantially reducing large test score gaps to both Swedish background boys and second generation immigrant girls by some 50-60 percent.[34] Immigrant boys see large gains in the post-implementation years. For immigrant girls, there is also some evidence of positive effects, in particular in years 3 and 4 (see Table A23).

[33] Another estimate of the total costs is based on the initial support program budget, which was SEK M450. If we divide this number by the number of students in our test score data during the 7-year follow up period, 205,387 (Table 4), we again arrive at a per student cost of around SEK 2000.

[34] When comparing the 0.112 point estimate to the control means for the respective groups.

Figure 6: Effects of targeted support on test scores by student gender



Notes: Main regression separately by gender, boys (left) and girls (right). Test scores in total (black), years 1-2 (gray squares), 3-4 (gray triangles), and 5-7 (gray circles). Estimations include the full set of controls as specified in Equation 1, excluding years 2019 and 2020.

We also explore heterogeneous program effects by estimating the effect of participation on the probability of achieving test scores above the 20th, 40th, 60th and 80th percentile of the test score distribution. The analysis shows that the gains are strongest among students at the lower end of the test score distribution. As shown in Appendix Figure B5, the overall positive effect is concentrated among low-performing students during the implementation period, and the magnitude of this improvement is sufficient for it to impact the full follow-up period, although the post-implementation period shows gains also in the middle of the distribution. Separating by Swedish and foreign background students in Figures B6-B7, the overall pattern is reflected among students with a Swedish background, who also display some gains in the middle of the distribution during the post-implementation period. For students with foreign background, however, the distribution is flatter, with less gains during the implementation period, and gains at the lower end in the post-implementation period. Finally, the distributional results by gender indicate that the initial gains at the lower end are present among boys, whereas the girls who benefit are found higher up in the distribution (Figures B8-B9).

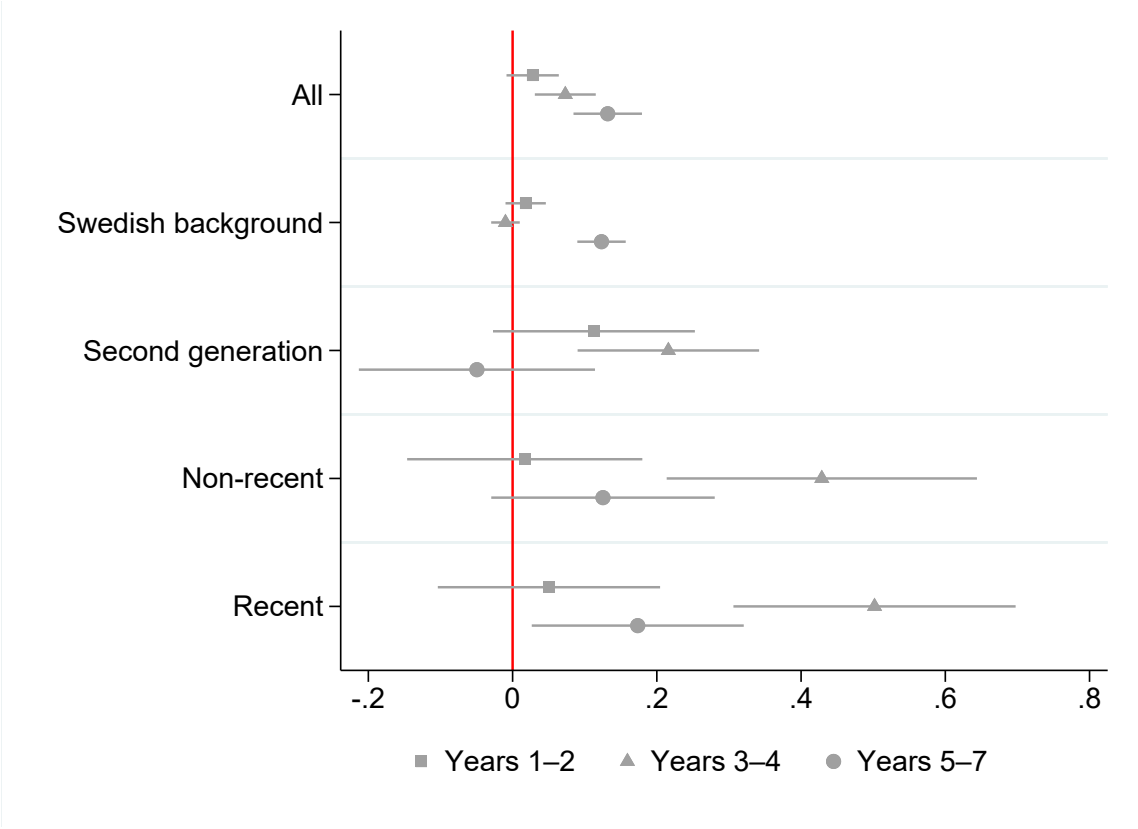
4.3 Mechanisms: Teaching resources and classroom organization

In order to understand how the intervention impacted students, we investigate changes in their access to teaching resources. Because the intervention was intended to facilitate integration of recent migrant students into schools and to improve the quality of education for Swedish learners, we are particularly interested in potential effects on student's access to various kinds of language-related pedagogical resources or special aid. We also investigate in what way the program affected school and classroom organization and if the quantity or quality of teaching staff was affected. It should to be stressed, however, that the program did not provide funding to employ more teachers or provide students with extra resources. Thus, any effects along these dimensions are reflecting shifts in priorities – allocation and organization of given resources – resulting from insights made as a result of professional development and organizational support. In addition, positive test score gains can stem from improved teaching quality resulting from professional development, rather than from reallocation of resources and organizational changes.

4.3.1 Pedagogical resources

First, we estimate program effects on student access to language enhancing pedagogical resources. We construct a standardized summary index, combining assignment to "Swedish as a second language" classes (as opposed to ordinary Swedish classes), home language instruction and receiving tutoring in the home language. Results presented in Figure 7 (see also Table A24 for details), show some indication of improved access in treated municipalities for second generation immigrant students already during the implementation period; the estimate is large but insignificant. In the post-period, access improves overall. Foreign background students gain more access to language enhancing resource in particular in the first years after implementation. In the longer term, there is increased access also for Swedish backgrounds students.

Figure 7: Effects on standardized language enhancing resources by migration status.



Notes: Language resources takes the values one for students receiving at least one of the following: Swedish as a second language, home language classes, tutoring in mother tongue. Resources in years 1-2 (squares), 3-4 (triangles), and 5-7 (circles). The measure is standardized. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Taking a closer look at different aspects of language enhancing resources in Appendix Table A25 and Figure B10, we can see that home language instruction increased for non-recent immigrants already during the implementation period. It is also revealed that the insignificant increase in access for second generation immigrant students during the implementation period was driven by assignment to Swedish as a second language, although the estimate is again insignificant. In the early post-implementation period, assignment to Swedish as a second language and participation in home language instruction increases for foreign background students (although not all estimates are significant), and immigrant students are more likely to have tutoring in their home language. Swedish background students instead got more of other types of special aid during this time period. In the longer term, assignment to Swedish as a second language is less common in treated municipalities, while immigrant students have higher access to tutoring in their home language. Access to home language instruction also remains higher in years 5-7, and interestingly also Swedish background students are more likely to participate.[35] Table A25

[35]This may at first seem surprising, but access to such classes should be provided if the student speaks a foreign language in the home, which could be the case if the student has a foreign born parent or immigrated grandparents.

shows that the effects on other, non-language related forms of special aid, were limited throughout, except for the increase for Swedish background students, implying that these were not crowded out by language oriented support.[36]

Further analysis suggests that there was an advantage in scaling up where a basic structure was already in place. The expansion of language-oriented pedagogical resources, in particular Swedish as a second language and home language instruction, was more rapid in municipalities where initial language resources were comparatively high before the intervention (to the left in Appendix Figure B11).

Connecting the evidence on language oriented resources to the effects on test scores for different groups of students, we infer that improved access to pedagogical resources matters. Figure B10 shows that foreign background students gained access to language enhancing resources in the post implementation period, relative to control municipalities, which may have contributed to the improved test scores in the post period. Since the test score improvements are present also in mathematics (Table 6), we can rule out that the improvement merely reflects more lenient grading and assessment of students taking Swedish as a second language. Instead, it suggests meaningful improvements in learning.[37]

4.3.2 Classroom and school organization

Second, we investigate program effects on the way students are assigned to classrooms, and other aspects of school organization relating to the reception and integration of migrant students.

We find that organization of the classroom changed in treated municipalities compared to control municipalities. Panel A in Figure 8 shows that there were minor increases in class size in schools attended by foreign background students during the implementation period, but that class size increased for all student groups in the post-implementation periods.[38] Panel B in Figure 8 also reveals that second generation immigrants had higher shares of recent immigrants in their classroom during the implementation period in treated municipalities than in control municipalities. Meanwhile, students with Swedish background and non-recent immigrant students had higher shares of recent immigrant classmates in the post-implementation period. The pattern possibly reflects initially segregated classrooms with respect to students' immigration background in the treated municipalities.[39]

In support of the interpretation of initial segregation, Panel C in Appendix Table A27 shows that during the implementation, immigrant students in treated municipalities were

[36] We have verified that the increase in special aid is not driven by more students being exempt from some subjects.

[37] Moreover, we find improved learning outcomes regardless of students' language background, suggesting that the program successfully contributed to improvements for all multilingual learners–not just those sharing language background with the recent immigrant students (see Appendix Table A26).

[38] This contrasts findings in Getik, Sjögren, and Sundberg (2024) of class size reductions in response to migration inflow during the 2015 refugee crisis.

[39] The corresponding estimates are presented in Panel A and B of Appendix Table A27.

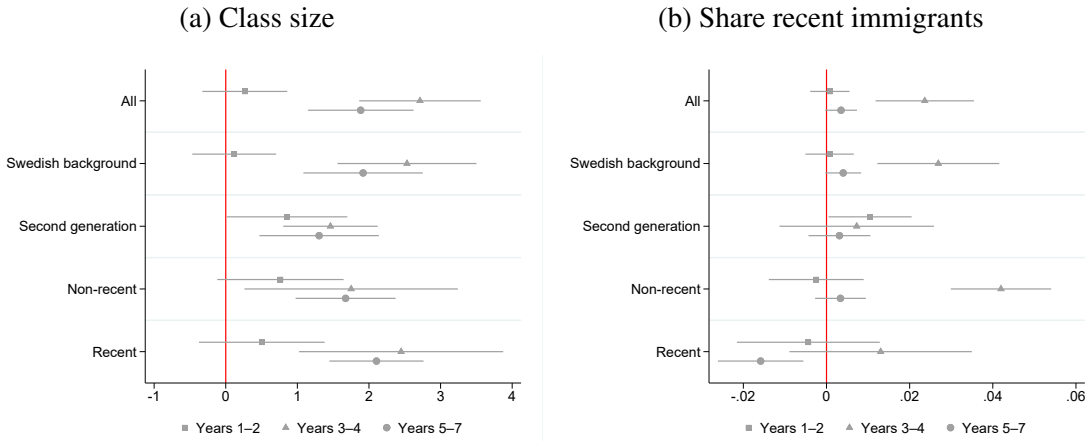
more likely to attend schools in which there was a reception class. During this period, the control mean reveals that reception classes were rather common in the schools of these students: almost a quarter of immigrant students in control municipalities attended schools where these were present. Placing recent migrant students into separate classes might initially have been seen as a way to focus the additional resources available during the intervention. In the post-implementation period, however, these special reception classes are more rare and even less prominent in treated municipalities than in the control municipalities. An interpretation is that treated municipalities in the longer term became more successful in integrating new migrants into regular classes, which is also consistent with the overall increasing shares of recent immigrants in the classroom in the post-period. Notably, there is variation in the incidence of reception classes across the student groups. In the schools where most second generation immigrant students attended, reception classes were less common and did not increase during the implementation period, instead they were exposed to more recent migrants in their class rooms. It is thus possible that this exposed them to more language enhancing teaching methods, contributing to their test score gains. Phasing out of introduction classes, which are typically small, may also have contributed to the increases in class size in the post-implementation period.

Panel D of Appendix Table A27 shows that during the implementation years, there was no difference between treated and control municipalities in the practice of teaching Swedish as a second language in a separate classroom. Instead, this increased in the post-period when most students in treated municipalities had an increased probability of attending a school where this was the practice, rather than teaching the two subjects in parallel in the same classroom. Thus, it seems that when reception classes were phased out in treated municipalities, they instead adopted the practice of teaching Swedish as a second language in separate classroom.

We also find that students with foreign background were more likely to attend an age appropriate grade in the post-intervention period (see Panel E of Appendix Table A27), but no difference during the implementation years, when segregating practices were more common.[40] As one would expect, the means of the dependent variable indicate that students with a foreign background are less likely to attend age-appropriate grades compared to those with a Swedish background. It is however worth noting that this gap decreases over time, and more so in treated municipalities. In treated municipalities, the practice of placing foreign-background students in age-appropriate classrooms remains more common even in years 5–7.

[40]There is also a statistically significant negative effect for Swedish-background students in years 3–4, but the effect is economically negligible relative to the group mean.

Figure 8: Effects on classroom organization by migration status.



Notes: Years 1-2 (squares), 3-4 (triangles), and 5-7 (circles). Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

To summarize, the evidence points to that the intervention had some influence on how municipalities organized classrooms and managed the reception of recent migrant students. While using more reception classes during the implementation period, the treated municipalities were more likely to phase these out and integrate recent migrant students in regular classrooms over time. In line with this development, they also became more likely to place students with foreign background in age-appropriate grades. At the same time, they maintained the practice of teaching Swedish as a second language in separate classrooms and although access to language enhancing resources increases, there is no sign that these crowded out other forms of special aid.

4.3.3 Teachers

Third, we investigate effects of participation in the support program on the quantity and quality of teaching resources. However, because the program focused on professional development and support of school management, we should perhaps not expect large effects along these dimensions. Data allows us to tie teachers to schools, but not to individual classrooms, hence any differential effects by students' migration background reflects segregation across schools, not within schools.

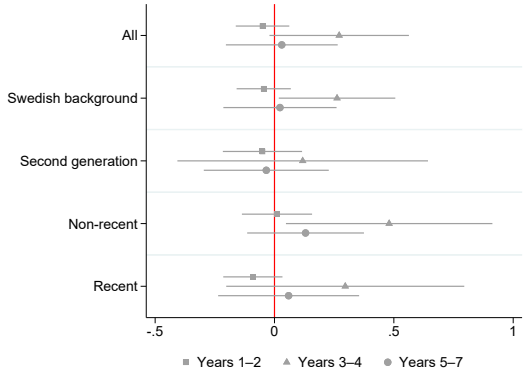
Panel A of Figure 9 reveals no effects on the teacher-student ratio in treated municipalities during the implementation period. There is a tendency of more teachers per student in the early post-period overall, and we find significant increases for students with Swedish background and non-recent immigrant students. However, the estimates are small relative to the control mean of 8.2 teachers per 100 students (see Panel A in Appendix Table A28).

Turning to the quality of the teachers, we do not find that the increasing number of teachers come at the expense of their credentials. Panel B of Figure 9 shows that the share of certified teachers increased during the implementation period, and there are no significant reductions at any time, for any group of students. Instead, the second gen-

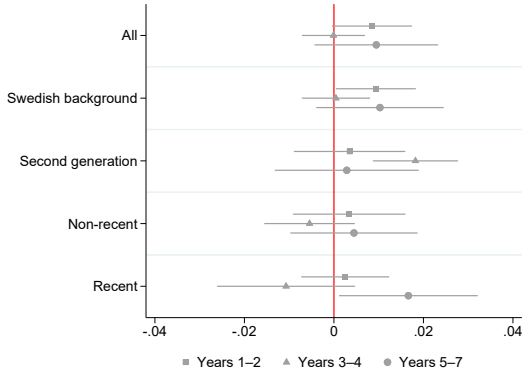
eration immigrants experience an increase in the early post-period. Average experience presented in Panel C of Figure 9 also increases during the intervention, but students with foreign background have slightly less experienced teachers in the post-period, at most half a year less than the control average around 13.5 years.

Figure 9: Effects on teacher student ratios and teacher, by migration status. qualifications

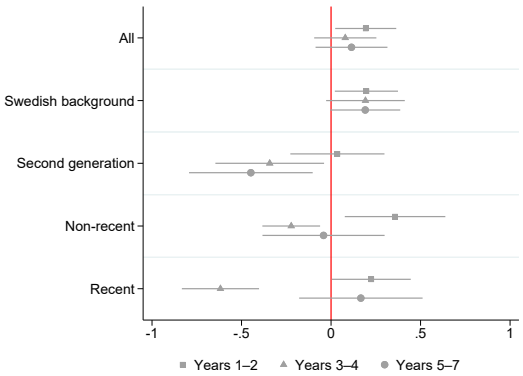
(a) Teacher-student ratio



(b) Share certified teachers



(c) Average experience



Notes: Years 1-2 (squares), year 3-4 (triangles), year 5-7 (circles). Teacher-student ratio (Panel A) is the number of teachers (full time) per 100 students. Certified (Panel B) is an indicator of teachers having the appropriate education, Average experience (Panel C) is the number of years being a teacher. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

4.4 Program content and spending

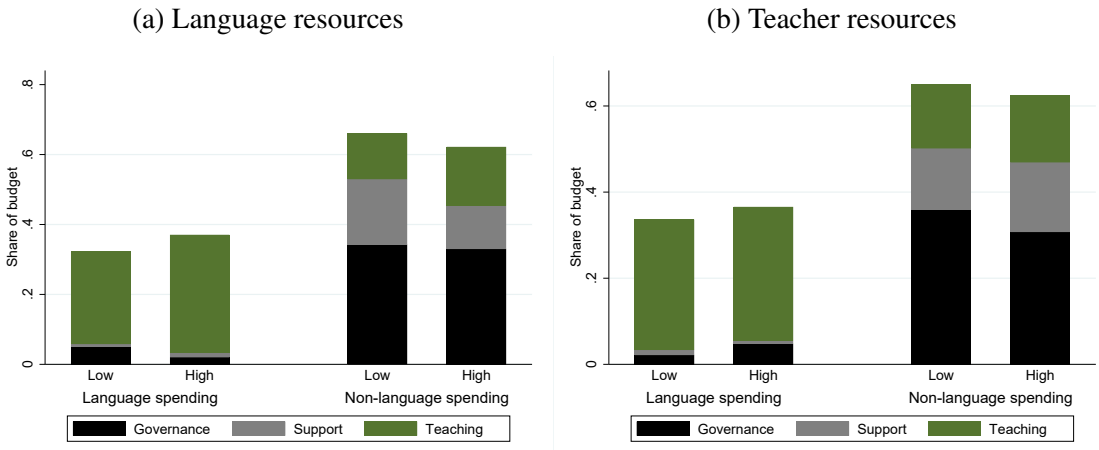
Next, we want to shed some light on the content of the intervention. Since the program was tailored to the specific needs of each participating municipality, it is not possible to make causal inference linking program content to outcomes. Nevertheless, we will present a description of the content in different types of municipalities, and how it relates to the effects found in terms of test scores and resources.

We first investigate how initial levels of a) language resources and b) teacher resources relate to program spending during the intervention (see Figure 10). Municipalities are

divided into an above (high) and a below (low) median pre-intervention resource group.[41] Program spending is first categorized as either language-oriented or not, and within these broad categories, spending is classified as oriented toward enhancing i) teaching, i.e., professional development and coaching of teachers, salaries for substitutes, in green, ii) governance, i.e., supporting organization and management, in black, or (iii) other types of support, in gray.

Figure 10 shows that low (high) pre-intervention language resources (a) and teaching resources (b) are correlated with lower (higher) shares of language oriented program spending, although differences are not stark. Similarly, a smaller share of the budget is directed towards teaching (green part of the bar) in municipalities where resources were initially low. Interestingly, municipalities with initially low language oriented resources have a relatively small share of program spending on language oriented teaching, and instead spend more on governance, suggesting that they needed to put more effort into organizing these types of services.

Figure 10: Allocation of program budget by initial resources



Notes: The share of budget spent for low/high initial resources at the municipal level. Language resources is an index of Swedish as 2nd language, home language class and tutoring. Teacher resources is an index of the share certified teachers, the teacher-student ratio and teacher experience. Resources are equally weighted and expressed at the municipal level, measured the two years before implementation. The different types of spending are mutually exclusive and can range between 0 and 1.

In line with the notion that municipalities with limited language-oriented resources required greater organizational support, we also find higher program spending per student in these municipalities (See Appendix Figure B12). Spending per student was, however, higher in municipalities with high initial teacher resources, as compared to municipalities with initially below median teaching resources.

We explore how the effects on test scores relate to theses spending patterns, and we find some variation by allocation of the budget. However, both levels and patterns of spending are endogenous, thus this analysis should be seen as descriptive rather than

[41] An index of individual language support is created based on the measures related to language enhancement in Table A25 and an index of initial teacher resources is based on the measures in Table A28.

causal. First, in panel a) of Figure B13, we split treated municipalities into two groups, above and below median (high, low) program spending per student and find that there is no difference in overall program effects on test scores. Effects on test scores are, however, larger in municipalities where more was spent on salaries–both language and non-language oriented modules. Non-language oriented spending on governance and development is instead associated with lower test score effects (see panel B of Figure B13).

In conclusion, the analysis of initial conditions and spending during the intervention provides suggestive evidence that effects of program participation were more positive in municipalities where the support efforts were able to focus on language enhancing and on teaching activities, as opposed to non-language oriented management support and governance structures.

5 Conclusion

Promoting immigrant student integration and improving teaching quality for multilingual students through national policy is challenging, both because evidence from small scale interventions and qualitative research does not necessarily scale up, sustained effects of organizational and professional development require local ownership and buy-in among school professionals, and because there may be fears that improving quality of education for some students comes at the expense of others. Yet, the dialogue based support program with this aim, launched in 2016 by the Swedish National Agency for Education(SNAE) targeting municipalities heavily impacted by the 2015 refugee crisis, seems to have overcome at least some of these challenges.

The SNAE targeted support program focused on professional development to improve language awareness and teaching practices, to integrate content learning and language development, and more broadly provide organizational and management support for the reception and integration of migrant students. Central to this form of dialogue based school development program is that a package of professional and organizational development measures –customized to the specific needs of the municipality– is the outcome of an analysis of local conditions, conducted in collaboration between a local team and process supporters from the SNAE. This package is then implemented by the municipality with financial and managerial support from the SNAE. Thanks to the pairwise randomized roll-out of the program, we have evaluated its impact.

In a 7-year follow-up, we find that average core subject test scores improved by 0.021 sd, mathematics test scores improved by 0.032 while the overall impact on Swedish test scores was small and insignificant. Gains in mathematics are present for all students both during and after the implementation. Swedish test scores took time to materialize and improved for all groups in the years just after the implementation. In the full 7 year follow-up period, the Swedish gains are present only for recent migrant students when we account for changes in student composition. Exploring heterogeneous effects, we find that the positive effects of the program for Swedish background and second generation immigrant students are larger for boys, while both immigrant boys and girls gained. We

also find that the benefits of the program were larger for low performing students.

We can conclude that the targeted support program had positive effects on learning in both mathematics and Swedish, and that it reduced test score gaps between foreign and Swedish background students and between boys and girls. The sustained gains are of an order of magnitude enough to pass a cost benefit test, giving thirteen euros in return for every euro spent. In spite of the program's focus on immigrant integration, the gains materialized earlier–and were also sustained–for second generation immigrant and Swedish background students. For immigrant students, the positive effects materialized only after the implementation period.

An analysis of program effects on access to language related resources, class room organization and teaching resources provides some insights into what may drive the results. First, second generation immigrant students likely benefited form improved language awareness among teachers, and there are some signs that access to language oriented resources improved, possibly contributing to their early test score gains. Post-implementation, treated municipalities expanded access to language enhancing resources for foreign background students compared to control municipalities. A reason could be that there was a strong focus in all municipalities on providing access to these types of resources for recent immigrants in the years just after the refugee crisis, such that i) the program made a difference in access for other groups of multilingual background student and ii) that access remained also as the crisis subsided. Second, we find that treated municipalities were initially more likely to engage in segregating school/and class room organization, placing recent immigrants in reception classes, but that they then became less inclined than control municipalities to do so. These reception classes, while allowing for a strong focus on initial language acquisition, may have contributed to an initial crowding out of mathematics instruction. Increased teacher-student ratios could also have contributed to positive effects, while increased class sizes likely reduced the positive impact of the program.

To conclude, our analysis shows that the dialogue based support program had some impact on resource allocation and organizational practices also in the longer term, moving away from segregating practices and maintaining access to language enhancing pedagogical resources. That the program was particularly successful in improving test scores for second generation immigrants, could reflect that the targeted municipalities had little previous experience of immigration, suggesting that this student group earlier had been deprived of the language enhancing resources they now benefited from. The more general test score gains in mathematics and for boys can be interpreted as supporting the importance of integrating language learning with content learning and that there are spillover effect of supporting strong language development.

References

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2019). "Family disadvantage and the gender gap in behavioral and educational outcomes". *American Economic Journal: Applied Economics* 11.3, 338–81.

Bertrand, Marianne and Jessica Pan (2013). "The trouble with boys: Social influences and the gender gap in disruptive behavior". *American Economic Journal: Applied Economics* 5.1. Publisher: American Economic Association, 32–64.

Björklund, Anders, Per-Anders Edin, Peter Fredriksson, and Alan Krueger (2004). *Education, equality, and efficiency: An analysis of Swedish school reforms during the 1990s*. Vol. 1. Citeseer.

Böhlmark, Anders, Helena Holmlund, and Mikael Lindahl (2016). "Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform". *Journal of Population Economics* 29.4. Publisher: Springer, 1155–1190.

Brandén, Maria, Gunn Elisabeth Birkelund, and Ryszard Szulkin (2019). "Ethnic composition of schools and students' educational outcomes: Evidence from Sweden". *International Migration Review* 53.2. Publisher: Sage Publications Sage CA: Los Angeles, CA, 486–517.

Breidbach, Stephan and Britta Viebrock, eds. (2013). *Content and Language Integrated Learning (CLIL) in Europe: research perspectives on policy and practice*. eng ger. Mehrsprachigkeit in Schule und Unterricht Band 14. Frankfurt am Main Bern Bruxelles New York Oxford Warszawa Wien: Peter Lang Edition.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (Apr. 2018). "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico". en. *Journal of Political Economy* 126.2, 635–687.

Bunar, Nihad (2010). "Nyanlända och lärande - mottagande och inkludering av nyanlända elever i den svenska skolan". PhD Thesis. Stockholms universitet.

Bunar, Nihad (2017). "Migration and education in Sweden: Integration of migrants in the Swedish school education and higher education systems". *NESET II ad hoc Question* 3, 1–19.

Bunar, Nihad and Päivi Juvonen (Nov. 2022). "'Not (yet) ready for the mainstream' – newly arrived migrant students in a separate educational program". en. *Journal of Education Policy* 37.6, 986–1008.

Clarke, Damian, Joseph P. Romano, and Michael Wolf (Dec. 2020). "The Romano–Wolf multiple-hypothesis correction in Stata". en. *The Stata Journal: Promoting communications on statistics and Stata* 20.4, 812–843.

Cummins, Jim (Nov. 2019). “Should Schools Undermine or Sustain Multilingualism? an Analysis of Theory, Research, and Pedagogical Practice”. en. *Sustainable Multilingualism* 15.1, 1–26.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler (2016). “The impact of immigration: Why do studies reach such different results?” *Journal of Economic Perspectives* 30.4, 31–56.

Farre, Lidia, Francesc Ortega, and Ryuichi Tanaka (Apr. 2018). “Immigration and the public–private school choice”. en. *Labour Economics* 51, 184–201.

Figlio, David, Paola Giuliano, Riccardo Marchingiglio, Umut Özek, and Paola Sapienza (2023). “Diversity in schools: Immigrants and the educational performance of US born students”. *Review of Economic Studies*.

Figlio, David and Umut Özek (2019). “Unwelcome guests? The effects of refugees on the educational outcomes of incumbent students”. *Journal of Labor Economics* 37.4. Publisher: The University of Chicago Press Chicago, IL, 1061–1096.

Fort, Margherita, Andrea Ichino, and Giulio Zanella (Jan. 2020). “Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families”. en. *Journal of Political Economy* 128.1, 158–205.

Getik, Demid, Anna Sjögren, and Anton Sundberg (2024). “Migration inflow and school performance of incumbents”. Working paper 2024:XX. IFAU.

Gibbons, Pauline (2015). *Scaffolding language, scaffolding learning: teaching English language learners in the mainstream classroom*. eng. Second edition. Portsmouth, NH: Heinemann.

Gould, Eric D, Victor Lavy, and M Daniele Paserman (2009). “Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence”. *The Economic Journal* 119.540. Publisher: Oxford University Press Oxford, UK, 1243–1269.

Green, Colin and Jon Marius Vaag Iversen (2022). “Refugees and the educational attainment of natives: Evidence from Norway”. *Economics of Education Review* 88. Publisher: Elsevier, 102258.

Grönqvist, Erik, Björn Öckert, and Olof Rosenqvist (2025). “Does the “Boost for Mathematics” Boost Mathematics? A Large-Scale Evaluation of the “Lesson Study” Methodology on Student Performance”. en. *American Economic Journal: Economic Policy* ().

Grönqvist, Hans and Susan Niknami (2017). “The school achievements of refugee children : Lessons from Sweden”. *Nordic Economic Policy Review* 520. Publisher: Stockholm University, The Swedish Institute for Social Research (SOFI), 159–183.

Gurevich, Tamara, Peter Herman, Farid Toubal, and Yoto Yotov (2024). *Domestic and International Common Language (DICL) Database*. Artwork Size: 2048, 6791889,

9219124, 4825, 5163, 41238, 41227 Pages: 2048, 6791889, 9219124, 4825, 5163, 41238, 41227.

Hassan, Said, Camilla Hvidtfeldt, Lars Højsgaard Andersen, and Rebecca Overgaard Udsen (2023). "Do refugee children impair the academic performance of native children in the school? Informative null results from Danish register data". *European Sociological Review* 39.3. Publisher: Oxford University Press UK, 352–365.

Holmlund, Helena, Josefin Häggblom, and Erica Lindahl (2024). "The Boost for reading". IFAU WP 2024:6.

Holmlund, Helena, Anna Sjögren, and Björn Öckert (2019). "Jämlikhet i utfall och möjligheter i den svenska skolan". Bilaga till Långtidsutredningen SOU2019:40. Finansdepartementet.

Kalinowski, Eva, Anna Gronostaj, and Miriam Vock (Jan. 2019). "Effective Professional Development for Teachers to Foster Students' Academic Language Proficiency Across the Curriculum: A Systematic Review". en. *AERA Open* 5.1, 2332858419828691.

Kraft, Matthew A., David Blazar, and Dylan Hogan (Aug. 2018). "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence". en. *Review of Educational Research* 88.4, 547–588.

Machin, Stephen and Sandra McNally (June 2008). "The literacy hour". en. *Journal of Public Economics* 92.5-6, 1441–1462.

Mörtlund, Tove (2020). "Arbetet med nyanlända elever i grundskolan". Working Paper 2020:2. IFAU.

OECD (2015). "Helping immigrant students to succeed at school – and beyond". Tech. rep. OECD.

Palik, Júlia and Gudrun Østby (2023). "Interventions to improve refugee children's access to education and quality learning: A scoping review of existing impact evaluations". *International Review of Education. Internationale Zeitschrift Fur Erziehungswissenschaft. Revue Internationale De Pedagogie* 69.1-2, 227–247.

Pinson, Halleli, Nihad Bunar, and Dympna Devine, eds. (2023). *Research handbook on migration and education*. Elgar handbooks in education. OCLC: on1391325824. Cheltenham, UK ; Northampton, MA, USA: Edward Elgar Publishing.

Redding, Christopher and Tuan D. Nguyen (Dec. 2020). "The Relationship Between School Turnaround and Student Outcomes: A Meta-Analysis". en. *Educational Evaluation and Policy Analysis* 42.4, 493–519.

Regeringen (1975). *Proposition 1975:26 - Riktlinjer för invandrar- och minoritetspolitiken*.

Regeringen (2014). "Utbildning för nyanlända elever – mottagande och skolgång". Proposition 2014/15:45.

Ruiz De Zarobe, Yolanda and Rosa María Jiménez Catalán, eds. (Dec. 2009). *Content and Language Integrated Learning: Evidence from Research in Europe*. Multilingual Matters.

SCB (Mar. 2025). *Sveriges befolkning i sammandrag 1960-2024*.

Schueler, Beth E., Catherine Armstrong Asher, Katherine E. Larned, Sarah Mehrotra, and Cynthia Pollard (Oct. 2022). "Improving Low-Performing Schools: A Meta-Analysis of Impact Evaluation Studies". en. *American Educational Research Journal* 59.5, 975–1010.

Sims, Sam, Harry Fletcher-Wood, Alison O'Mara-Eves, Sarah Cottingham, Claire Stansfield, Josh Goodrich, Jo Van Herwegen, and Jake Anders (Apr. 2025). "Effective Teacher Professional Development: New Theory and a Meta-Analytic Test". en. *Review of Educational Research* 95.2, 213–254.

Skollag (2010). *SFS 2010:800*. sv. Utbildningsdepartementet.

Skolöverstyrelsen (1980). *Läroplan för grundskolan Lgr 80*.

Skolverket (2015). *Studiehandledning på modersmål - att stödja kunskapsutvecklingen hos flerspråkiga elever*.

Skolverket (2016). *Utbildning för nyanlända elever - Vägledning för huvudmän och skolor*.

Skolverket (2021). "Utvärdering av riktade insatser för nyanländas lärande". Rapporter 2021:11.

Stolk, Yvonne, Ida Kaplan, and Josef Szwarc (May 2023). "Majority language acquisition by children of refugee background: a review". en. *International Journal of Inclusive Education*, 1–24.

Sun, Min, Alec I. Kennedy, and Susanna Loeb (Dec. 2021). "The Longitudinal Effects of School Improvement Grants". en. *Educational Evaluation and Policy Analysis* 43.4, 647–667.

Sun, Min, Emily K. Penner, and Susanna Loeb (Aug. 2017). "Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants". en. *American Educational Research Journal* 54.4, 607–643.

Taylor, Eric S. (2023). "Teacher evaluation and training". en. *Handbook of the Economics of Education*. Vol. 7. Elsevier, 61–141.

Tumen, Semih (2021). "The effect of refugees on native adolescents' test scores: Quasi-experimental evidence from PISA". *Journal of Development Economics* 150. Publisher: Elsevier, 102633.

Utbildningsdepartementet (1977). "Proposition 1976/77:118 Om hemspråksundervisning för invandrarbarn m.m." Tech. rep.

Utbildningsdepartementet (2013). "Utbildning för nyanlända elever". Tech. rep. DS 2013:6. Utbildningsdepartementet.

Vlachos, Jonas (2019). “Trust-based evaluation in a market-oriented school system”. *Neoliberalism and Market Forces in Education*. Routledge, 212–230.

Appendix A: Additional tables

Table A1: Sampling by sampling round

	(1)	(2)	(3)	(4)
Sample round	Start date (first)	Sampled pairs	Future T or G	Previous C
1	2016 May	13	12	0
2	2016 June	13	8	0
3	2016 December	4	3	8
4	2017 July	9	1	5
5	2018 January	6	0	8
6	2018 June	10	0	1
7	2018 December	8	0	2

Notes: The table displays the number of pairs. T refers to Controls randomized into treatment, and similarly G refers to Controls randomized into Guaranteed treatment.

Table A2: Sampling by initial and re-sampling round

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First sample round	Sampled R1	Sampled R2	Sampled R3	Sampled R4	Sampled R5	Sampled R6	Sampled R7	Total resampled	
Round 1	13	0	8(3+5)	3(1+2)	1(0+1)	0	0	12	
Round 2	-	13	0	0	5(3+2)	1(1+0)	2(1+1)	8	
Round 3	-	-	4	2(0+2)	1(0+1)	0	0	3	
Round 4	-	-	-	9	1(0+1)	0	0	1	
Round 5	-	-	-	-	6	0	0	0	
Round 6	-	-	-	-	-	10	0	0	
Round 7	-	-	-	-	-	-	8	0	

Notes: Each cell contains the total number of pairs where one is sampled to be either treated or guarantee. Total(T+G).

Table A3: Pairs by sampling round and time since first contract

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample round	First year	Second year	Third year	Fourth year	Fifth year	Sixth year	Seventh year
1	5	1	1	-	-	1	1
2	13	8	5	-	-	5	5
3	4	1	1	-	-	1	1
4	8	8	-	-	8	8	-
5	6	6	-	-	6	6	-
6	10	-	-	10	10	-	-
7	8	-	-	8	8	-	-

Notes: The number of pairs for which we can estimate impacts on test scores, in years since the first date of the wave.

Table A4: Test score, excluding independent schools

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.021***	0.019**	0.065**	0.025	0.040
	(0.008)	(0.009)	(0.028)	(0.029)	(0.024)
Obs	195,368	158,911	12,733	11,030	12,694
Control mean	-0.141	-0.075	-0.304	-0.347	-0.659
Panel B: Implementation period					
Years 1-2	0.011	0.016	0.117***	0.018	-0.018
	(0.009)	(0.010)	(0.041)	(0.036)	(0.032)
Obs	86,318	71,123	4,749	3,080	7,366
Control mean	-0.165	-0.094	-0.327	-0.363	-0.691
Panel C: Early post intervention period					
Years 3-4	0.059**	0.041	0.131***	0.336***	0.079*
	(0.027)	(0.025)	(0.040)	(0.064)	(0.045)
Obs	37,051	30,148	2,646	2,018	2,239
Control mean	-0.102	-0.031	-0.299	-0.325	-0.679
Panel D: Late post intervention period					
Years 5-7	0.045***	0.029**	0.070**	0.062**	0.182***
	(0.012)	(0.011)	(0.034)	(0.031)	(0.031)
Obs	71,999	57,640	5,338	5,932	3,089
Control mean	-0.133	-0.076	-0.287	-0.346	-0.567

Notes: Main analysis on average test scores in grades 3, 6, and 9, excluding independent schools. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A5: Effects on teacher assessment grades and when imputing missing test scores with grades

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: Test score					
Treated	0.026**	0.025**	-0.003	0.014	0.003
	(0.010)	(0.011)	(0.030)	(0.031)	(0.026)
Obs	118,358	97,958	6,932	7,013	6,455
Control mean	-0.169	-0.116	-0.315	-0.339	-0.659
Panel B: GPA					
Treated	0.040***	0.040***	0.022	0.023	0.002
	(0.010)	(0.009)	(0.039)	(0.034)	(0.038)
Obs	118,358	97,958	6,932	7,013	6,455
Control mean	-0.110	-0.013	-0.314	-0.474	-1.029
Panel C: Imputed					
Treated	0.018**	0.016*	0.030	0.021	-0.005
	(0.009)	(0.009)	(0.035)	(0.025)	(0.026)
Obs	211,988	173,283	12,468	12,825	13,412
Control mean	-0.162	-0.084	-0.300	-0.390	-0.862

Notes: The analysis spans the full 7-year period, but is limited to grade 6 and 9. Panel A and B includes only students for whom we have both test score and GPA. Panel C presents test score including year 2019 and 2020, which are imputed with GPA. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1.

Commentary: Missing data due to non-collection of test scores. The timing of the intervention relative to the pandemic implies that we are missing two years of outcomes; test scores for 2019 and 2020, affecting primarily the post intervention period. To deal with this, an alternative outcome measure is grades based on teacher assessments, GPA. While these are arguably less objective, they constitute the best available measure for these years. In Table A5, Panel A and B reports test scores and GPA for the limited sample of students for whom we have both GPA and test scores. As this removes all students in grade 3, the sample shrinks substantially, but the overall test score results are significant and positive. Importantly, the GPA and test score results are consistent. Imputing test scores with grades for the pandemic years when test scores are missing in Panel C of Table A5, the results are similar to the main estimation in Table 4. However, the imputed overall estimate is lower than test scores and GPA separately (Panel A and B in Table A5), thus the years of the pandemic seem to be influential. Possibly, the intervention was less successful during the chaotic years of the pandemic, but we cannot rule out that grading standards were different during those two years.

Table A6: Balance table, ranking score and school resources

	(1)	(2)	(3)	(4)	(5)
	Ranking score	Teacher exp.	Share teachers certified	Teacher-student ratio	Number of students
Treated	-0.011	0.019	-0.000	0.059	512.027**
	(0.008)	(0.273)	(0.010)	(0.135)	(240.929)
Obs	126	126	126	126	126
Control mean	1.717	13.688	0.860	8.759	1428.072

Notes: Measured in the 2 years before sampling in the wave. Teacher-student ratio is the number of teachers per 100 students.

Table A7: Balance table, student composition: The fraction of the respective group in the municipality's population of students

	(1)	(2)	(3)	(4)	(5)
	Swedish background	2nd gen.	Foreign born	Recent immigrant	Asylum seekers
Treated	-0.020	0.008	0.012	0.011	0.003
	(0.013)	(0.006)	(0.010)	(0.007)	(0.003)
Obs	126	126	126	126	126
Control mean	0.843	0.044	0.113	0.086	0.027

Notes: Measured in the 2 years before sampling in the wave. Swedish background are Swedish born students with at least one parent born in Sweden, Recent immigrants are students immigrating within 4 years, both referring to the shares as reported in the student registry.

Table A8: Main effects, unweighted

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.018**	0.010	0.076***	0.028	0.053***
	(0.007)	(0.008)	(0.020)	(0.023)	(0.019)
Obs	209,081	171,022	13,605	11,511	12,943
Control mean	-0.135	-0.071	-0.299	-0.340	-0.656
Panel B: Implementation period					
Years 1-2	0.011	0.012	0.095***	0.000	-0.012
	(0.007)	(0.008)	(0.034)	(0.030)	(0.027)
Obs	91,833	76,079	5,054	3,210	7,490
Control mean	-0.158	-0.090	-0.320	-0.354	-0.686
Panel C: Post intervention period					
Years 3-4	0.067**	0.032	0.153***	0.352***	0.125**
	(0.026)	(0.024)	(0.047)	(0.099)	(0.052)
Obs	39,880	32,633	2,852	2,110	2,285
Control mean	-0.096	-0.027	-0.297	-0.318	-0.676
Panel D: Post intervention period					
Years 5-7	0.029**	0.011	0.088***	0.069**	0.131***
	(0.012)	(0.012)	(0.027)	(0.028)	(0.030)
Obs	77,368	62,310	5,699	6,191	3,168
Control mean	-0.127	-0.071	-0.282	-0.341	-0.565

Notes: Unweighted analysis. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A9: Effects of Swedish test, unweighted

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.010	0.002	0.050**	0.014	0.072**
	(0.008)	(0.009)	(0.025)	(0.025)	(0.029)
Obs	202,372	166,388	13,315	11,198	11,471
Control mean	-0.122	-0.079	-0.184	-0.200	-0.634
Panel B: Implementation period					
Years 1-2	0.006	0.003	0.086**	0.018	0.040
	(0.010)	(0.010)	(0.039)	(0.035)	(0.041)
Obs	88,752	74,108	4,938	3,119	6,587
Control mean	-0.148	-0.093	-0.214	-0.225	-0.722
Panel C: Early post intervention period					
Years 3-4	0.093***	0.052***	0.099	0.386**	0.226***
	(0.022)	(0.018)	(0.075)	(0.148)	(0.066)
Obs	38,658	31,764	2,789	2,059	2,046
Control mean	-0.098	-0.051	-0.188	-0.217	-0.635
Panel D: Late post intervention period					
Years 5-7	0.002	-0.009	0.052*	0.026	0.072**
	(0.009)	(0.010)	(0.030)	(0.026)	(0.032)
Obs	74,962	60,516	5,588	6,020	2,838
Control mean	-0.103	-0.077	-0.156	-0.182	-0.420

Notes: Unweighted analysis. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A10: Effects on math test, unweighted

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.028***	0.022**	0.106***	0.042	0.042*
	(0.010)	(0.009)	(0.025)	(0.027)	(0.022)
Obs	199,602	163,463	13,091	10,937	12,111
Control mean	-0.115	-0.036	-0.395	-0.451	-0.619
Panel B: Implementation period					
Years 1-2	0.019**	0.023**	0.096**	-0.009	-0.024
	(0.009)	(0.009)	(0.041)	(0.041)	(0.026)
Obs	85,728	71,216	4,755	2,903	6,854
Control mean	-0.137	-0.062	-0.408	-0.488	-0.610
Panel C: Post intervention period					
Years 3-4	0.050	0.024	0.229***	0.331***	0.011
	(0.035)	(0.033)	(0.050)	(0.077)	(0.070)
Obs	38,523	31,525	2,776	2,023	2,199
Control mean	-0.064	0.021	-0.380	-0.387	-0.649
Panel D: Post intervention period					
Years 5-7	0.053***	0.032*	0.146***	0.088**	0.169***
	(0.018)	(0.017)	(0.031)	(0.035)	(0.035)
Obs	75,351	60,722	5,560	6,011	3,058
Control mean	-0.118	-0.037	-0.393	-0.456	-0.618

Notes: Unweighted analysis. Treatment is an indicator for being selected for treatment. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A11: Effects on test scores in mathematics and Swedish with individual controls

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.020**	0.015	0.046*	0.024	0.053**
	(0.010)	(0.010)	(0.026)	(0.029)	(0.023)
Obs	209,081	171,022	13,605	11,511	12,943
Control mean	-0.135	-0.071	-0.299	-0.340	-0.656
Panel B: Implementation period					
Years 1-2	0.014	0.017	0.089**	0.010	-0.001
	(0.010)	(0.011)	(0.035)	(0.034)	(0.032)
Obs	91,833	76,079	5,054	3,210	7,490
Control mean	-0.158	-0.090	-0.320	-0.354	-0.686
Panel C: Early post intervention period					
Years 3-4	0.032	0.016	0.098**	0.306***	0.083*
	(0.024)	(0.025)	(0.040)	(0.064)	(0.046)
Obs	39,880	32,633	2,852	2,110	2,285
Control mean	-0.096	-0.027	-0.297	-0.318	-0.676
Panel D: Late post intervention period					
Years 5-7	0.039***	0.020	0.059	0.073**	0.196***
	(0.013)	(0.012)	(0.038)	(0.029)	(0.031)
Obs	77,368	62,310	5,699	6,191	3,168
Control mean	-0.127	-0.071	-0.282	-0.341	-0.565

Notes: Individual controls are the same variables used to predict test scores. Treatment is an indicator for being selected for treatment. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A12: Effects on test scores in Swedish with individual controls

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.010	0.008	0.028	-0.014	0.053*
	(0.009)	(0.009)	(0.030)	(0.034)	(0.031)
Obs	202,372	166,388	13,315	11,198	11,471
Control mean	-0.122	-0.079	-0.184	-0.200	-0.634
Panel B: Implementation period					
Years 1-2	0.010	0.010	0.086**	-0.018	0.043
	(0.011)	(0.011)	(0.042)	(0.044)	(0.043)
Obs	88,752	74,108	4,938	3,119	6,587
Control mean	-0.148	-0.093	-0.214	-0.225	-0.722
Panel C: Early post intervention period					
Years 3-4	0.047***	0.025	0.079	0.325***	0.207***
	(0.017)	(0.016)	(0.060)	(0.094)	(0.074)
Obs	38,658	31,764	2,789	2,059	2,046
Control mean	-0.098	-0.051	-0.188	-0.217	-0.635
Panel D: Late post intervention period					
Years 5-7	0.001	-0.008	-0.002	0.019	0.095***
	(0.009)	(0.009)	(0.039)	(0.029)	(0.034)
Obs	74,962	60,516	5,588	6,020	2,838
Control mean	-0.103	-0.077	-0.156	-0.182	-0.420

Notes: Individual controls are the same variables used to predict test scores. Treatment is an indicator for being selected for treatment. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A13: Effects on test scores in mathematics with individual controls

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: In total					
Years 1-7	0.032**	0.026**	0.076**	0.070**	0.042*
	(0.013)	(0.012)	(0.032)	(0.032)	(0.025)
Obs	199,602	163,463	13,091	10,937	12,111
Control mean	-0.115	-0.036	-0.395	-0.451	-0.619
Panel B: Implementation period					
Years 1-2	0.022*	0.025*	0.099**	0.054	-0.016
	(0.012)	(0.013)	(0.041)	(0.038)	(0.034)
Obs	85,728	71,216	4,755	2,903	6,854
Control mean	-0.137	-0.062	-0.408	-0.488	-0.610
Panel C: Post intervention period					
Years 3-4	0.019	0.010	0.128*	0.301***	-0.064
	(0.034)	(0.035)	(0.066)	(0.053)	(0.054)
Obs	38,523	31,525	2,776	2,023	2,199
Control mean	-0.064	0.021	-0.380	-0.387	-0.649
Panel D: Post intervention period					
Years 5-7	0.074***	0.049***	0.146***	0.108***	0.246***
	(0.020)	(0.018)	(0.043)	(0.039)	(0.034)
Obs	75,351	60,722	5,560	6,011	3,058
Control mean	-0.118	-0.037	-0.393	-0.456	-0.618

Notes: Individual controls are the same variables used to predict test scores. Treatment is an indicator for being selected for treatment. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A14: Grade specific effects on test score in grades 3,6 and 9, during years 1-7.

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: Grade 3					
Treated	0.014	0.007	0.128***	0.001	0.024
	(0.015)	(0.013)	(0.046)	(0.050)	(0.056)
Obs	71,935	58,508	5,532	2,875	5,019
Control mean	-0.080	-0.003	-0.305	-0.355	-0.621
Panel B: Grade 6					
Treated	0.034**	0.031**	0.053	0.048	-0.006
	(0.015)	(0.015)	(0.033)	(0.035)	(0.028)
Obs	70,344	57,678	4,536	4,047	4,080
Control mean	-0.146	-0.092	-0.316	-0.309	-0.604
Panel C: Grade 9					
Treated	0.019*	0.022*	-0.027	-0.015	0.119***
	(0.011)	(0.013)	(0.037)	(0.030)	(0.038)
Obs	66,802	54,836	3,532	4,585	3,844
Control mean	-0.181	-0.122	-0.272	-0.359	-0.750

Notes: Test scores years 1-7. Treatment is an indicator. Recent includes Asylum seekers. Treatment is an indicator for being selected for treatment. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A15: Test score p-values corrected for multiple hypothesis during years 1-7.

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Average test score	0.068	0.128	0.072	0.709	0.446
Mathematics test score	0.040	0.048	0.032	0.207	0.709
Swedish test score	0.518	0.570	0.570	0.725	0.570

Notes: The corrected p-values applies the Romano-Wolf adjustment, see details in (Clarke, Romano, and Wolf, 2020). Test scores years 1-7, including the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A16: Test score effects in an integrated model, during years 1-7.

	(1) Average test score	(2) Mathematics	(3) Swedish
Treated	0.026***	0.037***	0.015
	(0.010)	(0.012)	(0.010)
Treated X second gen.	0.018	0.028	0.012
	(0.036)	(0.041)	(0.041)
Treated X non-recent	-0.031	-0.025	-0.032
	(0.033)	(0.035)	(0.042)
Treated X recent	-0.055	-0.061	-0.043
	(0.039)	(0.042)	(0.048)
Second generation	-0.228***	-0.365***	-0.096***
	(0.025)	(0.028)	(0.028)
Non-recent	-0.185***	-0.333***	-0.042
	(0.027)	(0.027)	(0.033)
Recent	-0.495***	-0.487***	-0.469***
	(0.029)	(0.030)	(0.036)
Obs	209,081	199,602	202,372
Control mean	-0.071	-0.036	-0.079

Notes: Test scores years 1-7. Treatment is an indicator for being selected for treatment. The omitted group is Swedish background. Recent includes Asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A17: Asylum seekers, total effects

	(1) Test score	(2) Math score	(3) Swe score	(4) Swe 2nd	(5) Test taking	(6) Pred. score	(7) Pred. score test-takers
Years 1-7	0.092	0.167*	-0.056	0.002	-0.023	-0.006	0.003
	(0.089)	(0.085)	(0.108)	(0.023)	(0.031)	(0.007)	(0.006)
Obs	1,420	1,290	1,169	2,438	2,438	2,438	1,420
Control mean	-0.799	-0.639	-0.967	0.435	0.598	-1.439	-1.501

Notes: Estimation for the full 1-7 year period. Column 1-3 show test scores pooled, in math and in Swedish respectively. Swed 2nd is an indicator of taking Swedish as a second language. Test participation is an indicator taking 1 if the child took the test. Predicted test scores are based on year, age at immigration, gender, migration status, as well as parental information about earnings and education. In column 7, the prediction is restricted to the population of students taking the national test. All estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A18: Effects on the composition of students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	During intervention					Post period (3-4 years)					Post period (5-7 years)				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent
Panel A: Test participation															
Treated	-0.001	0.000	0.001	0.009	-0.007	-0.007***	-0.007***	-0.006**	0.020***	0.033**	-0.001	-0.007***	0.008**	0.015***	0.020
	(0.004)	(0.003)	(0.007)	(0.008)	(0.016)	(0.002)	(0.001)	(0.003)	(0.006)	(0.015)	(0.002)	(0.002)	(0.004)	(0.005)	(0.012)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	0.956	0.975	0.975	0.957	0.789	0.968	0.973	0.981	0.980	0.874	0.964	0.967	0.978	0.972	0.859
Panel B: Predicted test scores															
Treated	-0.004	-0.005	-0.031	0.050**	-0.010	-0.002	-0.005	0.004	-0.029	0.008	-0.003	-0.003	0.007	0.009	-0.011
	(0.004)	(0.004)	(0.021)	(0.019)	(0.010)	(0.005)	(0.005)	(0.016)	(0.021)	(0.016)	(0.005)	(0.005)	(0.013)	(0.011)	(0.011)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	-0.820	-0.754	-1.052	-0.874	-1.235	-1.105	-1.043	-1.307	-1.376	-1.493	-1.218	-1.156	-1.386	-1.503	-1.597
Panel C: Predicted test scores, test takers															
Treated	-0.005	-0.005	-0.037*	0.049**	-0.015	-0.003	-0.004	0.011	-0.037*	-0.018	-0.004	-0.003	0.005	0.007	-0.011
	(0.004)	(0.005)	(0.020)	(0.019)	(0.015)	(0.005)	(0.006)	(0.015)	(0.021)	(0.018)	(0.005)	(0.005)	(0.013)	(0.011)	(0.010)
Obs	91,833	76,079	5,054	3,210	7,490	39,880	32,633	2,852	2,110	2,285	77,368	62,310	5,699	6,191	3,168
Control mean	-0.818	-0.759	-1.055	-0.881	-1.255	-1.099	-1.039	-1.304	-1.370	-1.488	-1.211	-1.149	-1.383	-1.499	-1.591

Notes: During intervention refers to years 1-2, Post period is split into 3-4 years and 5-7 years, after implementation. Test participation in Panel A is an indicator taking 1 if the child took the test. In Panel B, test scores are predicted using year, age at immigration, gender, migration status, as well as parental information about earnings and education. In Panel C, the prediction is restricted to the population of students taking the national test.

Table A19: Student mobility in & out from the municipality, municipal level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number children	Share new	Share Asylum	Moving in		Moving out	
				Share	Predicted test score	Share	Predicted test score
Panel A: Implementation period							
Years 1-2	-542.286***	0.003	0.009***	-0.001	0.031	0.000	-0.026*
	(123.592)	(0.003)	(0.003)	(0.001)	(0.021)	(0.001)	(0.015)
Obs	88,621	88,621	88,621	88,621	88,621	88,621	88,621
Control mean	3169.842	0.102	0.022	0.025	-1.155	0.027	-1.152
Panel B: Post implementation period							
Years 3-7	-371.881***	0.004	0.000	0.003	-0.017	0.003	-0.024
	(85.194)	(0.003)	(0.001)	(0.002)	(0.012)	(0.002)	(0.015)
Obs	121,606	121,606	121,606	121,606	121,606	121,606	121,606
Control mean	3493.059	0.047	0.004	0.026	-1.301	0.027	-1.333

Notes: Mobility is defined relative to the previous year. In mobility reflects the students who were new in the municipality, the test year of interest. Out mobility reflects the students who are missing in the municipality, i.e., those who moved out by the test year of interest. Averages at the municipal level, including all grades (1-9 for in, 0-8 for out). Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A20: Mobility in and out from the municipality, individual level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	During intervention					Post period				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent
Panel A: Moving to the municipality										
Treated	-0.002	-0.002*	-0.006	0.006	-0.000	0.000	0.001	-0.005	0.001	-0.023**
	(0.001)	(0.001)	(0.009)	(0.006)	(0.006)	(0.002)	(0.002)	(0.007)	(0.006)	(0.011)
Obs	91,865	77,903	5,160	3,308	5,494	118,877	97,275	8,662	8,433	4,507
Control mean	0.023	0.018	0.037	0.037	0.069	0.023	0.019	0.035	0.045	0.062
Panel B: Moving from the municipality										
Treated	-0.003	0.000	-0.001	-0.008	-0.027***	0.000	0.001	0.012***	0.004	0.008
	(0.002)	(0.002)	(0.006)	(0.007)	(0.007)	(0.002)	(0.001)	(0.005)	(0.009)	(0.008)
Obs	86,139	72,595	5,005	2,327	6,212	118,919	97,131	8,647	6,431	6,710
Control mean	0.026	0.019	0.036	0.047	0.097	0.023	0.017	0.028	0.051	0.076

Notes: Mobility is defined relative to the previous year. In mobility reflects the children who were new in the municipality, the test year of interest. Out mobility reflects the children who are missing in the municipality, i.e., those who moved out by the test year of interest. The analysis is restricted to children observed in the student register two consecutive years, with a test score.

Table A21: Program cost per student

	(1) Mean	(2) Median	(3) Minimum	(4) Maximum
Per student and year	278	199	62	1267
Per unique student	1106	793	283	4593
Per student and year (in 1-2 years)	1605	1051	373	7888
Per unique student (in 1-2 years)	2152	1630	633	7901

Notes: Costs are expressed in SEK, according to the final program reports of the 61 municipalities that participated in the program (2 of the 63 municipalities that were offered treatment never participated).

Table A22: Heterogeneous test score effects by gender of the student, interaction model

	(1) Average test score	(2) Mathematics	(3) Swedish
Treated	0.026***	0.037***	0.016
	(0.009)	(0.012)	(0.010)
Treated X Girl	-0.011	-0.010	-0.009
	(0.012)	(0.013)	(0.014)
Obs	209,081	199,602	202,372
Control mean	-0.071	-0.036	-0.079

Notes: For the full period (years 1-7). Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A23: Heterogeneous test score effects by gender of the student

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Boys					Girls				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent			Second gen.	Non-recent	Recent
Panel A: In total										
Years 1-7	0.030***	0.024**	0.112***	0.040	0.061**	0.015	0.013	0.016	-0.003	0.045
	(0.009)	(0.010)	(0.038)	(0.032)	(0.031)	(0.009)	(0.009)	(0.037)	(0.035)	(0.035)
Obs	106,401	87,514	7,026	5,888	5,971	101,257	83,507	6,575	5,623	5,550
Control mean	-0.246	-0.192	-0.389	-0.431	-0.713	-0.007	0.056	-0.205	-0.240	-0.550
Panel B: Implementation period										
Years 1-2	0.024**	0.017	0.187***	0.064	-0.013	0.014	0.020**	0.066	-0.062	0.024
	(0.011)	(0.013)	(0.049)	(0.044)	(0.034)	(0.010)	(0.010)	(0.049)	(0.045)	(0.049)
Obs	46,632	38,898	2,627	1,647	3,458	44,178	37,180	2,423	1,563	3,010
Control mean	-0.291	-0.232	-0.453	-0.505	-0.739	-0.003	0.057	-0.175	-0.191	-0.567
Panel C: Early post intervention period										
Years 3-4	0.053*	0.043	0.214***	0.319***	0.208***	0.030	0.010	0.092	0.373***	0.095**
	(0.029)	(0.026)	(0.065)	(0.077)	(0.060)	(0.023)	(0.025)	(0.108)	(0.090)	(0.042)
Obs	20,267	16,694	1,512	1,062	999	19,381	15,939	1,340	1,048	1,054
Control mean	-0.202	-0.148	-0.376	-0.333	-0.735	0.022	0.098	-0.212	-0.302	-0.594
Panel D: Late post intervention period										
Years 5-7	0.059***	0.035***	0.066	0.111***	0.233***	0.010	0.000	0.043	-0.008	0.149***
	(0.010)	(0.010)	(0.060)	(0.037)	(0.040)	(0.014)	(0.014)	(0.064)	(0.038)	(0.038)
Obs	39,502	31,922	2,887	3,179	1,514	37,698	30,388	2,812	3,012	1,486
Control mean	-0.216	-0.166	-0.338	-0.426	-0.640	-0.028	0.030	-0.227	-0.244	-0.483
-0.028	0.030	-0.227	-0.244	-0.483						

Notes: During intervention refers to years 1-2, Post period refers year 3-7. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A24: Standardized language enhancing pedagogical resources

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Panel A: Implementation period					
Years 1-2	0.028	0.018	0.113	0.017	0.050
	(0.018)	(0.014)	(0.071)	(0.082)	(0.078)
Obs	96,029	78,006	5,183	3,332	9,508
Control mean	-0.011	-0.328	0.495	1.022	2.074
Panel B: Post intervention period					
Years 3-4	0.073***	-0.010	0.216***	0.429***	0.502***
	(0.021)	(0.010)	(0.063)	(0.107)	(0.097)
Obs	41,274	33,571	2,903	2,163	2,637
Control mean	-0.037	-0.329	0.491	1.459	2.122
Panel C: Post intervention period					
Years 5-7	0.132***	0.123***	-0.049	0.125	0.174**
	(0.024)	(0.017)	(0.082)	(0.078)	(0.074)
Obs	80,332	64,473	5,830	6,372	3,657
Control mean	-0.001	-0.336	0.582	1.684	2.253

Notes: During intervention regards the first two years since the program was initiated, Post period is presented for 2-4 years and 5-7 years separately.Language resources takes the value one for students receiving at least one of the following: Swedish as a second language, home language classes, tutoring in mother tongue. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A25: Pedagogical resources

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	During intervention					Post period (3-4 years)					Post period (5-7 years)				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent
Panel A: Swedish as 2nd language															
Treated	0.001	-0.001**	0.045	-0.016	-0.001	0.010**	-0.003	0.067**	0.053	0.041	-0.009**	-0.002**	-0.094**	-0.045	-0.034*
	(0.003)	(0.001)	(0.033)	(0.027)	(0.016)	(0.005)	(0.002)	(0.032)	(0.066)	(0.036)	(0.005)	(0.001)	(0.037)	(0.031)	(0.019)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	0.106	0.003	0.286	0.471	0.765	0.108	0.006	0.322	0.631	0.824	0.126	0.006	0.393	0.714	0.862
Panel B: Home language class															
Treated	0.020*	0.014	0.019	0.054**	0.011	0.009	-0.006	0.094***	0.069**	0.078**	0.111***	0.120***	0.095***	0.074***	0.080***
	(0.010)	(0.012)	(0.016)	(0.024)	(0.013)	(0.011)	(0.008)	(0.023)	(0.029)	(0.030)	(0.014)	(0.016)	(0.013)	(0.020)	(0.021)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	0.069	0.020	0.215	0.313	0.327	0.092	0.042	0.228	0.404	0.342	0.071	0.035	0.131	0.294	0.248
Panel C: Tutoring															
Treated	-0.000	0.001**	0.003	-0.014	-0.002	0.021***	0.000	-0.003	0.101***	0.158***	0.012**	-0.000	0.002	0.047*	0.059*
	(0.003)	(0.000)	(0.006)	(0.014)	(0.028)	(0.006)	(0.000)	(0.012)	(0.016)	(0.049)	(0.006)	(0.000)	(0.016)	(0.028)	(0.030)
Obs	93,888	78,006	5,183	3,332	7,367	40,878	33,571	2,903	2,163	2,241	79,895	64,473	5,830	6,372	3,220
Control mean	0.031	0.001	0.038	0.097	0.331	0.031	0.001	0.042	0.179	0.356	0.038	0.001	0.064	0.228	0.378
Panel D: Special resources															
Treated	-0.009*	-0.010**	0.001	-0.015	-0.017	0.015*	0.021***	0.009	-0.005	-0.015	-0.005	-0.005	-0.013	-0.008	-0.015
	(0.005)	(0.004)	(0.013)	(0.012)	(0.020)	(0.009)	(0.007)	(0.011)	(0.014)	(0.022)	(0.006)	(0.005)	(0.012)	(0.012)	(0.016)
Obs	94,195	78,006	5,183	3,332	7,674	40,935	33,571	2,903	2,163	2,298	80,067	64,473	5,830	6,372	3,392
Control mean	0.078	0.073	0.083	0.099	0.116	0.075	0.071	0.083	0.099	0.103	0.086	0.083	0.090	0.106	0.118

Notes: During intervention regards the first two years since the program was initiated, Post period is presented for 2-4 years and 5-7 years separately. Swedish as a second language (Panel A) is an indicator of having a grade in Swedish as a second language (as opposed to following the regular curriculum). Home language (Panel B) is an indicator of participating in extra instruction to learn their home language. Tutoring (Panel C) is an indicator of receiving tutoring in the class room on their home language. Special (Panel D) is a composite measure taking 1 if the student is on an adapted study path or receiving specialized tutoring either individually or in group. Information about tutoring and special resources is available only in the student registry, thus these analysis includes only students linked between the student registry and test score data (excluding all asylum seekers, who would otherwise be included in column 1, 5, and 7). Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Table A26: Effects of targeted support on average test scores by language background, 1-7 years

	(1)	(2)	(3)	(4)	(5)
	All	Swedish background	Foreign background		
			Second gen.	Non-recent	Recent
Treated x Same language	-0.010	0.009	-0.097	-0.042	-0.053
	(0.027)	(0.030)	(0.072)	(0.081)	(0.074)
Treated	0.023**	0.018**	0.122**	0.056	0.077
	(0.009)	(0.008)	(0.047)	(0.070)	(0.060)
Same language indicator	-0.059***	-0.098***	0.090	-0.127**	-0.049
	(0.021)	(0.018)	(0.056)	(0.056)	(0.048)
Obs	209,081	171,022	13,605	11,511	12,943
Control mean	-0.135	-0.071	-0.299	-0.340	-0.656

Notes: Same language is constructed using country of origin (of the child and/or parents) and the common official language indicator from the DICL database (Gurevich et al., 2024). Our constructed indicator takes the value 1 if the student's language is common to that of recent immigrants. The relevant reference group of recent immigrants are those arriving to the municipality up to 2 years before the intervention, with at least 5 students from the same country.

Table A27: School and classroom organization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	During intervention					Post period (3-4 years)					Post period (5-7 years)				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent
	Panel A: Class size														
Treated	0.266	0.118	0.855**	0.763*	0.504	2.712***	2.531***	1.462***	1.752**	2.449***	1.885***	1.918***	1.304***	1.674***	2.105***
	(0.299)	(0.296)	(0.425)	(0.445)	(0.443)	(0.422)	(0.482)	(0.328)	(0.741)	(0.709)	(0.370)	(0.418)	(0.419)	(0.351)	(0.330)
Obs	95,808	78,006	5,183	3,332	9,287	41,220	33,571	2,903	2,163	2,583	80,292	64,473	5,830	6,372	3,617
Control mean	23.373	23.370	24.029	23.968	22.843	24.408	24.345	25.537	24.526	23.726	23.099	23.128	23.441	22.744	22.635
	Panel B: Share recent immigrants in class room														
Treated	0.001	0.001	0.010**	-0.002	-0.004	0.024***	0.027***	0.007	0.042***	0.013	0.003*	0.004*	0.003	0.003	-0.016***
	(0.002)	(0.003)	(0.005)	(0.006)	(0.009)	(0.006)	(0.007)	(0.009)	(0.006)	(0.011)	(0.002)	(0.002)	(0.004)	(0.003)	(0.005)
Obs	95,808	78,006	5,183	3,332	9,287	41,220	33,571	2,903	2,163	2,583	80,292	64,473	5,830	6,372	3,617
Control mean	0.099	0.083	0.110	0.131	0.223	0.058	0.047	0.070	0.082	0.175	0.043	0.037	0.049	0.060	0.126
	Panel C: School has a reception class														
Treated	0.052	0.044	0.030	0.127***	0.070*	-0.043	-0.031	-0.074**	-0.055	-0.120**	-0.024	-0.018	-0.036*	-0.057**	-0.065**
	(0.034)	(0.033)	(0.041)	(0.039)	(0.042)	(0.033)	(0.029)	(0.036)	(0.037)	(0.056)	(0.020)	(0.020)	(0.019)	(0.028)	(0.030)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	0.164	0.151	0.169	0.247	0.241	0.066	0.062	0.068	0.080	0.119	0.036	0.032	0.042	0.053	0.061
	Panel D: School has class with only Swedish as 2nd language														
Treated	-0.001	-0.001	-0.011	0.009	-0.007	0.090***	0.089***	0.128***	0.063***	0.086***	0.056*	0.053*	0.065**	0.061	0.039
	(0.007)	(0.006)	(0.012)	(0.008)	(0.012)	(0.016)	(0.017)	(0.017)	(0.018)	(0.018)	(0.032)	(0.031)	(0.028)	(0.040)	(0.040)
Obs	95,808	78,006	5,183	3,332	9,287	41,220	33,571	2,903	2,163	2,583	80,292	64,473	5,830	6,372	3,617
Control mean	0.025	0.022	0.031	0.033	0.050	0.013	0.011	0.017	0.018	0.027	0.044	0.044	0.023	0.039	0.079
	Panel E: Age for grade														
Treated	-0.008*	0.000	0.010	-0.024	-0.031	0.003	-0.003***	0.009	-0.021	0.105***	0.004**	0.000	0.022***	0.043***	-0.018
	(0.004)	(0.001)	(0.007)	(0.020)	(0.021)	(0.002)	(0.000)	(0.005)	(0.027)	(0.023)	(0.002)	(0.001)	(0.004)	(0.009)	(0.020)
Obs	96,029	78,006	5,183	3,332	9,508	41,274	33,571	2,903	2,163	2,637	80,332	64,473	5,830	6,372	3,657
Control mean	0.946	0.987	0.975	0.837	0.618	0.967	0.992	0.982	0.887	0.665	0.968	0.992	0.985	0.866	0.680

Notes: A special reception class (Panel C) is an indicator that the school has a class with more than 90% recent migrants. Panel D is an indicator with value 1 if the school has a separate class for Swedish as a second language (more than 90% in that grade). Panel E presents effects on age for grade which is an indicator variable taking the value 1 it the student does attends a grade at the intended (or lower) age. Class identity is retrieved from the student registry. Note that not all students in test score data are identified in the student registry, thus they are excluded from this analysis. Excluded students are primarily asylum seekers taking the test. However, asylum seekers are included in the outcome variable share of recent immigrants in Panel B, but asylum seekers in the student registry are accounted for in the share of foreign born students. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

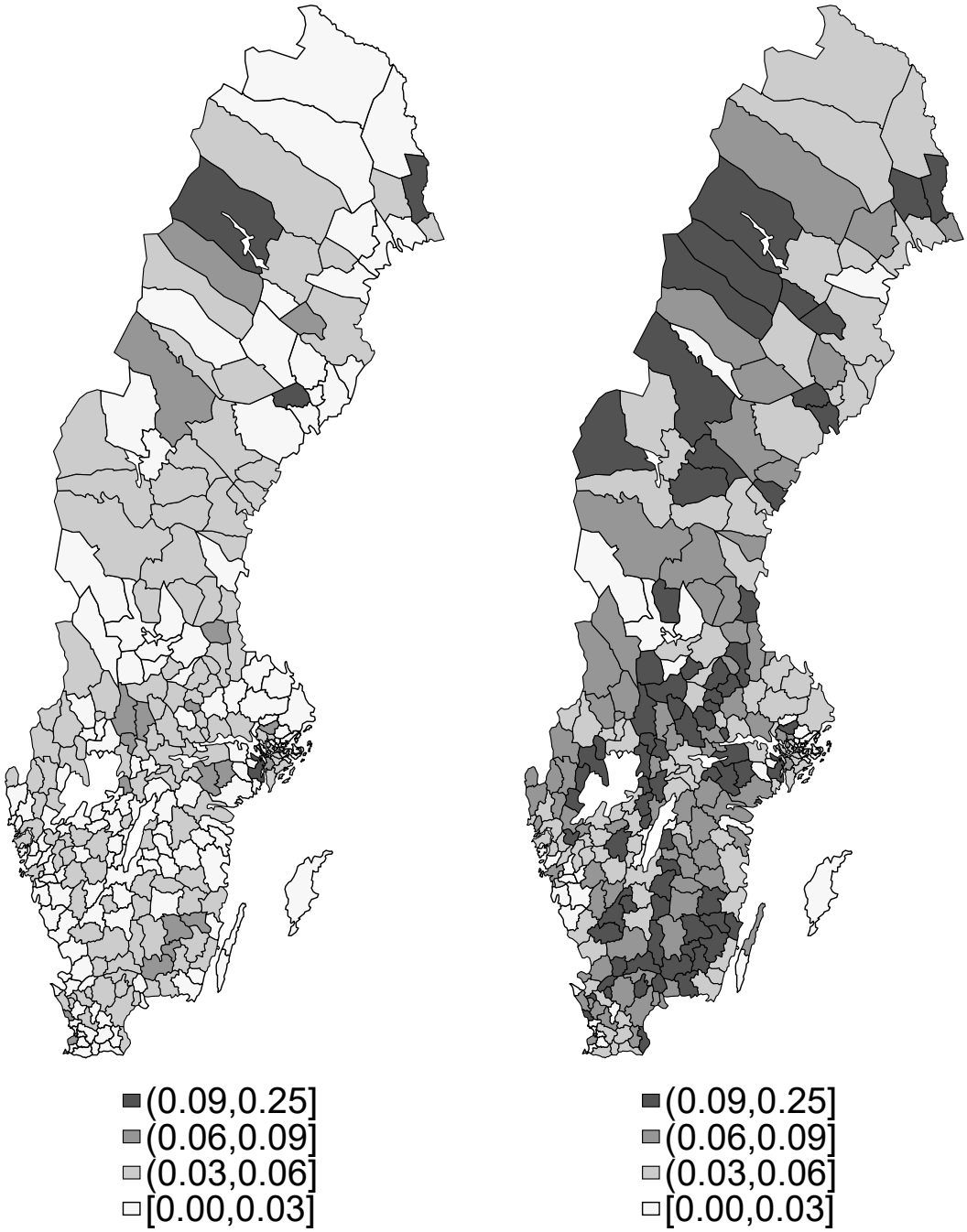
Table A28: Teacher characteristics in school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	During intervention					Post period (3-4 years)					Post period (5-7 years)				
	All	Swedish background	Foreign background			All	Swedish background	Foreign background			All	Swedish background	Foreign background		
			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent			Second gen	Non-recent	Recent
Panel A: Teacher-student ratio															
Treated	-0.050	-0.045	-0.050	0.010	-0.091	0.271*	0.262**	0.118	0.480**	0.296	0.031	0.023	-0.034	0.130	0.059
	(0.057)	(0.057)	(0.084)	(0.074)	(0.063)	(0.145)	(0.121)	(0.261)	(0.215)	(0.248)	(0.118)	(0.119)	(0.131)	(0.123)	(0.148)
Obs	95,665	77,704	5,168	3,324	9,469	41,193	33,519	2,897	2,162	2,615	80,156	64,337	5,824	6,364	3,631
Control mean	8.598	8.545	8.594	8.930	8.933	8.241	8.192	8.262	8.481	8.704	8.278	8.206	8.240	8.761	8.797
Panel B: Share certified teachers															
Treated	0.008*	0.009**	0.004	0.003	0.003	-0.000	0.000	0.018***	-0.005	-0.011	0.009	0.010	0.003	0.004	0.017**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.003)	(0.004)	(0.005)	(0.005)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)
Obs	95,812	77,819	5,178	3,329	9,486	41,238	33,557	2,903	2,162	2,616	80,299	64,464	5,830	6,371	3,634
Control mean	0.841	0.844	0.857	0.818	0.820	0.866	0.869	0.866	0.852	0.834	0.850	0.852	0.862	0.834	0.833
Panel C: Average experience															
Treated	0.194**	0.198**	0.036	0.358**	0.225**	0.079	0.192*	-0.343**	-0.222***	-0.618***	0.114	0.191*	-0.448**	-0.042	0.167
	(0.086)	(0.089)	(0.133)	(0.142)	(0.111)	(0.086)	(0.109)	(0.151)	(0.080)	(0.107)	(0.101)	(0.098)	(0.173)	(0.171)	(0.173)
Obs	95,812	77,819	5,178	3,329	9,486	41,238	33,557	2,903	2,162	2,616	80,299	64,464	5,830	6,371	3,634
Control mean	12.962	12.951	13.196	12.855	12.963	13.754	13.763	13.699	13.971	13.524	13.747	13.744	13.832	13.823	13.548

Notes: During intervention regards the first two years since program was initiated, Post period is year 3-7. Teacher-student ratio (Panel A) is the number of teachers (full time) per 100 students. Certified (Panel B) is an indicator of teachers having the appropriate education, Average experience (Panel C) is the number of years being a teacher. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Appendix B: Additional figures

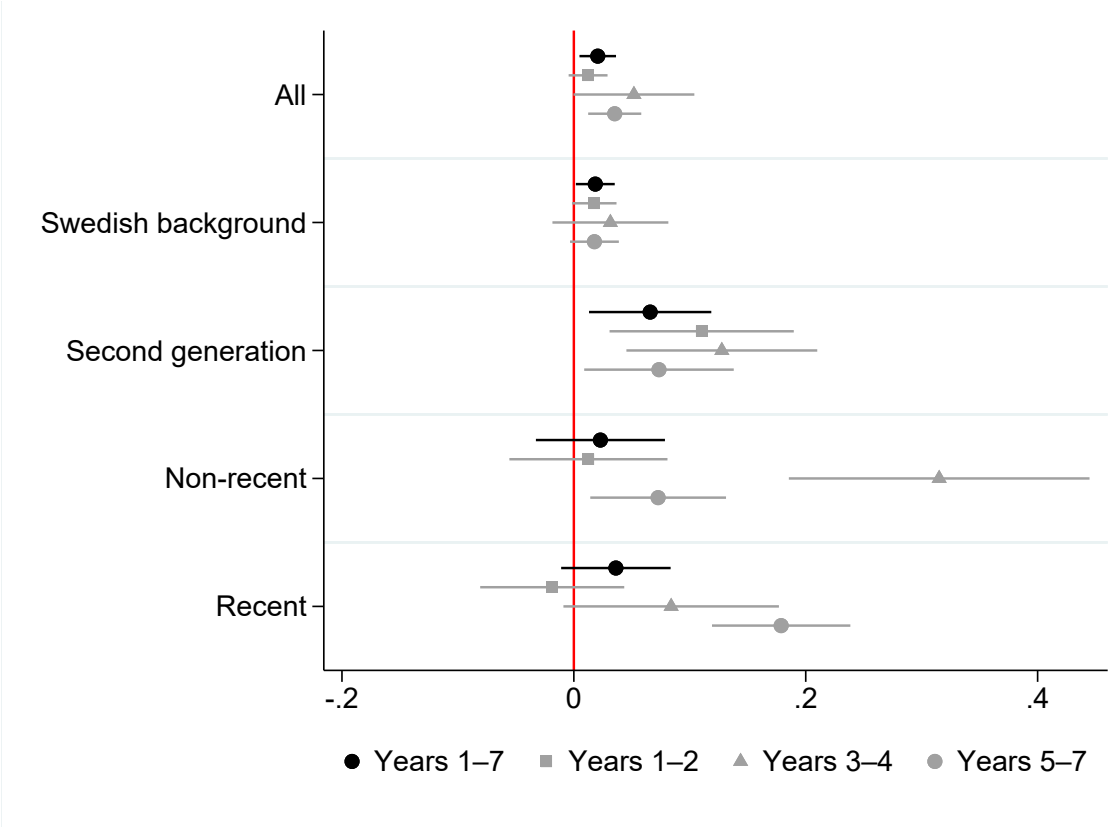
Figure B1: The fraction of recent migrant and asylum seeking students in Swedish municipalities before and during the refugee crisis



(a) 2008-2014 (b) 2015-2017.

Note: The figures (a) and (b) show the fraction of recent migrant and asylum seeking students in Swedish municipalities in 2008-2014 and during the crisis years 2015-2017

Figure B2: Effect on test scores

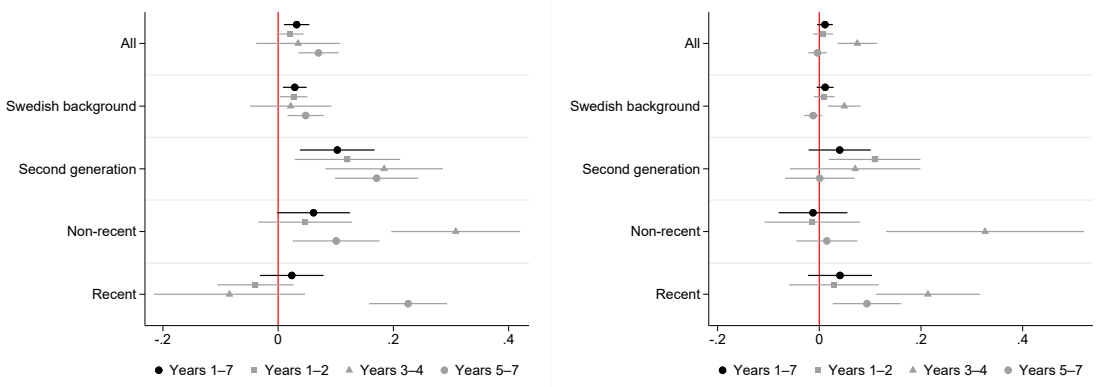


Notes: Test scores in total (black), years 1-2 (gray squares), 3-4 (gray triangles), and 5-7 (gray circles). Recent include asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Figure B3: Effect on test scores by subject

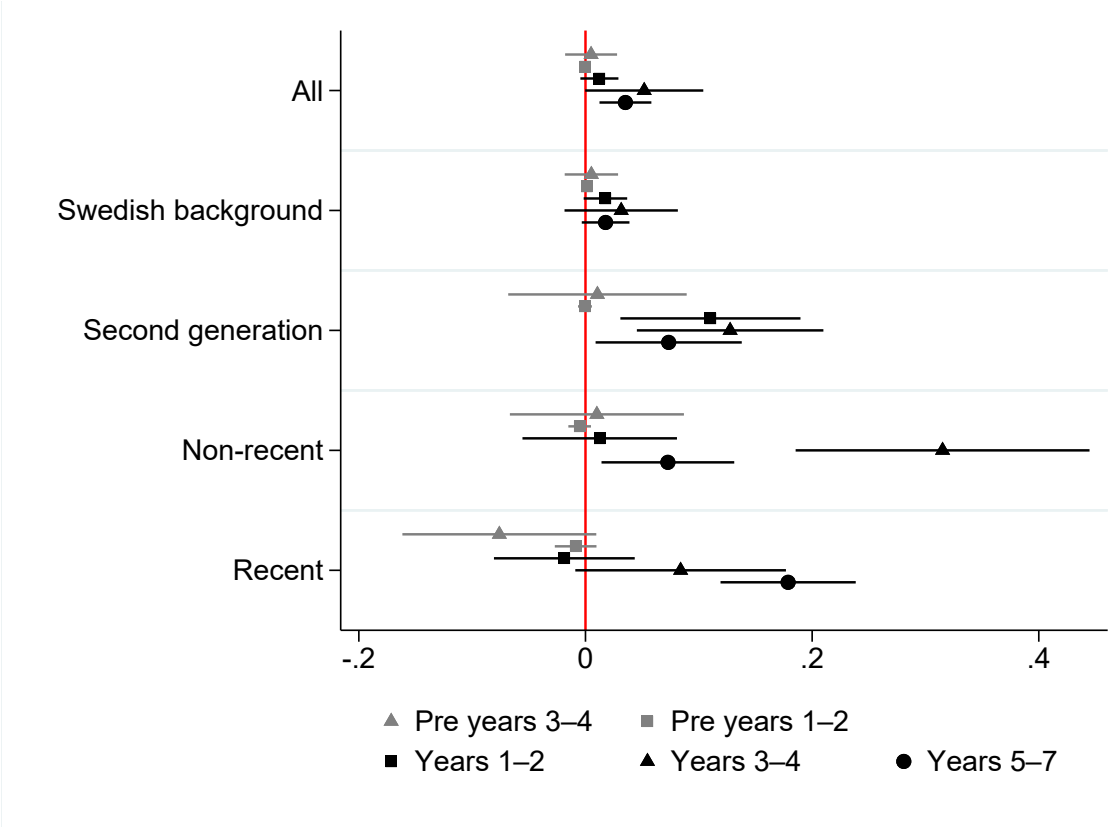
(a) Math test score

(b) Swedish test score



Notes: Test scores in total (black), years 1-2 (gray squares), 3-4 (gray triangles), and 5-7 (gray circles). Recent include asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

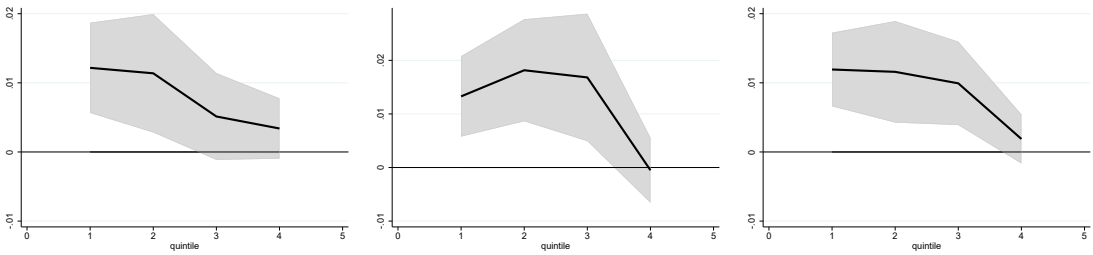
Figure B4: Effect on test scores, including pre-periods



Notes: Test scores before the intervention are indicated in gray; 1-2 years before (gray squares) and 3-4 years before (gray triangles). For post intervention test scores, years 1-2 (black squares), 3-4 (black triangles), and 5-7 (black circles). Recent include asylum seekers. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

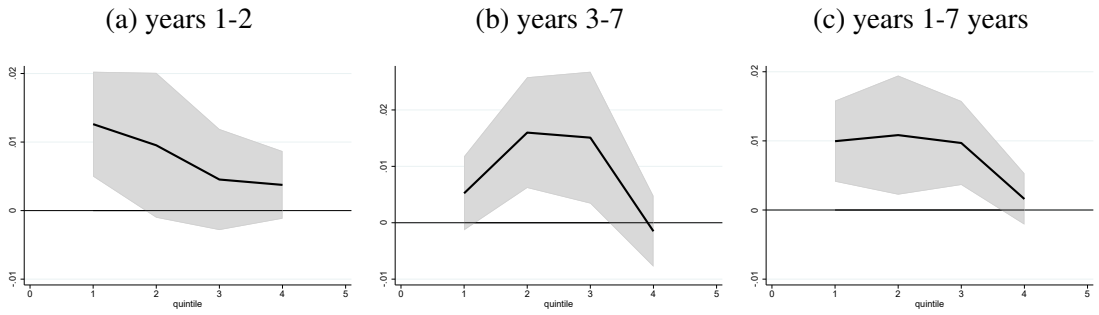
Figure B5: Effects on test scores over the test score distribution

(a) years 1-2 (b) years 3-7 (c) year 1-7



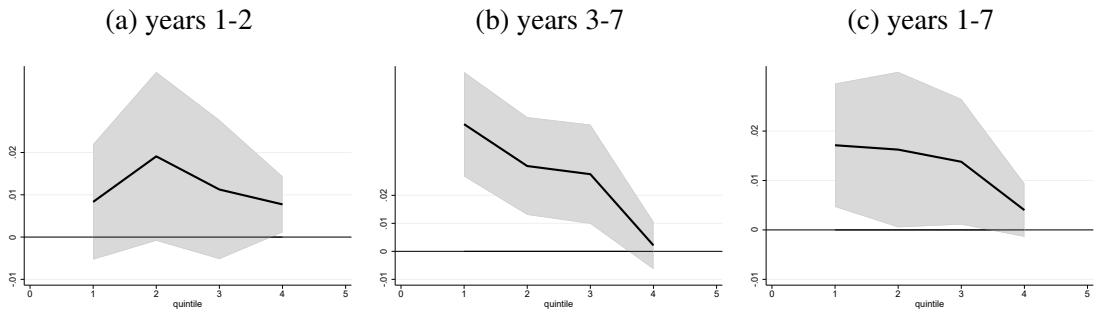
Notes: Figures display the estimates and confidence intervals from separate regressions where dependent variable is indicator for having test scores above quintile i=1,...,4.

Figure B6: Effects on test scores over the test score distribution, Swedish background students



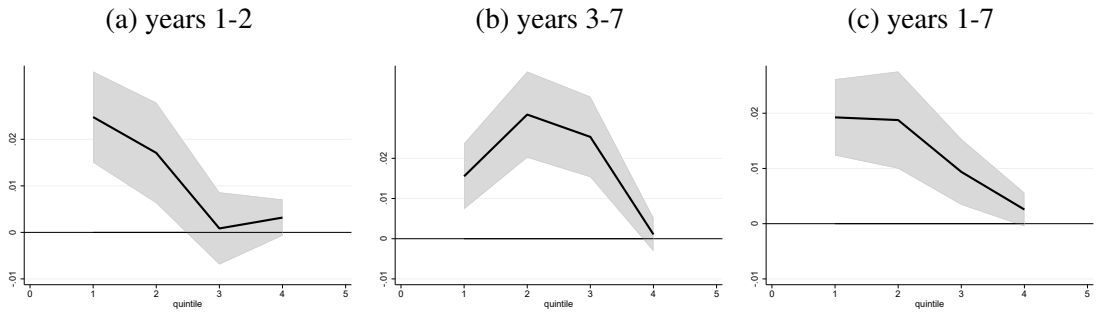
Notes: Figures display the estimates and confidence intervals from separate regressions where dependent variable is indicator for having test scores above quintile i=1,...,4.

Figure B7: Effects on test scores over the test score distribution, foreign background students



Notes: Figures display the estimates and confidence intervals from separate regressions where dependent variable is indicator for having test scores above quintile i=1,...,4. Foreign background include first and second generation immigrants.

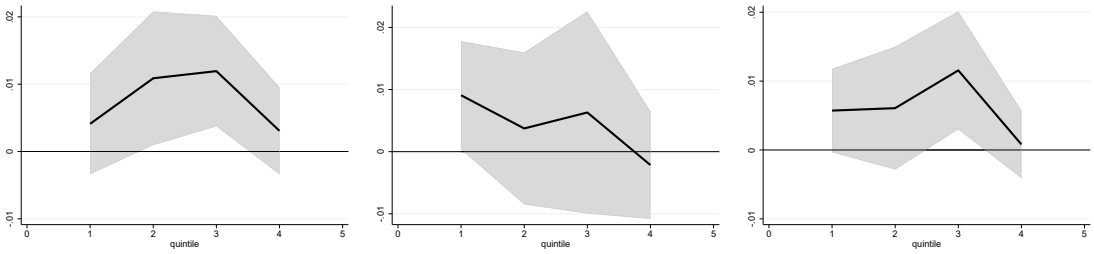
Figure B8: Effects on test scores over the test score distribution, boys



Notes: Figures display the estimates and confidence intervals from separate regressions where dependent variable is indicator for having test scores above quintile i=1,...,4.

Figure B9: Effects on test scores over the test score distribution, girls

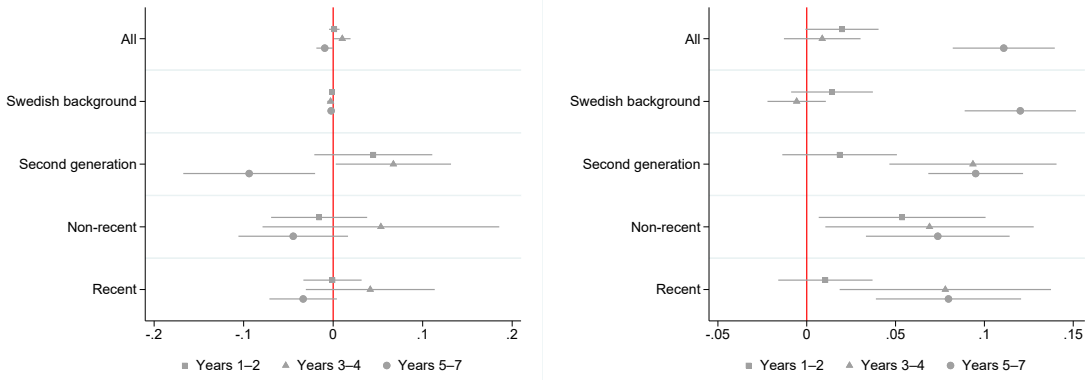
(a) 1-2 years (b) 3-7 years (c) 1-7 years



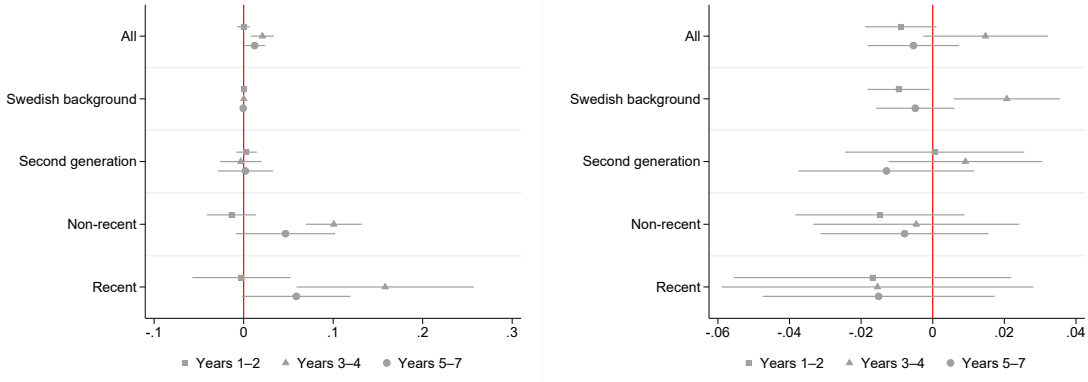
Notes: Figures display the estimates and confidence intervals from separate regressions where dependent variable is indicator for having test scores above quintile i=1,...,4.

Figure B10: Effects on students' access to pedagogical resources by migration status

(a) Swedish as 2nd language (b) Home language

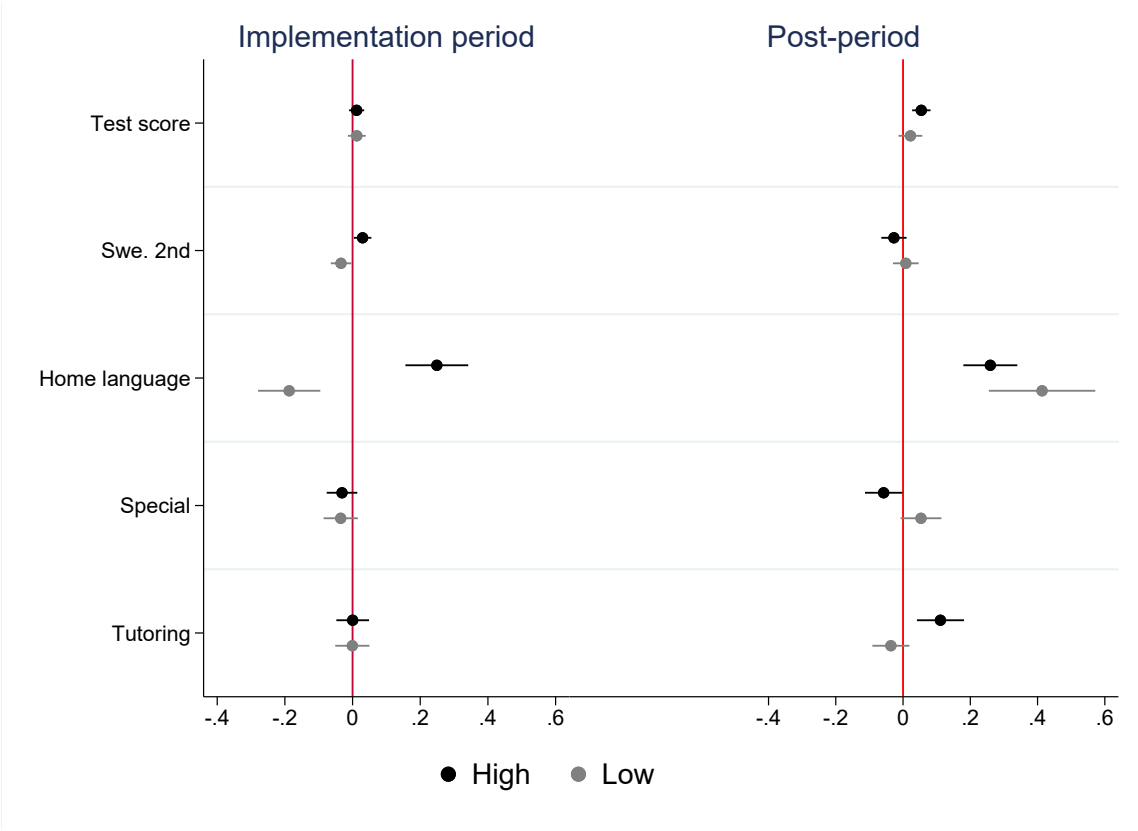


(c) Tutoring in mother tongue (d) Special aid



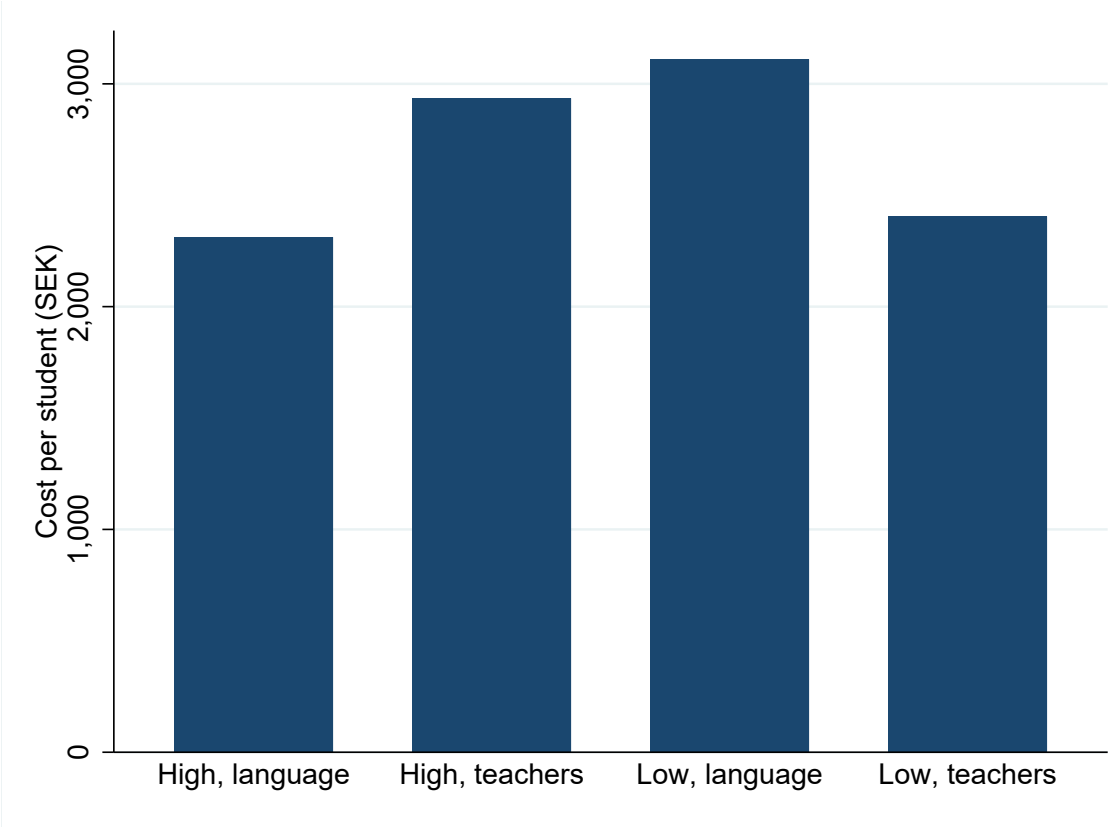
Notes: Years 1-2 (squares), year 3-4 (triangles), year 5-7 (circles). Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Figure B11: Effects on resources, by initial language supporting resources



Notes: Implementation period is years 1-2, post-period is years 3-7. Initial resources in terms of language support is an index of Swedish as 2nd language, home language class and tutoring. Initial resources in terms of special aid reflects receiving any type of special support that is not specifically targeting language. Resources are equally weighted and expressed at the municipal level, measured the year before implementation. The plotted estimates are the treatment status interacted with an indicator variable of high initial resources (black) and low initial resources (gray), respectively, defined by the median. The effects are expressed in terms of standard deviations. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

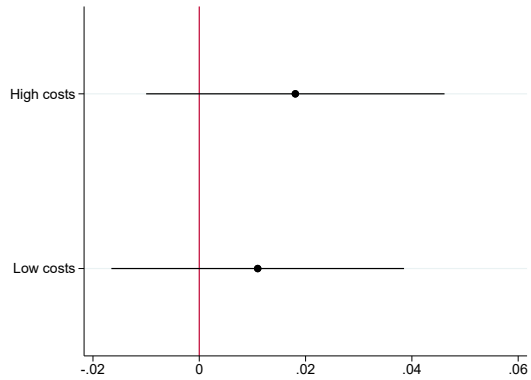
Figure B12: Program spending per student by initial resources



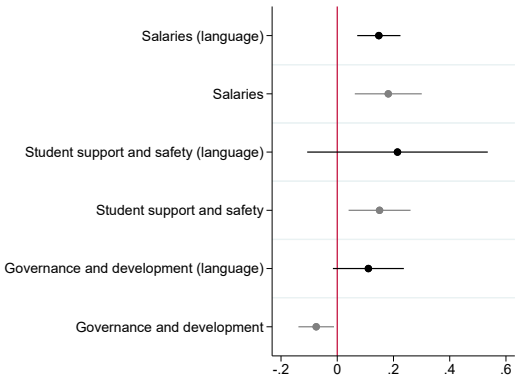
Notes: Language resources is an index of Swedish as 2nd language, home language class and tutoring. Teacher resources is an index of the share certified teachers, the teacher-student ratio and teacher experience, equally weighted and expressed at the municipal level. Both resources and number of students are measured in the two years before implementation. High reflects levels above the median, low reflects levels at the median and below.

Figure B13: Effect on test scores by level and allocation of program spending

(a) By high/low spending per student



(b) By allocation of budget



Notes: Each panel represents a single regression. In panel A, number of students are measured in the two years before implementation. High reflects levels above the median, low reflects levels at the median and below, both are interacted with the treatment dummy. In Panel B, each type of spending is included as the share interacted with treatment. Language related spendings in black, spendings unrelated to language in gray. Estimations include the full set of municipality controls as specified in Equation 1, excluding years 2019 and 2020.

Appendix C: Program modules

The program mainly focused on various types of professional development and training for teachers. The table below shows modules included in the program and whether the content focus on in-service training for teachers and school professionals, language learning, or if it's main focus was on administration and organization.

Table C1: Program modules

	(1) Training	(2) Language	(3) Admin
Language and learning enhancing	x	x	
Team and process management			x
Tutoring	x	x	
Other training	x	x	
Routines			x
Student health	x		
Quality enhancing practicies	x		x
Intercultural classrom	x		x
Parental involvement			x
Swedish as 2nd language	x	x	
Home language instruction	x	x	
Grading and assessment	x		
IT and digitalization	x		
Student transitions			x
Collegial learning	x		
Adult students	x	x	
Integration	x		
Learning difficulties	x	x	
Inclusion	x		

Language and learning enhancing

The primary component of our intervention involved teacher training in knowledge and language-enhancing teaching strategies aimed at supporting the language development and learning of migrant students. This approach, known as Scaffolding Language (Språk-

och kunskapsutvecklande arbetssätt, SKUA), draws upon various theoretical models and encompasses research from diverse fields. Courses have been provided at universities or colleges, such as Malmö University, Jönköping University, or the National Center for Swedish as a Second Language at Stockholm University (NC). SKUA can be offered to teachers in all grades, preschool, after-school care, and special education. Since SKUA is not a unified research field, it does not involve a single method but is described more as an approach or attitude geared towards teaching students whose first language is not the language of instruction. The purpose of this approach is to provide students with the conditions to develop their language skills as much as possible while integrating language and subject content. Another goal is for students to be as linguistically active as possible during lessons.

Team and process management

Most municipalities established a local team to lead, support, and coordinate the implementation of interventions in the action plan. This usually also involves the employment of a project leader or process support with a similar function.

Tutoring

Training or other initiatives for study counselors, who tutor students in their native language.

Other training

Other training includes among other things, professional development for study and career counselors, in-service training regarding laws regulating the learning of newly arrived students for various target groups (including politicians), language development in preschool; external lectures on cultural competence; in-service training in individual study plans; a course in cultural analysis; professional development for principals to enhance the learning of newly arrived students; a conference on multilingualism; a lecture on "Newly arrived's Learning"; multilingualism in preschool; leadership training for school leaders; communication tools for preschool staff; value-based work; academic success of newly arrived students (course in collaboration with Karlstad University); an inspiration day focusing on newly arrived and multilingual students' learning; cultural identity in preschool; trauma informed care (training in collaboration with SNEA, Save the Children, and the municipality).

Routines

Routines involve interventions within various areas, e.g. developing concrete measures to reduce students' school absenteeism, more individualized planning of instruction for newly arrived students, or developing municipality-wide routines for each school form

(primarily related to integration procedures). Overall, it may involve analysis and mapping at the organizational level regarding the learning of newly arrived students from various aspects and developing structures for mapping and creating quality-assured individual study plans.

Student health

In-service training or other initiatives for student health personnel. This may involve guidance for the staff, trauma-informed care, and regular training.

Quality enhancing practices

The systematic quality work is a requirement in the Education Act. According to SNEA, systematic quality work is a development effort that should be linked to conditions, work processes, and the organization of teaching. It begins with a description of the current situation and information about various types of results and goal fulfillment, such as students' results on national tests, other exams and tests, and grades that can be presented at the group-, class- and school levels. It may also include documentation of what implemented development efforts have led to and what conditions exist in the organization to carry out the mission based on national goals, requirements and guidelines. Within the program, investments in systematic quality work may involve courses or guidance for operational managers, school leaders, teachers and other educational personnel to develop the ability to follow up and evaluate teaching, sometimes with a special focus on the learning of newly arrived immigrants. It may also involve appointing (or allocating part of) a position to follow this. This includes in-service training and new methods for monitoring and analyzing the results of newly arrived and multilingual students in preschool, primary school and upper-secondary education.

Intercultural classroom

Intercultural education or cultural competence is conducted as courses at the university level and is offered at several universities or colleges.

Parental involvement

This module is about efforts to involve parents or guardians. These may include initiatives to enhance parent meetings, providing information to parents or guardians about the Swedish school system, parent education for guardians of newly arrived students, expanded parental support within preschool with targeted information and dialogue regarding children's basic needs, the parental role and issues related to collaboration between preschool and home. This could also involve parental collaboration or providing education for guardians on Swedish school culture and educational perspectives to increase participation in parent meetings, absence notifications and the use of school platforms.

Swedish as a 2nd language

In-service training or other initiatives for teaching Swedish as a second language. It often involves costs for substitutes during the training period and aims to provide teachers who teach the subject with formal qualifications in Swedish as a second language.

Home language instruction

In-service training or other initiatives for teachers who teach pupils in their home language.

Grading and assessment

Under grades and assessment, among other things, we find training in the SNEA's assessment material "Building Swedish". "Building Swedish" is an assessment tool with descriptions of students' language development in five stages. We also include lectures and guidance for teachers regarding grades and assessment under this heading. The purpose of this training is to strengthen teachers' ability to make equitable assessments and grading of newly arrived students.

IT and digitalization

Interventions related to information and communication technology (ICT). This may involve in-service training for teachers to enhance the quality of using ICT in the instruction of newly arrived students, as well as reviewing software and digital services for remote teaching.

Student transitions

Under this category, we include initiatives for collaboration between different organizations within the municipality. It may involve collaboration between primary and secondary schools regarding newly arrived students. It may also include instances where collaboration between different professional groups is discussed to reach students and parents for specific purposes.

Collegial learning

Under this heading we include initiatives involving collegial learning. Primarily, it seems to involve senior teachers implementing development initiatives among their colleagues, or involve groups of employees within a school form collectively implementing an educational effort, or similar.

Adult students

This heading includes, for example, in-service training for teachers in adult education or interventions related to Swedish for Immigrants (SFI) instruction.

Coaching

Coaching seems to be a tool to support teachers and preschool educators in their professional roles and practices. For example, it may involve being able to professionally meet the various needs of students and parents whitin the framework of the preschool and school's mission, but it is also mentioned as support during reorganization.

After school care

When municipalities invest in after school centers or their staff.

Integration

Learning difficulties

In-service training regarding reading and writing difficulties or basic literacy.

Inclusion

Integration into wider society.