Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1474

Does in-service training in special education increase the effectiveness of general classroom teachers? Evidence from the large-scale "Special Education Pedagogy for Learning" program in Sweden

Leah Natasha Glassow
Goteborgs Universitet

Nils Kirsten
Uppsala University

Causal evidence on the effectiveness of general teacher training in special education is scarce. This study evaluates the "Special Education Pedagogy for Learning" (SFL) program, a professional development program for general classroom teachers in Sweden involving approximately 60,000 teachers. Linking program administrative data to full-population register data, we estimate dynamic difference-in-differences effects of school-level participation on mathematics and Swedish grades and national test scores. Effects were generally small and statistically indistinguishable from zero, including for students with special educational needs, a key target group. Effects on Swedish national test scores for newly arrived immigrants were statistically significant but not robust to alternative specifications or multiple-testing corrections. The study therefore finds little evidence that this whole-class, inclusion-oriented program improved student achievement.

VERSION: November 2026

Suggested citation: Glassow, Leah Natasha, and Nils Kirsten. (2026). Does in-service training in special education increase the effectiveness of general classroom teachers? Evidence from the large-scale "Special Education Pedagogy for Learning" program in Sweden. (EdWorkingPaper: 26-1474). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/p0s9-9131

Does in-service training in special education increase the effectiveness of general classroom teachers? Evidence from the large-scale "Special Education Pedagogy for Learning" program in Sweden

Leah Natasha Glassow* Nils Kirsten†

November 11, 2026

Abstract

Causal evidence on the effectiveness of general teacher training in special education is scarce. This study evaluates the "Special Education Pedagogy for Learning" (SFL) program, a professional development program for general classroom teachers in Sweden involving approximately 60,000 teachers. Linking program administrative data to full-population register data, we estimate dynamic difference-in-differences effects of school-level participation on mathematics and Swedish grades and national test scores. Effects were generally small and statistically indistinguishable from zero, including for students with special educational needs, a key target group. Effects on Swedish national test scores for newly arrived immigrants were statistically significant but not robust to alternative specifications or multiple-testing corrections. The study therefore finds little evidence that this whole-class, inclusion-oriented program improved student achievement.

Keywords: large-scale teacher professional development, special and inclusive education, special educational needs, childhood immigrants, event study, difference-in-differences

*University of Gothenburg. Email: leah.glassow-hill@gu.se
†Uppsala University

1 Introduction

High-quality teaching has long-lasting impacts on student outcomes (Chetty, Friedman, & Rockoff, 2014)—and consequently, in Sweden and many countries around the world, government bodies have employed large-scale teacher professional development (PD) programs in response to declining test scores (Dee & Jacob, 2011; Grönqvist, Öckert, & Rosenqvist, 2025; Holmlund, Häggblom, & Lindahl, 2024). Many studies have evaluated programs targeting vulnerable student groups, including those with a foreign language or immigrant background, as well as low-achieving students more broadly (Gamse, Bloom, Kemple, & Jacob, 2008; Jacob & Lefgren, 2004; Machin & McNally, 2008; Yoshikawa et al., 2015). Less often, however, have large-scale PD programs targeting students with special educational needs or using special or inclusive educational approaches been rigorously evaluated (Donath, Luke, Graf, Tran, & Götz, 2023).

On-the-job training in this endeavor may provide much needed support to teachers, given that many students with special educational needs attend general classrooms (Cooc, 2019; OECD, 2015b, 2021). Additionally, expanding migration flows have led to increasingly linguistically and culturally diverse classrooms in many countries, including in Sweden (Pinson, Bunar & Devine, 2023). Inclusive education principles and approaches, designed to accommodate a wide range of learning needs within the general classroom, may be increasingly relevant not only for students with disabilities, but for supporting migrant and multilingual students as well (Eurydice, 2019, 2023).

Although the evidence for supplementary teaching and special education designation is strong (Coffey et al., 2026; Gersten, Haymon, Newman-Gonchar, Dimino, & Jayanthi, 2020; O'Hagan & Stiefel, 2024; Schwartz, Hopkins, & Stiefel, 2021), the ways in which general education teachers can be aided to support students with special educational needs or multilingual backgrounds are less clear (Bal & Perzigian, 2013; Carroll et al., 2017; D. Fuchs, Gilmour, & Wanzek, 2025; L. S. Fuchs et al., 2015). Some special education researchers have advocated that general education teachers choose methods that avoid placing unnecessary demands on skills like working memory and phonological processing, which some students may struggle with (L. S. Fuchs, Sterba, Fuchs, & Malone, 2016). Others advocate for within-class differentiation in which students with varying needs may be directed to learn different content, using different methods, working in different learning environments, or demonstrating their knowledge in different ways (Kahmann, Droop, & Lazonder, 2022). In contrast, still others caution that excessive within-class differentiation may undermine inclusion and instead advocate common instructional goals, combined with responsive teaching and integrated forms of support (Florian, 2019; Jordan, Schwartz, & McGhie-Richmond, 2009; Kalinowski, Gronostaj, & Vock, 2019).

Teacher survey responses as well as qualitative evidence from a range of different countries show that

teachers in general classrooms struggle to implement special and inclusive education pedagogical approaches and support students with special educational needs or multilingual backgrounds in their classrooms (Bölte, Leifler, Berggren, & Borg, 2021; Cooc, 2019; Deunk, Smale-Jacobse, de Boer, Doolaard, & Bosker, 2018; Eurydice, 2019; Savolainen, Engelbrecht, Nel, & Malinen, 2012; Söderlund, Thoutenhoofd, Westman Andersson, Gerrbo, & Berhanu, 2024). Moreover, though many PD programs have aimed to support students with special educational needs, such as reading difficulties, dyslexia or ADHD (e.g., Chao, Sze, Chow, Forlin, & Ho, 2017; Didion et al., 2021; Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008; Zentall & Javorsky, 2007), causal evidence of more general special or inclusive education PD is rare—and all the rarer for programs implemented at scale (Donath et al., 2023).

In this study, we address this gap by evaluating the effectiveness of one such inclusive education PD program: the large-scale "Special Education Pedagogy for Learning" (*Specialpedagogik för Lärande,* hereafter, SFL) program, which was launched in 2016 and formally ended in 2025. SFL represented a national policy effort to embed special and inclusive education principles within general classroom teaching and thereby strengthen teacher capacity to address diverse student needs across a range of subjects. Tasked by the Swedish government, both the National Agency for Education (SNAE) and the National Agency for Special Needs Education (SPSM) were responsible for the development and rollout of the SFL program, in which about 60,000 Swedish teachers participated, and which cost 680 million SEK (roughly 70 million USD; Swedish National Agency for Education, 2025).

Via a dynamic difference-in-differences estimation strategy (Callaway & Sant'Anna, 2021) and by linking program administrative data to national full population register data across 4936 schools and just under 2 million students, we exploit the staggered rollout and determine whether the positive effects of SFL, should they exist, persisted over time or tapered off, as our data extend up to 8 years prior to and up to 6 years after the SFL program intervention. The main outcomes of interest are grades and national test scores in mathematics and Swedish reported in the national register at Statistics Sweden. We test average effects of SFL as well as program effects for students with higher likelihoods of benefitting from the intervention, such as those with individualized education programs, a foreign language background, or those newly arrived in Sweden. We also test heterogeneity based on gender, school achievement quartiles, average teacher qualification levels in the school, and school type. Causal identification of the difference-in-differences analysis relies on the assumption that achievement trajectories in treated and control schools would have evolved similarly over time in the absence of SFL. We address several threats to causal identification by conducting a series of sensitivity analyses to explore the robustness of the findings, using alternative model and treatment specifications, and examining grading leniency, program dosage, student compositional changes, and cohort- and grade-specific effects.

When effects were averaged across all students, the findings indicate that the SFL program produced statistically significant but very small effects on Swedish grades, but no effects in mathematics grades or in either subject in the national test scores, implying unstable effects or that the SFL program may have increased grading leniency. Moreover, the confidence intervals suggest that any average effects in outcomes pooled across all students were unlikely to be more than small in magnitude. Analyses based on the program's target subgroups (students with IEPs and students with foreign language backgrounds) suggested no effect of the program for students with IEPs across any outcome of interest, and no statistically significant effects for students with two foreign born parents. However, for childhood immigrants (arriving after birth or within 5 years of the outcome of interest), the analysis suggests that the SFL program led to improvements in Swedish national test scores of between 0.077 and 0.09 standard deviations, though these did not remain statistically significant after corrections for multiple hypothesis testing.

Beyond assessing the effectiveness of the large public investment in the SFL program, our study advances several strands of literature which we outline from most to least proximal. First, despite the proliferation of PD programs aimed at supporting students with special educational needs, rigorous evaluations are rare and typically small in scale. For example, while a meta-review by Donath et al. (2023) included 324 studies, only 17 percent of these had an experimental or quasi-experimental design and the review did not differentiate results from these studies from the cross-sectional or single-group studies. Similarly, a recent review of experimental and quasi-experimental studies on special education PD by Didion et al. (2026) identified only 9 studies which included effects on students. Evidence is further fragmented with reviews of differentiated-instruction PD documenting highly variable effects across programs and little explanation for why (Kahmann et al., 2022). Our study thus provides a rare evaluation of a large-scale special/inclusive education PD program targeted towards general classroom teachers, adding more robust evidence to a literature in which effects have seldom been assessed through high-quality research designs.

Second, we contribute to the empirical literature on the effectiveness of inclusive education practices within general classrooms. Across reviews, research on how best to support diverse learners in mainstream settings remains inconclusive. Research syntheses of differentiated instruction show that, although some studies report small to moderate positive effects, the evidence base is heterogeneous and difficult to generalize because implementations, outcome measures, and contrasts vary widely (Deunk et al., 2018; Smale-Jacobse, Meijer, Helms-Lorenz, & Maulana, 2019). Similarly, Carroll et al. (2017) note that while strong evidence exists for targeted interventions, much less is clear regarding effective everyday teaching and adaptations. In contrast, credible evidence points to the effectiveness of intensive, supplemental, and highly explicit instruction delivered in small groups or one-to-one formats (D. Fuchs et al., 2025). Our findings align with the broader pattern of unclear effects of mainstream approaches to support diverse learners: despite students with

individualized education programs being a stated target group of the SFL program, we observe consistently null effects for these students across all outcomes, persistent across a range of sensitivity checks.

At the same time, exploratory subgroup analyses suggested possible benefits for newly arrived childhood immigrants, another group emphasized in the program's aims. While positive effects were observed for Swedish national test scores, these findings were not robust across all specifications and were no longer statistically significant after correction for multiple hypothesis testing. Although a wide range of policies and practices are promoted across OECD countries to support immigrant students' integration, inclusion-oriented approaches among them (Choi & Lee, 2020; Choi, Mao, & Park, 2025; Eurydice, 2019; OECD, 2015a), the existing knowledge base is dominated by descriptive or correlational studies and provides little rigorous causal evidence on which pedagogical approaches actually improve academic outcomes for this group (Bal & Perzigian, 2013). While our findings should indeed be interpreted cautiously given the exploratory nature of the subgroup analysis, the present study nevertheless provides one of the few large-scale causal evaluations of an inclusion-oriented teacher professional development program targeting immigrant students.

Third, the mostly null effects from our study add to a growing body of research indicating that maintaining robust effects of PD programs at scale is challenging. Reviews such as Kraft, Blazar, and Hogan (2018) interpret the weaker results of large-scale initiatives as stemming from lower intensity, reduced targeting, and implementation heterogeneity. The dynamic estimates in our study show that even the positive subgroup effects found (e.g., for newly arrived immigrants) tapered off within the post-treatment period. Together, these findings add to evidence regarding the difficulty of achieving and maintaining robust instructional improvements through large-scale PD models. They also complement evidence from prior national PD programs in Sweden, which similarly displayed small effects (0.01 to 0.03 SD), albeit reaching statistical significance (Grönqvist et al., 2025; Holmlund et al., 2024).

The article is structured as follows. In section 2 we present an overview of the context and content of the SFL program in Sweden, section 3 presents the data sources, treatment definition(s), and balancing and pre-trend tests, as well as the outcome variables. Section 4 presents the main identification strategy, heterogeneity analyses, and planned sensitivity checks. In section 5, we present the main results for the pooled analyses, heterogeneity analyses, and discuss the results from a battery of sensitivity checks which are documented in detail in the supplementary materials. Section 6 discusses and contextualizes the findings and provides some guidelines for future PD oriented policy efforts.

2 The SFL program

2.1 Context

Between the years 2003 and 2012, student performance in Sweden declined more rapidly than in any other country participating in PISA (OECD, 2015b). Moreover, in 2014, almost 15% of students in Sweden failed to qualify for upper-secondary school beginning at the age of 16—and overrepresented among this group were students with special educational needs, students with foreign language backgrounds, and newly arrived students (Regeringen, 2015, Regeringsbeslut U2015/05783/S). In response to this decline in achievement, the Swedish government tasked the SNAE with developing a series of one-year large-scale professional development programs, including the Boost for Mathematics (BFM), the Boost for Reading (BFR), and the SFL program which began in 2016.

Like both BFM and BFR, the SFL program was large in scale—and the Swedish government allocated approximately 680 million SEK (roughly 70 million USD) towards its completion. Moreover, if each participating teacher spent at least 60 hours on SFL activities and roughly 60,000 teachers took part between 2016 and 2025, the program required about 3.6 million teacher-hours during this period. Participation in the SFL program was voluntary for schools and occurred during regular school working hours and locally at each school, supported through a targeted state grant administered by the SNAE. In addition to costs for developing the PD material and training local coaches, most funding was used to compensate for the time that coaches spent leading the collegial learning cycles. Depending on whether the school applied for participation with one or two teacher groups, the grant covered either 10 percent or 20 percent of the coach's employment position over one academic year.

The SNAE states in their description of the program that it does not advocate any one particular teaching method. Rather, the aim is to provide special education-oriented material that aids professional dialogues in schools about how to improve their work. The content of the SFL program reflects an ambition to teach students with a broad range of educational needs in the general classroom, including students with formal special educational needs, language difficulties, or who have recently immigrated to Sweden (Swedish National Agency for Education, 2025). In fact, the program explicitly states that all students form the target group, though the formal budgetary justification for SFL lists students with special educational needs and students with a foreign language background as target groups.

One large (Thoutenhoofd, Söderlund, Westman Andersson, Berhanu, & Gerrbo, 2020) and three small (Swedish National Agency for Education, 2025) evaluations of SFL have been conducted, all of them focusing on teacher attitudes and none of which involved a control group. The largest evaluation (Thoutenhoofd et al., 2020) used pre- and post-questionnaires from over 3,000 teachers, along with observations and interviews in 30

schools, to examine changes in attitudes, self-efficacy, and collaborative practices related to inclusion, as well as to analyze program content. Teachers generally reported more positive views of inclusive education and valued the collegial discussions, though many found the program content difficult to translate into practice. The three smaller evaluations also reported positive views on the program among supervisors and participants (Swedish National Agency for Education, 2025).

The SFL program has also been criticized in public debate especially for its mention of Universal Design for Learning (UDL), allegedly for lacking a firm evidence base (Expressen, 2023). Nevertheless, the SFL program was broad in scope, and included modules ranging from inclusion and participation, multilingualism, neurodevelopmental disorders, attention and the learning environment, and mathematics didactics, among others, which we expand upon in the following section.

2.2 SFL program content

In line with past nationally implemented PD programs in Sweden, SFL adopted the so-called 'collegial learning' model developed by SNAE in 2011 (Kirsten & Carlbaum, 2020). This model borrows elements from several professional development frameworks, including lesson study (Chen & Zhang, 2019), learning study (Pang & Ling, 2012), and Nordic professional development models such as 'research circles' (Holmstrand & Härnsten, 2003). In this collaborative, practice-oriented approach, the aim is that teachers improve their teaching practices in groups taking part in learning modules.

Each participating teacher group took part in SFL for one academic year. During that year, modules were undertaken sequentially, with compulsory schools needing to complete at least 2 modules per year. If schools did not implement the required number of modules, they could be forced to pay back the state administered grant. At participating schools, one or more teacher groups (on average consisting of 10 teachers, see Table 2) met weekly throughout the academic year. The sessions followed structured modules provided on the program website. Program materials included theoretical readings, videos, examples, and step-by-step guidance for both individual preparation and teacher meetings. Each module followed a structured cycle of (a) reading, (b) collegial discussion of the assigned text and joint lesson planning of teaching activities related to its theme, (c) classroom implementation of a teaching activity, and (d) joint reflection based on teachers' self-reports of the implemented activities (see Figure 1). Schools selected which modules to engage with based on local needs, though all needed to complete at least 1 of 3 of the foundational modules on inclusion. Typically, the school principal selected participating teachers, and the principal, the coach, and/or the teacher group decided which modules to complete.

A central component of SFL was the use of trained coaches who guided the teachers' professional

learning. Although the module content was housed within an online learning portal, the coaches and weekly meetings ensured a high level of face-to-face teacher training. Coaches were usually special education teachers, but certified teachers without a special education degree were also eligible provided they received additional training from the SNAE, though all coaches were offered training (Swedish National Agency for Education, 2025). The coaches' role was to lead teacher discussions according to instructions specified in the module content. While the coaches were typically employed at the school, this was not always the case, and sometimes groups of several schools collaborated together via the principal organizer (a municipality or the owner of the independent school[1]).

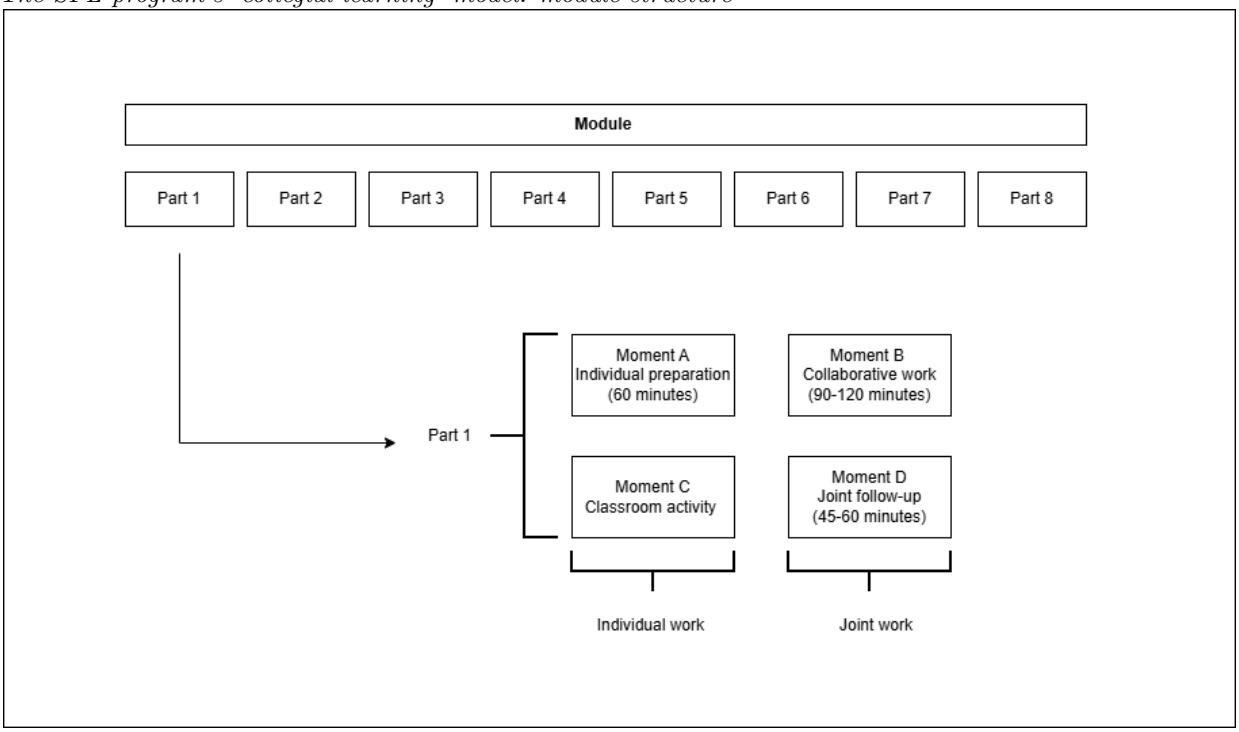
The materials of the SFL program were developed based on cooperation between the SNAE, the SPSM, and commissioned university researchers. The program materials included a wide variety of inclusive and special educational approaches and topics, though the vast majority had a strong focus on inclusive education topics and strategies. For descriptions of select modules, see Table 1. Examples of some of the intended outcomes of the foundational and mandatory modules on inclusion and participation include increased ability to conduct accessible and adapted teaching, strengthened student-teacher relationships, and a safer and more friendly school climate, which in turn, according to the program, should foster learning (Swedish National Agency for Education, 2025).

Examples of optional modules included 1) Participation in the Learning Environment, 2) Inclusion in School Practice, 3) Attention, Interaction, and Communication, 4) Multilingualism, 5) Neurodevelopmental Conditions, 6) Mathematics Didactics and Special Education, 7) Accessible Learning with Digital Tools, and 8) Systematic Work with Support in Teaching. Some explicit aims of these modules state for example, increased awareness of: how the environment shapes conditions for participation and strategies to foster a supportive learning environment, how varied instruction can meet diverse student needs, methods and strategies which support learning for children with attention difficulties, methods to promote motivation and school attendance, and methods to support students with language difficulties either due to a foreign language background or delays in second-language acquisition (Swedish National Agency for Education, 2025).

[1]In Sweden, independent schools may not charge fees for their students but can apply their own admission rules based on residential proximity or capacity.

Figure 1

The SFL program's 'collegial learning' model: module structure



Note. The figure shows the structure of SFL within an example module, which is divided into 8 subparts in which the cycle between moments (a) through (d) occur.

Table 1

Overview of content in select SFL modules provided by the SNAE

Module	SNAE description
Inclusion and participation (foundational)	The module aims to provide in-depth knowledge and understanding of what inclusion entails and how an inclusive learning environment and students' participation influence learning. The module offers tools for reflecting on, observing, testing, analysing, and discussing one's own teaching practice and organisational context. Furthermore, the module aims to develop the ability to design accessible and adapted instruction.
Interaction for inclusion (foundational)	A starting point of the module is the idea that strong relationships and effective communication form the foundation of a school that promotes well-being and learning for all students, as well as for school staff. Teachers' continuous learning and collaboration to improve and develop the school (collective learning) are foregrounded. Special educational challenges are viewed as a collective responsibility rather than as individual problems. When teachers jointly develop their communication and relational competence across the whole school, key teacher-student relationships can be strengthened, a safe and supportive social climate can be fostered, and a positive learning environment can be established.
Participation in the learning environment	The module is based on a model of participation that supports identifying, reflecting on, discussing, analysing, and developing the learning environment. Well-developed work on participation can support the creation of a successful learning environment that includes all students. Increased awareness and knowledge of how the learning environment shapes conditions and sets boundaries for students' opportunities for participation also enhances the potential for change and development.
Attention, interaction and communication	This module aims to provide in-depth knowledge of how teaching and learning environments can be designed in relation to students' variation in perception, cognitive abilities, executive functions, and communication. The content enhances understanding of the pedagogical consequences that difficulties in different areas may entail, while also presenting approaches, methods, and attitudes that support learning and interaction. The module also aims to develop and deepen professional relational competence that promotes motivation and school attendance.
Multilingualism	The module aims to provide in-depth knowledge and understanding of how teaching and learning environments can be adapted to students' variation from a multilingual perspective. The content presents approaches, methods, and attitudes that support the learning conditions of multilingual students, both in relation to expected second-language development and in cases where second-language development is delayed.
Neurodevelopmental conditions and accessible education	The purpose of the module is to enable teachers and staff involved in teaching to develop their knowledge of neurodevelopmental conditions. The module presents selected research within the neuropsychiatric field that provides guidance on how teaching and learning environments can be designed in relation to students' needs and prerequisites.

Note. The figure displays select modules from the SFL program with modules summaries provided by the SNAE (Swedish National Agency for Education, 2025).

3 Data and empirical approach

3.1 Data sources and measures

The study combines administrative data from the Swedish National Agency for Education on SFL program participation between the years 2016 and 2022 and full population register data from several national population registers from Statistics Sweden between 2014 and 2022. The first is the pedagogical register which contains individual level annually updated information on all individuals employed in Swedish schools. It includes teacher credentials, age, gender, employment information, position and role. We also link information from the pupil register, the grade point register and the national test register, which includes information on student migration background, grades and national test scores, as well as whether students have individualized education programs. The teacher and student data are linked at the school level, meaning although we cannot link individual students to teachers, we can identify school-level treatment effects. The school-level variables include its unique identifier, type (independent or public), and size.

We use data for each participating and never-participating compulsory school across 9 academic school years (2014/2015-2022/2023), amounting to 1,967,114 individual students in 4936 schools, as well as the full population of teachers (whose characteristics are averaged at the school level) in these schools.

As is typical in large-scale administrative datasets, some variables contain missing observations. Missingness is negligible or less than 3% for most variables (e.g., gender and individualized education programs) and modest for some background and outcome measures such as parental education and national test scores (approximately 5-10 percent).

3.2 Treatment definition and SFL program characteristics

The SNAE tracked which schools received the SFL state grants in terms of the number of coaches assigned to each school. Based on these data, combined with the population register on the full population of students, teachers, and schools, we can identify which schools were treated and how many coaches were assigned per school. The main treatment definition identifies a school as being treated if they have at least one coach assigned to the school, as identified by the SNAE. As shown in Table 2, treated schools hosted on average approximately 1.7 coaches, 2.4 teacher groups, and about 16 participating teachers.

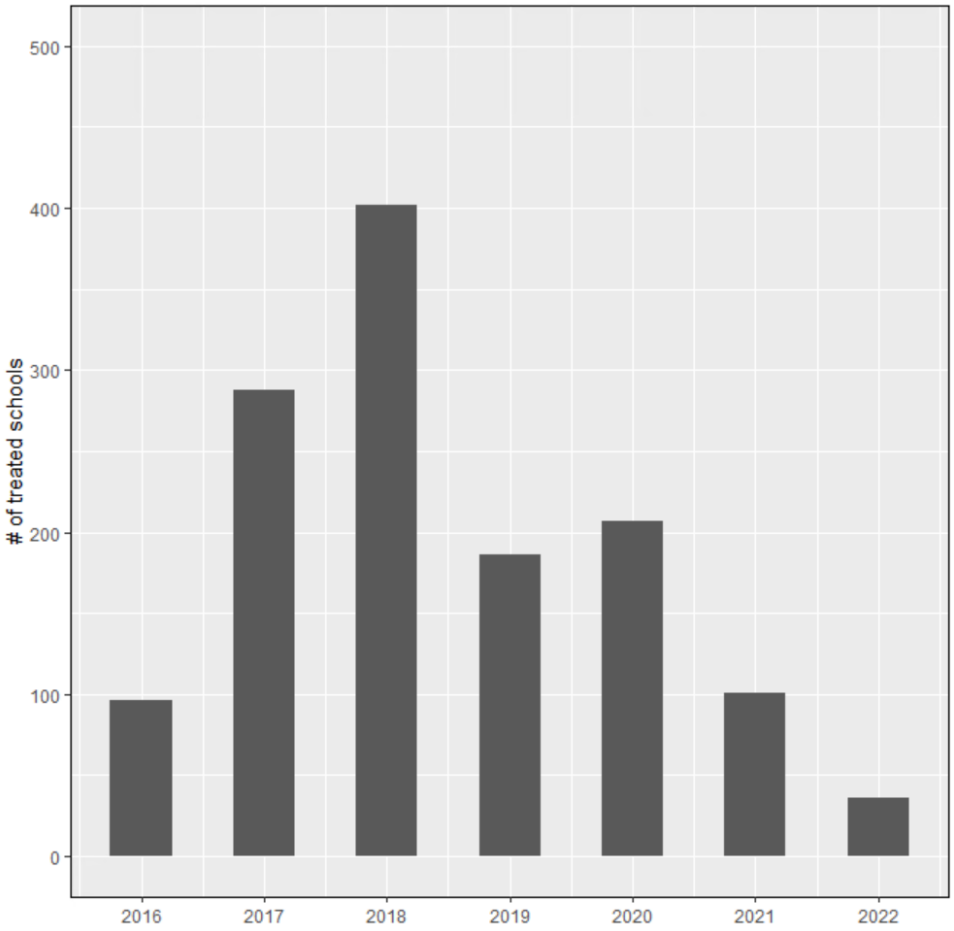
Based on this treatment definition, Figure 2 shows the staggered rollout in the total number of schools participating in SFL per year. The SFL dataset provided by the SNAE lists 1356 mainstream compulsory grade 1-9 schools which participated in SFL during the period studied, which corresponds to roughly one quarter of the schools in our data. One school was missing a school identification code and therefore could

not be identified. Additionally, 108 schools were identified by SNAE as having received the treatment but no information on whether the school was assigned a coach could be found. We therefore exclude these schools so that the number of treated schools in our analysis is 1247. However, we include these schools in an alternative treatment definition in a sensitivity test.

Unfortunately, we cannot test whether the effectiveness varied as a function of the proportion of participating teachers in a school, as these data are only available from the SNAE for a select few number of years (2019-2022). However, schools which served grade 6 or 9 students had an average of 25 teachers in total working at a school based on teacher register data aggregated to the school level. If the average number of teachers per school is representative of participation rates for the other SFL program years, this indicates that on average roughly 65% of teachers in the SFL schools participated in SFL. Consequently, any school-level effects estimated in this study are likely to be attenuated relative to a scenario in which all teachers in participating schools had received the intervention.

Figure 2

Participating grade 1-9 schools and rollout of SFL program



Note. The figure shows the number of schools participating in each wave of the SFL program between 2016 and 2022.

Table 2 shows the number of coaches per school, the average group size, the number of participating teachers, and the number of groups for participating mainstream grade 1-9 schools, based on data provided by the SNAE. When data for a variable were not available for all years, we based our calculations on the years with data and specify which years were used in the Table.

Table 2

SFL program characteristics of treated schools provided by the SNAE

	Data availability	Minimum	Maximum	Mean	SD
# of coaches	2016-2022	1	16	1.693	1.222
# of participants	2018-2022	1	86	16.45	11.47
# of groups	2018-2019	1	16	2.389	1.624
# teachers per group	2019-2022	1	40	9.996	5.664

Note. Participants refers to participating teachers.

3.3 Balancing tests and examination of pre-treatment trends

Table 3 presents a balancing test that compares the means of the pre-treatment treated schools as well as the never-treated schools prior to 2016. As expected in a large administrative dataset, several differences are statistically significant due to the large sample size and indicate slightly higher achievement, higher levels of parental education and lower levels of migrant and newly arrived students in the treated schools. However, the standardized magnitude of these differences was small. By contrast, the proportion of students with an individualized education program was virtually identical across treated and untreated schools.

For school-related variables, independent schools were substantially more common among treated schools than among never-treated schools. Treated schools were also more likely to be in the top achievement quartile and less likely to be in the bottom quartile of school achievement between 2014-2015. In addition, treated schools had slightly lower shares of certified teachers as well as slightly lower average years of teacher experience than never treated schools.

Table 3

Pre-intervention characteristics and balance tests of treated and never treated schools

	Never treated	Treated	*p*-Value	SMD
Achievement measures				
Math (national test score)	-0.01	0.03	<0.001	0.042
Math (grade)	-0.01	0.03	<0.001	0.039
Swedish (national test score)	-0.01	0.03	<0.001	0.051
Swedish (grade)	-0.01	0.03	<0.001	0.049
Student characteristics				
Individualized education program (ever)	0.20	0.19	0.034	0.077
Individualized education program (current)	0.06	0.06	0.121	0.002
Parental education (1= post-secondary education)	0.60	0.62	<0.001	0.050
Migrant background (2 foreign born parents)	0.19	0.17	0.077	0.062
Childhood immigrant	0.11	0.10	0.010	0.092
Arrived within 5 years	0.07	0.06	0.003	0.111
School characteristics				
School type (1=independent school)	0.13	0.25	<0.001	0.302
Share certified teachers	0.74	0.72	<0.001	0.142
Teacher experience (years)	13.46	13.11	0.004	0.099
Schools (n)	3689	1247		

Note. Parental education indicates the share of students whose parents have post-secondary degrees. Migrant background indicates they have two foreign born parents or are not born in Sweden. Both the national test scores and grades are standardized by each yearly cohort. SMD = standardized mean difference.

Although the patterns shown in Table 3 indicate that SFL-participating schools differed slightly from non-participating schools in several school- and student-related characteristics, difference-in-differences estimates remain unbiased in the presence of such baseline differences, so long as the parallel trends assumption is met. This assumption implies that, absent SFL, the trajectory of grades and test scores in mathematics and Swedish would have been the same for treated and never-treated schools—even though their levels may differ.

While there is no formal test for the parallel trends assumption, we assess its plausibility by examining pre-treatment trends across treated and never-treated schools for each outcome variable, using both visual inspection of the event study plots, pre-trend tests, and plot outcomes for treatment and control groups over the study period (see supplementary material section A).

3.4 Outcome measures

3.4.1 Grades and national test scores in mathematics and Swedish

To measure student achievement, we rely on two distinct measures: teacher-assigned grades in school years 6 and 9 and standardized national test scores in school years 6 and 9 in the subjects Swedish and mathematics. Teacher grades and national tests are scored at the scale F-A, where F is "fail." We translate this so that F is 0 and A is 5 and standardize both grades and student test scores (mean 0 and standard deviation 1) by year. We lack national test scores for 2019/2020 and 2020/2021 as no national tests were conducted in Sweden these years due to the COVID-19 pandemic. The missing test years reduce the number of observations for national test outcomes, and thus the precision of the estimates for post-treatment years, but should not otherwise bias the model estimates, since the cancellations applied nationwide and are therefore unrelated to treatment status.

We use both grades and national test scores as the outcome measures of interest for several reasons. In theory, teacher-assigned grades should reflect an overarching assessment of student general academic ability, based on continuous observation and evaluation over the academic year. They are also used for eligibility to upper-secondary education and from upper-secondary school to tertiary education, making them highly consequential for students' further education (Vlachos, 2019). By contrast, national tests are provided at the end of an academic year and graded according to detailed instructions.

While national tests are not externally graded, schools are required to anonymize student responses before grading (SFS 2011:185, 2011, Chapter 9, Section 22) using nationally provided grading rubrics. Thus, national tests are less prone to grading leniency than grades. For our empirical strategy, permanent differences in grading leniency across schools would not cause bias. However, if treatment changes the extent of grading

leniency, it would be a concern. We examine this by modeling grading leniency as an outcome in a sensitivity analysis.

3.4.2 Swedish as a second language

In Sweden, students with a first language other than Swedish may be assigned to "Swedish as a Second Language" (*svenska som andraspråk,* SVA) rather than the standard Swedish curriculum. Given the partial focus of the SFL program on students with a foreign language background and considering that the learning needs may differ within this group, we classify three categories of individuals who may fall into this category and may or may not study SVA—including students with 2 foreign born parents, students who were born outside of Sweden (arriving anytime before the SFL program) or students who were born outside of Sweden but arrived within 5 years of their respective outcome measures (grades or national test scores).

SVA-enrollment is determined by the principal's ongoing assessment of whether a student requires second-language-oriented instruction, a decision that is meant to be reviewed periodically rather than fixed at first enrollment (Skolverket, 2023). Because this assessment is conducted at the school level and is not centrally standardized, students with similar migration backgrounds may be placed differently depending on their school's assessment practices and available instructional capacity (Skolverket, 2023). In the analytic sample, 82.5% of newly arrived students take the Swedish national test in SVA, compared with 58.5% of childhood immigrants and 49.6% of students with 2 foreign born parents. Students who are enrolled in SVA take the same national test as students who are enrolled in Swedish as a first language. However, because the grading rubrics differ slightly, we also estimate separate analyses for students enrolled in Swedish and SVA as an additional sensitivity check.

4 Identification strategy

Due to the staggered rollout of SFL, approximately 25% of mainstream compulsory schools had participated in the SFL teacher professional development program between its inception in 2016 and the end of the 2022/2023 academic year. We exploit this staggered rollout to estimate dynamic effects of the SFL program on grade 6 and 9 achievement using the estimator proposed by Callaway and Sant'Anna (2021). This approach addresses concerns raised in recent evaluations of difference-in-differences analyses based on two-way fixed-effects models (Roth, Sant'Anna, Bilinski, & Poe, 2023). Specifically, it accommodates staggered treatment adoption and treatment effect heterogeneity, avoiding the bias that arises when earlier-treated schools serve as controls for later-treated schools in the presence of heterogeneous treatment effects. It does so by constructing separate cohort- and time-specific DiD comparisons rather than pooling all variation into a single regression coefficient.

The logic of the estimator can be illustrated with the canonical case of two groups (treated and control) and two periods (before and after), where the average treatment effect D is given by:

$$D = (\mathbb{E}[Y_{post} \mid T = 1] - \mathbb{E}[Y_{pre} \mid T = 1]) - (\mathbb{E}[Y_{post} \mid T = 0] - \mathbb{E}[Y_{pre} \mid T = 0]) \tag{1}$$

and T indicates treatment status. In the SFL setting, as in many others, however, there are multiple pre- and post-treatment periods, many treated units, and treatment begins at different times for different schools. Effects may also differ across schools and over time. To accommodate this, Callaway and Sant'Anna (2021) apply the same DiD logic separately to each treatment cohort and period, comparing each cohort to a common comparison group.

Formally, let Y_{ist} denote the outcome of interest for student i tested in school s in year t. Let G_s denote the year in which school s first participated (its treatment cohort g), with $G_s = 0$ for schools that never participated during the sample period. The data form a repeated cross-section of students nested within a panel of schools: each school is observed repeatedly, but the students tested differ across years. The expectations below are therefore taken over the population of students tested in each period. For each cohort g and year $t \geq g$, the comparison group consists of never-treated schools ($G_s = 0$), a choice supported by their large share of the sample. The cohort- and time-specific version of equation 1 is then:

$$ATT_{g,t} = (\mathbb{E}[Y_{ist} \mid G_s = g] - \mathbb{E}[Y_{is,g-1} \mid G_s = g]) - (\mathbb{E}[Y_{ist} \mid G_s = 0] - \mathbb{E}[Y_{is,g-1} \mid G_s = 0]) \tag{2}$$

where $ATT_{g,t}$ is the effect for cohort g in year t. The first term is the change in average outcomes among students tested in cohort-g schools between the pre-treatment reference year $(g-1)$ and year t; the second term is the same change among never-treated schools, which serves as the counterfactual trend. Each $ATT_{g,t}$ is thus a canonical 2×2 DiD comparison, estimated separately for every cohort and post-treatment year.

In a final step, we aggregate these cohort-time estimates into an overall average treatment effect on the treated,

$$ATT = \sum_{g} \sum_{t \geq g} w(g,t) \times ATT_{g,t} \tag{3}$$

where the weights $w(g,t)$ are proportional to cohort size. Cohorts treated earlier therefore receive more total weight, since they are observed for more post-treatment years (an issue to which we return in the sensitivity analysis). When we instead report effects by event time $\tau = t - g$, we average the $ATT_{g,t}$ across the cohorts observed at each τ, weighting each by its size. Standard errors are clustered at the school level. Finally, because treatment is defined by whether a coach was assigned to the school, and schools assigned a coach may have differed in how the program was implemented, we interpret the estimates as intent-to-treat (ITT)

effects.

4.1 Heterogeneity analyses

Given the broad mandate of the SFL program and its likelihood to disproportionately benefit several heterogeneous student groups, we conduct a number of heterogeneity analyses. First, we examine subgroups of students who are explicitly stated as the target of the SFL intervention: students with special educational needs (who have an individualized education program recorded in the national register) and students with a foreign language background (who either have two foreign born parents or who emigrated to Sweden at different times according to the national register). Given that these subgroups are stated targets of the SFL program, we consider these part of the main policy-relevant analysis to determine the overall effectiveness of the SFL-program.

We also conduct a series of more exploratory heterogeneity analyses, including across gender, given that boys are more frequently identified as having special educational needs (Giota, Lace, & Emanuelsson, 2023), school type (independent or publicly owned), school achievement quartiles, and the average level of teacher qualifications in the school. We identify schools in the top and bottom quartiles of student achievement and average teacher qualifications (experience and teaching certification) based on measures in the years prior to the SFL intervention (a two-year average based on 2014 and 2015).

4.2 Sensitivity analyses

We assess the plausibility of the parallel trends assumption (that treated and never-treated schools would have followed similar outcome trajectories in the absence of SFL) by testing for differential pre-treatment trends using an event study approach and a formal pre-trend test. Over and above this, we conduct several additional sensitivity checks to determine the robustness of our findings. Specifically, we investigate the degree to which our estimations of SFL program effects may be vulnerable to changes based on treatment definition and model specification (in the latter, we test models with school type as a covariate and therefore change the identifying parallel trends assumption to conditional on school type), student compositional changes, program dosage (by investigating SFL effects in small schools), as well as cohort- and grade- specific effects. We also investigate whether SFL influenced grading leniency by using grading leniency as the outcome in the difference-in-differences estimation.

5 Results

For each outcome measure, including mathematics and Swedish grades and national test scores, we present the results for the pooled sample. Next, we display subgroup analysis results for the main target groups of the SFL program: students with IEPs, with foreign language backgrounds or who are newly arrived to Sweden. Next, we present results from more exploratory heterogeneity analysis across school achievement quartile, school type and school-level teacher qualifications. Last, we present the results of the robustness and sensitivity tests.

5.1 Pooled analysis of SFL

As evidenced by the results shown in Table 4, we detect no clear pattern of effects across outcome measures when effects are averaged across all students. This is reflected by the very small effects in most cases, which are present for mathematics grades and national test scores as well as Swedish national test scores. We however find a small but statistically significant effect on students' Swedish grades, amounting to 0.023 standard deviations ($p = 0.026$).

Regarding the parallel trends assumption, Table 4 shows that the pre-trend test rejects the null hypothesis of no differential pre-trends for mathematics national test scores. However, the individual pre-treatment coefficients shown in the event study plots are very small in magnitude and not stable over time, suggesting that the rejection reflects minor changes rather than meaningful pre-trend differences.

In Figure 3, we present the estimated dynamic effects of SFL by event time (the time periods in relation to the treatment) in both Swedish and mathematics by grades (panel A) and national test scores (panel B). As in Table 4, Figure 3 shows that for all outcomes except Swedish grades, the effect of SFL on mathematics and Swedish is negligible when averaged across all students and grade 1-9 schools. Although the event-study plots show positive estimates after the treatment for both mathematics and Swedish grades which persist for several years, the post-SFL increases in grades are small, ranging from 0.01-0.04 standard deviations, and in most cases not reaching statistical significance. This pattern, however, is not reflected in the national test scores, which show consistently null estimates in the case of mathematics and very unstable estimates in the case of Swedish.

Table 4

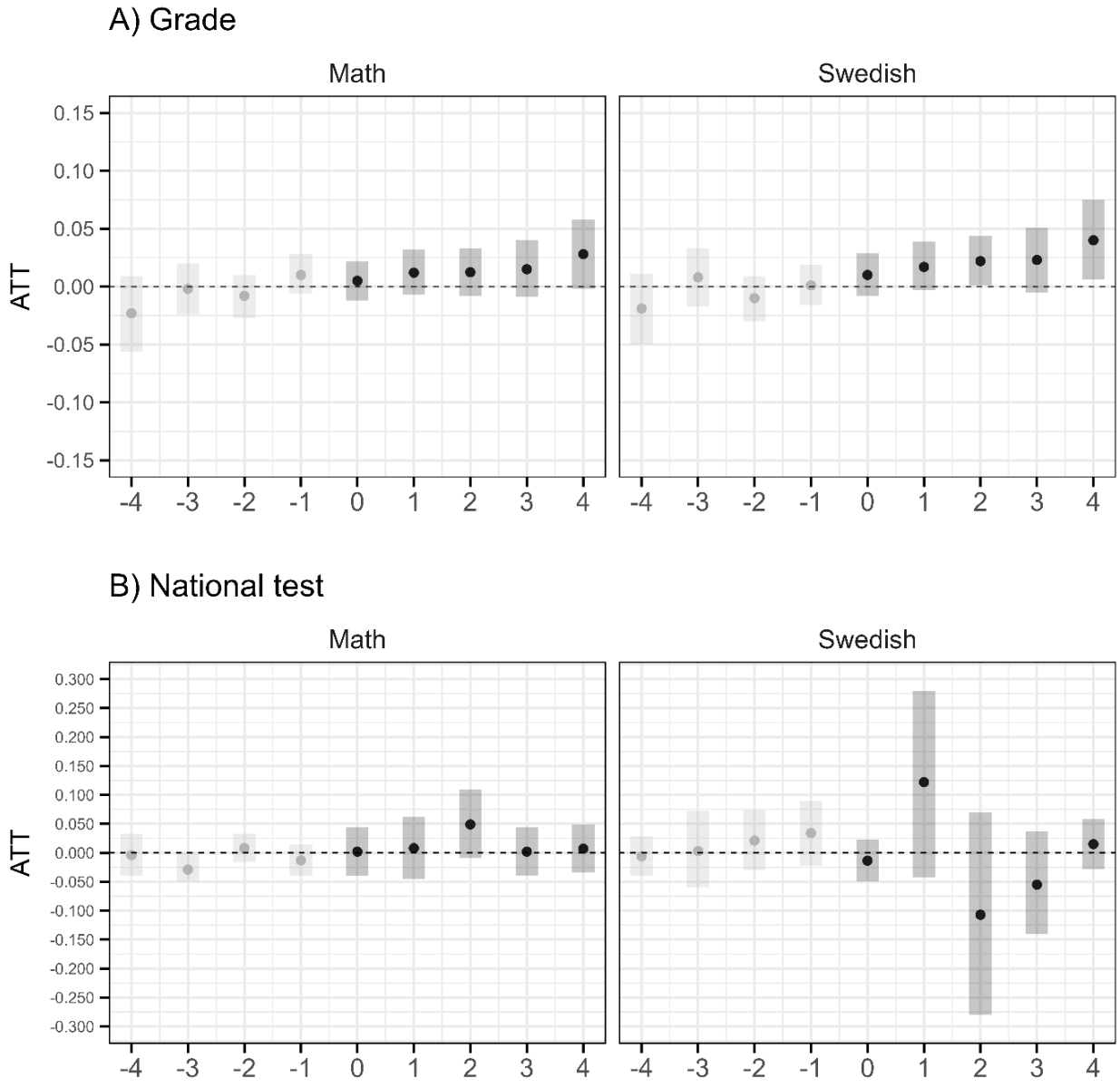
Estimated effects of SFL for all students

	ATT	Standard Error	p-Value	Lower CI (95%)	Upper CI (95%)	Observations	Pre-trend p-Value
Grades							
Math	0.0121	0.0089	0.173	−0.0053	0.0296	1,967,107	0.557
Swedish	0.0228	0.0102	0.026	0.0027	0.0429	1,967,107	0.667
National test scores							
Math	0.0119	0.0162	0.462	−0.0199	0.0439	1,380,376	0.037
Swedish	−0.0046	0.0242	0.847	−0.0520	0.0427	1,380,376	0.658

Note. The table shows the estimated effects of SFL on pooled Grade 6 and Grade 9 math and Swedish grades and national test scores when pooled across all students. Scores have been standardized by outcome measure and year. Standard errors clustered by school.

Figure 3

Event-time SFL effects on all students



Note. The figure displays event-time estimates where period 0 on the x-axis denotes the first year of SFL participation. Negative x-axis values indicate years prior to treatment and positive values indicate years after treatment. Light grey markers correspond to pre-treatment periods and dark grey markers correspond to post-treatment periods (point estimates and 95% confidence intervals). ATT = average treatment effects by event time (Callaway & Sant'Anna, 2021).

5.2 Effects of SFL for target subgroups

5.2.1 Students with IEPs

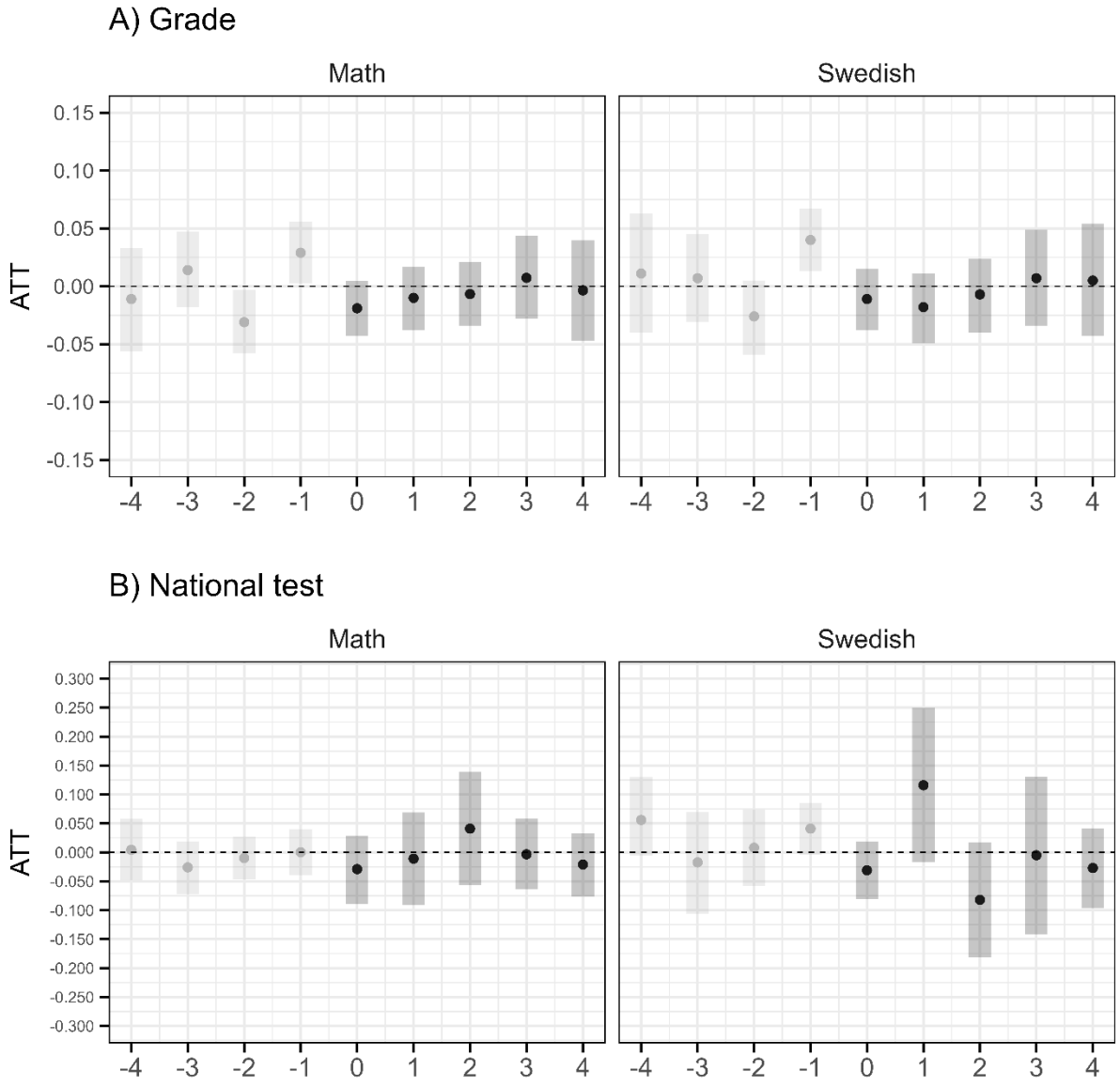
We now turn to reporting the effects of participation for stated subgroups of the SFL program, including students with special educational needs. Our primary specification (shown in Figure 4) defines students as having an IEP if they held an IEP at the time of the outcome measurement (i.e., in the same year as the grade or national test score). Figure 4 shows average treatment effects of 0.016 and 0.012 standard deviations for mathematics and Swedish grades, respectively, and average treatment effects of 0.005 and -0.006 standard deviations for mathematics and Swedish national test scores—none of which reach statistical significance, indicating null effects for this subgroup.

The event study plots in Figure 4 correspond to students who have a current IEP, which generates fewer observations and lower statistical power, particularly because IEPs are relatively uncommon in the lower grades. We therefore also subset the analysis to students who received an IEP at any point in their educational history, to capture a larger proportion of students who may be likely to receive special needs support, even in cases where no current IEP is reported. However, this approach also has the disadvantage that IEP status later in a student's educational trajectory may be influenced by a school's participation in the SFL program. Consequently, the results of this analysis should be interpreted with caution. The average treatment effects, along with their standard errors and sample sizes for each specification are reported in Table B1 in supplementary material section B, indicating similarly null results.

While the estimated effects cannot rule out a positive effect of the SFL program, a z-test for the equality of coefficients showed that the effect for IEP students did not differ significantly from the effect for the rest of the sample. Nevertheless, given the focus of the SFL program on general teacher training in special educational approaches, these null results (regardless of IEP definition) are noteworthy.

Figure 4

Event-time SFL effects on students with IEPs



Note. The figure displays event-time estimates where period 0 on the x-axis denotes the first year of SFL participation. Negative x-axis values indicate years prior to treatment and positive values indicate years after treatment. Light grey markers correspond to pre-treatment periods and dark grey markers correspond to post-treatment periods (point estimates and 95% confidence intervals). ATT = average treatment effects by event time (Callaway & Sant'Anna, 2021).

5.2.2 Students with foreign language backgrounds

Given that students with foreign language backgrounds were also a stated target group of the SFL program, we test average treatment effects for students who have 2 foreign born parents, are born outside of Sweden (arriving any time after birth) or who arrived within 5 years. Table 5 reports these average treatment effects, alongside average treatment effects for both specifications for students with IEPs.

Table 5 shows that just one estimate is positive and reaches (nominal) statistical significance: newly arrived students taking Swedish as a second language show a positive effect on national test scores (0.077 standard deviations, $p = 0.047$). However, given that Table 5 alone comprises 20 estimates, we implement a Romano-Wolf stepwise multiple testing correction using a cluster-corrected bootstrap (199 replications, resampled at the school level) to construct the joint null distribution. The Romano-Wolf–corrected p-values are reported in Table 5, and show that all previously statistically significant estimates lose their statistical significance after correction for multiple hypothesis testing.

The estimates in supplementary materials section B (Table B1) show the estimates for students with a foreign language background disaggregated by enrollment in Swedish as a second language in grades or national tests. These results show that newly arrived students taking Swedish as a first language display a positive effect on national test scores (0.341 standard deviations, $p = .018$); and childhood immigrants taking Swedish as a second language show a positive effect on national test scores (0.090 standard deviations, $p = .035$). No effects reach significance for course grades in any of the six migration-background-by-language subgroups, and no significant effects emerge for students with two foreign-born parents in either outcome. However, after Romano-Wolf correction, all previously statistically significant estimates lose their significance as for the estimates displayed in Table 5.

Table 5

Estimated SFL effects by student subgroups

	(1)	(2)	(3)	(4)	(5)
	IEP (ever)	IEP (current)	2 foreign-born parents	Childhood immigrant	Newly arrived
Grades					
Math					
ATT (SE)	−0.001 (0.012)	0.016 (0.019)	0.003 (0.013)	0.009 (0.015)	−0.006 (0.021)
[p; RW p]	[0.940; 1.000]	[0.436; 1.000]	[0.829; 1.000]	[0.563; 1.000]	[0.763; 1.000]
Observations	308,364	145,135	466,722	289,622	140,281
Pre-trend p-Value	0.293	0.468	<0.001	0.448	0.641
Swedish					
ATT (SE)	0.007 (0.014)	0.012 (0.019)	0.009 (0.018)	0.015 (0.023)	0.023 (0.029)
[p; RW p]	[0.617; 1.000]	[0.524; 1.000]	[0.620; 1.000]	[0.517; 1.000]	[0.426; 1.000]
Observations	308,364	145,135	466,722	289,622	140,281
Pre-trend p-Value	0.015	0.004	0.934	0.708	0.445
National test scores					
Math					
ATT (SE)	−0.009 (0.021)	0.005 (0.030)	0.039 (0.024)	0.043 (0.026)	0.042 (0.031)
[p; RW p]	[0.679; 1.000]	[0.881; 1.000]	[0.106; 0.790]	[0.099; 0.885]	[0.188; 0.910]
Observations	182,805	77,441	312,555	186,980	81,771
Pre-trend p-Value	0.292	0.574	0.341	0.603	0.474
Swedish					
ATT (SE)	−0.012 (0.026)	−0.007 (0.034)	0.025 (0.030)	0.060 (0.040)	0.078 (0.039)
[p; RW p]	[0.650; 1.000]	[0.848; 1.000]	[0.418; 1.000]	[0.138; 0.785]	[0.047; 0.540]
Observations	182,805	77,441	312,555	186,980	81,771
Pre-trend p-Value	0.002	0.235	0.829	0.722	0.488

Note. The table shows the estimated effects of SFL on pooled Grade 6 and Grade 9 math and Swedish grades and national test scores when pooled across all students. Scores have been standardized by outcome measure and year. p = unadjusted p-Value, RW p = Romano-Wolf adjusted p-Value.

5.3 Heterogeneity analyses

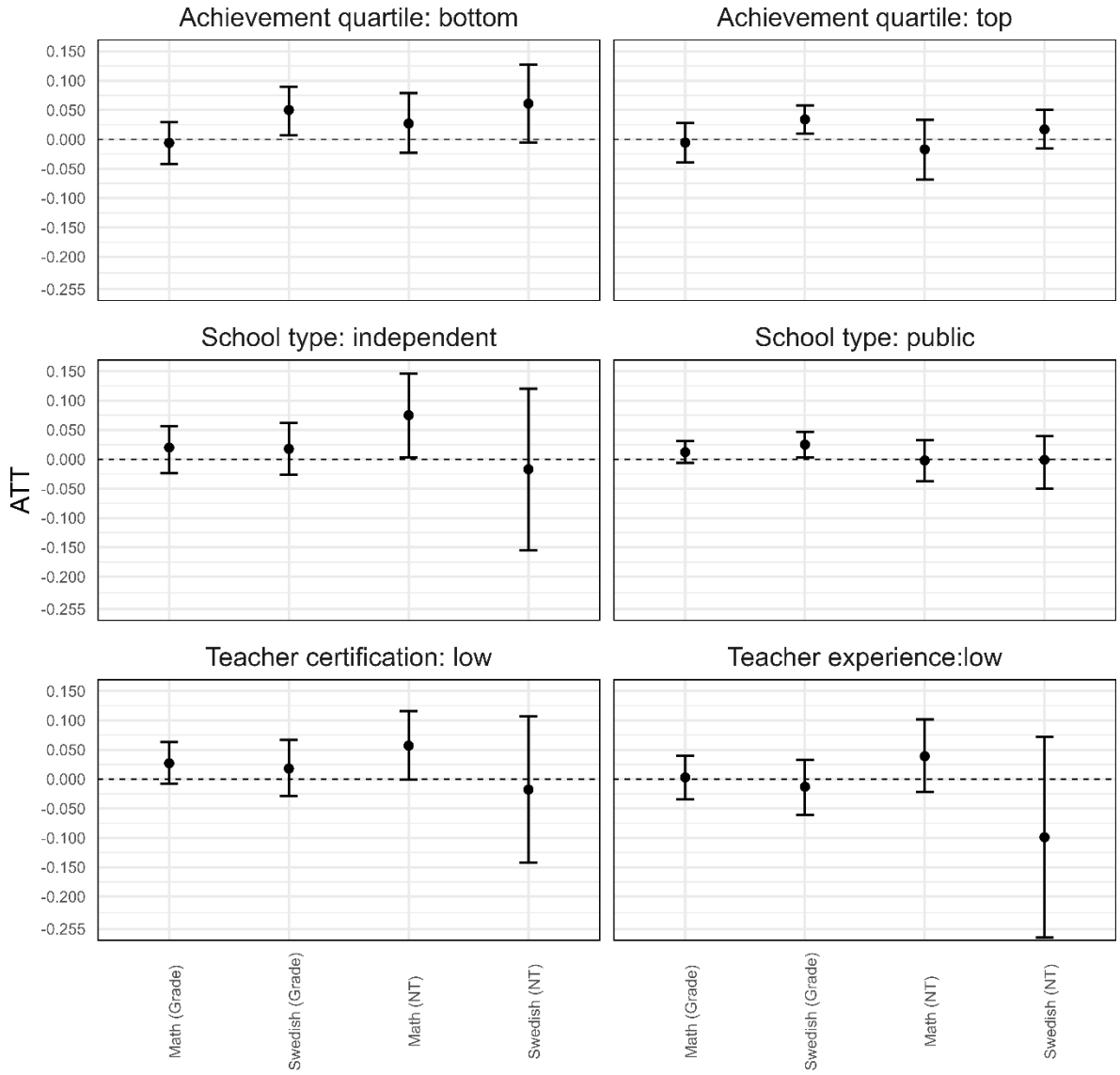
Figure 5 shows ATT estimates by gender, school achievement quartile, average teacher qualifications in the school, and school type. The pre-intervention school achievement quartiles are determined based on the subject in focus. As can be seen from Figure 5, the null pattern of results continues for the exploratory heterogeneity analyses. Most notably, we find positive and statistically significant effects of the SFL intervention on student grades in Swedish for those in the bottom achievement quartile, amounting to about 0.05 standard deviations for grades ($p = 0.022$) and 0.06 standard deviations for national test scores (though not statistically significant). Schools in the top achievement quartile also show small positive effects for Swedish grades (0.035 standard deviations, $p < 0.005$), but the effects in Swedish national test scores likewise do not reach statistical significance.

We find a similar pattern of null effects when subsetting the sample by average teacher qualifications,

either in the proportion of certified teachers or by average years of experience, as shown in Figure 5. One exception to this is a marginally insignificant treatment effect in mathematics national test scores for students in schools with a low share of certified teachers of 0.057 standard deviations ($p = 0.052$). In all other cases, effects of SFL do not differ by the average teacher qualification level in a school. Finally, we do not find any differential effects when analyses are subset by school type (see Figure 5), save for a similar pattern of positive effects in mathematics national test scores for students in independent schools (0.075 standard deviations, $p = 0.04$). The similarity of these estimates with those for schools with a low share of certified teachers may be due to that independent schools employ uncertified teachers at a higher rate. At the same time, these findings should be interpreted with caution, as the pre-trend test rejected the null hypothesis that the pre-treatment effects in mathematics were jointly equal to zero (pre-trend statistics alongside point estimates are also reported in supplementary materials section C Table C1). Event study plots for heterogeneity analyses may also be found in supplementary materials section C.

Figure 5

Estimated SFL effects by average school characteristics



Note. Point estimates and 95% confidence intervals are displayed by school characteristics.

5.4 Sensitivity analyses

We report the results from several sensitivity checks to determine the robustness of our findings in the study supplementary materials. First, to assess whether observed treatment effects might reflect changes in the demographic and socioeconomic composition of schools' student bodies rather than genuine program effects, we follow Grönqvist et al. (2025) in examining whether the demographic and socioeconomic composition of schools changes systematically following SFL implementation. We operationalize student body composition

using students' predicted test scores (a composite measure constructed from students' demographic and socioeconomic characteristics, including student migration background, gender, parental education, and school fixed effects) and test whether this measure shifts around the time of SFL adoption. These estimates can be found in supplementary material section D, and show no indication of changes in student composition which could serve as an alternative explanation.

Next, we consider an alternative treatment specification where we include all schools identified by the SNAE as having participated in the SFL program despite the lack of information on whether a coach was assigned to the school. The event study plots and ATTs for this analysis can be found in supplementary material section E, showing average pooled estimates highly similar to those we find in Table 4, indicating general stability of findings across treatment specifications. We also test the alternative treatment specification for the statistically significant findings from the heterogeneity analyses and find effects of similar magnitude but which in all cases fail to reach conventional levels of statistical significance.

Although we do not have access to implementation data across each year, we run a sensitivity test to determine whether the findings may be sensitive to variation in treatment dosage. Specifically, we examine whether treatment dosage has an impact on SFL success by investigating whether small schools showed larger effects of SFL (based on the assumption that participating teacher shares will be higher in small schools). These results can be found in supplementary material section F, showing slightly larger effects in mathematics though they do not reach statistical significance. Next, we examine the concern of grading leniency as an SFL outcome in section G of the supplementary materials. Effects are negligible for mathematics, while effects for Swedish are positive but statistically insignificant at approximately 0.02 standard deviations. The latter suggests that increased grading leniency remains a plausible explanation for the observed improvements in Swedish grades, particularly given the absence of corresponding effects on national test scores.

Given that earlier participating schools may differ systematically from schools which participated in later SFL cohorts, and to account for any minor changes to program design and content over time, we also examine cohort specific effects to examine whether SFL effectiveness varied as a function of the cohort of participating schools, and these estimates can be found in supplementary materials section H, and show that the null effects of SFL cannot be explained by effect heterogeneity across cohorts.

Considering that independent schools were more likely to participate in the SFL program, we incorporate school type as a covariate in the estimation and therefore change the identifying parallel trends assumption to conditional on school type. The results are presented in supplementary materials section I, showing nearly identical estimates to those in the main specification, including null effects for students with IEPs and significant positive effects for childhood immigrants. Last, we estimate separate effects by grade and find that the statistically significant effects for the pooled analysis in Swedish grades are driven by grade

9 students (supplementary materials section J).

6 Discussion

Through a difference-and-differences estimation strategy, this study evaluated the causal effect of a large-scale teacher professional development program in Sweden aimed at training general classroom teachers in special and inclusive education—a program in which 60,000 teachers participated and which cost the equivalent of roughly 70 million USD. In relation to previous research on PD programs in special and inclusive education for ordinary classroom teachers, our analysis constitutes rare and comparatively more robust evidence on the dynamic effects of such programs—evidence which is particularly scarce for large-scale programs such as SFL (Donath et al., 2023).

The effects of the SFL program are relevant to a global educational policymaking audience for several reasons. Like most countries, Sweden has a diverse body of pupils, ranging from students with special educational needs, foreign language backgrounds, or who may otherwise struggle in school, the lion's share of whom attend general classrooms. General classroom teachers repeatedly request additional training in this area (Cooc, 2019; Eurydice, 2019). Causal evidence on in-service initiatives aimed at improving teaching in general classrooms can therefore tell us which approaches are successful and whether their continued or wider use is justified. Moreover, using a dynamic difference-in-differences approach can tell us if effects of such programs persist in the post-intervention period.

Overall, the evidence regarding the effectiveness of the SFL program is mixed. When averaged across all students, SFL yielded small and statistically significant positive effects in Swedish grades of 0.023 standard deviations, persistent for several years after the intervention. However, the average effects of SFL were not confirmed by the national test scores. We are therefore unable to conclude that the program was successful in raising student achievement for the general student body. Despite this, given the specific focus of SFL in targeting (1) students with individualized education programs, and (2) students with foreign language backgrounds, we argue that the effect heterogeneity is at least as important for determining the success of SFL.

For students with individualized educational programs, we find consistently null or even slightly negative effects of the SFL program. This finding is indeed discouraging given the explicit focus of the SFL program on special educational approaches. On the other hand, the results suggest that a subset of childhood immigrants may have benefited from SFL—specifically those arriving in Sweden within 5 years. Students attending SFL participating schools in this group showed improved test scores of 0.077 standard deviations in national test scores in Swedish, a magnitude that is often described as moderate relative to effects commonly observed in educational interventions (Evans & Yuan, 2020; Kraft, 2020). Although the effects remained consistent in models conditioned on school type, they failed to reach statistical significance

at conventional levels in the alternate treatment specification—and failed to remain statistically significant after Romano-Wolf correction.

Due to the null effects on students with individualized education programs, we conclude that our analysis does not add empirical support for the type of broad PD focusing on whole-class approaches for students with special educational needs used in the SFL program, which included a host of diverse aims, for example including improved student-teacher relationships or a more safe and supportive school climate, via enhanced teacher awareness, reflection and practice with respect to a wide range of (typically inclusion-oriented) topics in special education. This may be due to either the PD design or the pedagogical approaches advocated by the program. If the latter is the case, the results are consistent with D. Fuchs et al. (2025), who argue that research offers inconclusive evidence regarding how to best support students with disabilities in general classrooms, while the strongest and most reliable findings instead point to the effectiveness of intensive supplementary instruction, typically delivered outside the general classroom. An alternative interpretation is that there are more effective whole-class practices than those promoted within the SFL program (Mccrea, Goodrich, & Barker, 2025). Indeed, while the findings are more scarce and the evidence more mixed, studies have documented whole-class programs that improve outcomes for struggling students (Neitzel, Lake, Pellegrini, & Slavin, 2022).

Although the main analyses suggest positive effects for newly arrived immigrant students, these findings were not robust across all model specifications and did not remain significant after correction for multiple hypothesis testing and should therefore be interpreted with caution. In addition, though it is certainly plausible that newly arrived students may benefit to a greater degree from the pedagogical strategies promoted within the SFL program, it is also possible that schools serving different student bodies implemented different modules of SFL. For example, any positive effects for migrant students could be due to schools serving larger shares of these students implementing the multilingualism module.

The null or weak effects of the SFL program may not necessarily be due to the special education approaches advocated by the program. It may, alternatively, be the form of the program that reduced its success. Like previous large-scale PD programs in Sweden, the SFL program followed a 'collegial learning' approach, where teachers follow modules and meet weekly to discuss module content, apply it in their classrooms, and reflect on the classroom application together under the guidance of a trained coach. Although the program included some effective PD elements (e.g., coaching and theory-practice links; Kraft et al., 2018; Sims et al., 2023), it placed relatively little emphasis on deliberate modelling, rehearsal, and feedback focused on specific instructional practices, and also did not include teaching observations or video-recorded teaching (Banks et al., 2025; Cohen, Wong, Krishnamachari, & Erickson, 2024; Sims, Fletcher-Wood, Godfrey-Faussett, Mccrea, & Meliss, 2026). This was not an accidental omission: the program deliberately avoided promoting

specific teaching methods, instead offering materials intended to stimulate professional dialogue within schools about improvement. While this approach may have merits, it may also have limited the program's impact on classroom practice. At the same time, evidence from two earlier Swedish programs using a similar 'collegial learning' model (BFR and BFM) suggests that such interventions tend to yield only small effects (0.01 to 0.03 standard deviations in national test scores), yet these evaluations reported statistically significant positive impacts nonetheless, despite facing similar issues related to treatment dilution. The absence of detectable effects on national test scores for SFL may therefore point to differences in program content as an explanation for the different results across programs. However, this interpretation is inconclusive: our point estimates are of similar magnitude, and the confidence intervals of our main effects include the effect sizes found for BFR and BFM.

Future policy efforts to boost student test scores may nevertheless wish to refine certain elements of the Swedish 'collegial learning' approach. For example, considering that a long duration has not been shown to be important for PD success (Didion, Toste, & Filderman, 2020; Kraft et al., 2018; Lynch, Gonzalez, Hill, & Merritt, 2025) and that teachers' time is limited, a shorter and narrower approach which includes more direct guidance, modelling, and observation-based feedback may be warranted. Furthermore, given the inherent challenges in scaling programs while maintaining effects, it would be valuable to integrate rigorous evaluation into program rollout from the outset, enabling A/B comparisons of design alternatives and evidence-based refinements before nationwide implementation (Angrist et al., 2026; List, 2024).

Though we do not have insight into the teaching practices which may have led to the documented effects, we investigate results across a range of school categories and student characteristics in search of explanatory mechanisms for our findings. In most cases, the null pattern of results continues, including across school type and when the analysis was subset to smaller schools. Sensitivity analyses also suggested that certain SFL cohorts may have been more successful (e.g. 2017) and that the positive effects of SFL may have been driven by effects for grade 9 students. Furthermore, while we observed positive effects in Swedish across both achievement measures in schools which were in the bottom achievement quartile prior to the SFL intervention, this only reached statistical significance for grades. Last, while a positive effect of 0.075 standard deviations was found in mathematics national test scores in independent schools and 0.06 in schools with a high share of uncertified teachers, possibly due to independent schools employing a higher share of uncertified teachers, these estimates were reduced considerably in magnitude and no longer statistically significant under model specifications with an alternative treatment definition and pre-trend tests reached statistical significance for mathematics in several cases.

Beyond the statistical significance of the estimates, the confidence intervals are informative regarding the range of plausible effects. For the pooled analyses, the estimated effects were generally close to zero and

the confidence intervals ruled out large positive impacts on student achievement. For example, the upper confidence bounds ranged from approximately 0.03 to 0.04 standard deviations for mathematics outcomes and 0.04 standard deviations for Swedish outcomes. These bounds suggest that, even if the true average effect of SFL were positive, it was likely modest in magnitude. Moreover, the confidence intervals largely overlap with the effect sizes previously reported for other large-scale Swedish professional development initiatives, such as the Boost for Mathematics and Boost for Reading programs, which produced effects on the order of 0.01-0.03 standard deviations. For students with IEPs, upper bounds were mostly around 0.04–0.06 standard deviations, so moderate effects remain unlikely. For students with a foreign-language background, the confidence intervals include both zero and moderately large effects, so these results are inconclusive. At the same time, the estimates reflect the impact of offering SFL at the school level. Considering that not all teachers in participating schools took part in SFL, the reported effects may underestimate any improvement that might have been observed under full participation.

6.1 Limitations and conclusion

The study has some limitations. The main limitations have to do with opacity regarding SFL program implementation, as beyond the foundational mandatory modules on inclusion, we cannot say for certain which PD modules were implemented in which schools and to what degree the modules were followed. We however argue that this presents a trade-off ultimately outweighed by the ecological validity of the findings due to the large scale and detailed analyses across subgroups and time. Though the estimates undoubtedly reflect considerable implementation heterogeneity, we partially address this limitation by examining differential impacts based on school size, teacher qualifications, and school type. In addition, it is possible that the SFL program may have been beneficial for certain subgroups of special educational needs, but we are not able to examine this based on the national register data. Furthermore, it is also possible that the program might have produced effects on other outcome measures than those tested in this study, such as more narrow dimensions of reading or mathematics, or behavioural outcomes. Finally, although the pre-trend tests held in most cases and we found no effects of the SFL program on student composition (proxied by predicted test scores), participation in the SFL program was voluntary for schools; as a result, participating schools may differ in unobserved characteristics, potentially violating the parallel trends assumption, which suggests that the certainty of causal interpretations are weaker than under randomized assignment to program participation.

Despite these limitations, in light of the still limited causal evidence on large-scale PD in special or inclusive educational approaches, our study is indeed useful. Although teacher PD programs often fail to produce significant positive effects on student outcomes (Burgess, Rawal, & Taylor, 2021; Lindvall et al.,

2021), they remain an important lever for improving teaching practices and student academic outcomes, requiring substantial financial and time commitments on the part of school actors and school staff. Rigorous evaluations are therefore important not only for informing decisions about investment and implementation, but also for advancing understanding of whether such programs are effective, for whom they are effective, and under what conditions they may contribute to improved outcomes. This is particularly true for special and inclusive education initiatives, where high-quality evidence remains scarce despite that general education teachers request more PD in this area (Cooc, 2019; Didion et al., 2026; Eurydice, 2019). Our results show that one such special education PD program in Sweden implemented at scale did not improve academic outcomes, either for students overall or for students with special educational needs. Although some analyses suggested positive effects on Swedish outcomes for newly arrived immigrant students, the evidence was not consistent across specifications. The study is therefore unable to conclude that the inclusion oriented, whole-classroom PD approach of the SFL program realized its intended aims of improving student achievement.

References

Angrist, N., et al. (2026). Iterative A/B testing for social impact: Rigorous, rapid, regular. *Stanford Social Innovation Review*. Retrieved from `https://ssir.org/articles/entry/iterative-a-b-testing-social-impact`

Bal, A., & Perzigian, A. B. T. (2013). Evidence-based interventions for immigrant students experiencing behavioral and academic problems: A systematic review of the literature. *Education and Treatment of Children*, *36*(4), 5–28. Retrieved from `http://www.jstor.org/stable/42900224`

Banks, B., Sims, S., Curran, J., Meliss, S., Chowdhury, N., Altunbas, H. G., et al. (2025). Decomposition and recomposition: Effects on early-career teachers' enactment and adaptive transfer of behaviour management practices. *European Journal of Teacher Education*, 1–24. doi: 10.1080/02619768.2025.2511867

Bölte, S., Leifler, E., Berggren, S., & Borg, A. (2021). Inclusive practice for students with neurodevelopmental disorders in Sweden. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, *9*, 9–15. doi: 10.21307/sjcapp-2021-002

Burgess, S., Rawal, S., & Taylor, E. S. (2021). Teacher peer observation and student test scores: Evidence from a field experiment in English secondary schools. *Journal of Labor Economics*, *39*(4), 1155–1186.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*, 200–230. doi: 10.1016/j.jeconom.2020.12.001

Carroll, J. M., Bradley, L., Crawford, H., Hannant, P., Johnson, H., & Thompson, A. (2017). *SEN support: A rapid evidence assessment* (Tech. Rep.). Coventry University.

Chao, C. N. G., Sze, W., Chow, E., Forlin, C., & Ho, F. C. (2017). Improving teachers' self-efficacy in applying teaching and learning strategies and classroom management to students with special education needs in Hong Kong. *Teaching and Teacher Education*, *66*. doi: 10.1016/j.tate.2017.05.004

Chen, X., & Zhang, Y. (2019). Typical practices of lesson study in East Asia. *European Journal of Education*, *54*, 189–201. doi: 10.1111/ejed.12334

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. doi: 10.1257/aer.104.9.2633

Choi, S., & Lee, S. W. (2020). Enhancing teacher self-efficacy in multicultural classrooms and school climate: The role of professional development in multicultural education in the United States and South Korea. *AERA Open*, *6*(4). doi: 10.1177/2332858420973574

Choi, S., Mao, X., & Park, S. (2025). Promoting school belonging for immigrant students: The interplay of

inclusive school climate and multicultural education. *School Effectiveness and School Improvement*, *36*(4), 522–542. doi: 10.1080/09243453.2025.2492012

Coffey, S. G., Goodman, J., Schwartz, A. E., Stiefel, L., Winters, M. A., & Yoon, Y. H. (2026). *Special education substantially improves learning: Evidence from three states* (NBER Working Paper No. 34998). National Bureau of Economic Research. doi: 10.3386/w34998

Cohen, J., Wong, V. C., Krishnamachari, A., & Erickson, S. (2024). Experimental evidence on the robustness of coaching supports in teacher education. *Educational Researcher*, *53*(1), 19–35.

Cooc, N. (2019). Teaching students with special needs: International trends in school capacity and the need for teacher professional development. *Teaching and Teacher Education*, *83*, 27–41. doi: 10.1016/j.tate.2019.03.021

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*, 418–446. doi: 10.1002/pam.20586

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54.

Didion, L., Filderman, M. J., Hart, E. J., Payne, S. B., Vitale, J., Olmstead, C., & Wexler, J. (2026). High-quality special education professional development improves teacher and student outcomes. *The Journal of Special Education*. doi: 10.1177/00224669261423924

Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness*, *13*(1), 29–66. doi: 10.1080/19345747.2019.1670884

Donath, L., Luke, T., Graf, E., Tran, U., & Götz, T. (2023). Does professional development effectively support the implementation of inclusive education? A meta-analysis. *Educational Psychology Review*, *35*. doi: 10.1007/s10648-023-09752-2

Eurydice. (2019). *Integrating students from migrant backgrounds into schools in Europe: National policies and measures*. Luxembourg: Publications Office of the European Union.

Eurydice. (2023). *Promoting diversity and inclusion in schools in Europe*. Publications Office of the European Union. Retrieved from `https://eurydice.eacea.ec.europa.eu/publications/promoting-diversity-and-inclusion-schools-europe`

Evans, D. K., & Yuan, F. (2020). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*. (01623737221079646)

Expressen. (2023, April 29). *Skolverket sprider myter och duckar för kritiken.* Expressen. Retrieved from `https://www.expressen.se/debatt/skolverket-sprider-myter-och-duckar-for-kritiken/`

Florian, L. (2019). On the necessary co-existence of special and inclusive education. *International Journal of Inclusive Education*, *23*(7–8), 691–704. doi: 10.1080/13603116.2019.1622801

Fuchs, D., Gilmour, A. F., & Wanzek, J. (2025). Reframing the most important special education policy debate in 50 years: How versus where to educate students with disabilities in America's schools. *Journal of Learning Disabilities*, *58*(4), 257–273. doi: 10.1177/00222194251315196

Fuchs, L. S., Fuchs, D., Compton, D. L., Wehby, J., Schumacher, R. F., Gersten, R., & Jordan, N. C. (2015). Inclusion versus specialized intervention for very-low-performing students: What does access mean in an era of academic challenge? *Exceptional Children*, *81*(2), 134–157.

Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. S. (2016). Does evidence-based fractions intervention address the needs of very low-performing students? *Journal of Research on Educational Effectiveness*, *9*(4), 662–677. doi: 10.1080/19345747.2015.1123336

Gamse, B. C., Bloom, H. S., Kemple, J. J., & Jacob, R. T. (2008). *Reading first impact study: Interim report* (Tech. Rep. No. NCEE 2008-4016). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Gersten, R., Haymon, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, *13*. doi: 10.1080/19345747.2019.1689591

Giota, J., Lace, I., & Emanuelsson, I. (2023). School achievement and changes in inclusive vs exclusive support over 50 years in Sweden regarding students with intellectual disabilities and special educational needs. *Scandinavian Journal of Educational Research*, *67*(6), 997–1011. doi: 10.1080/00313831.2022.2115129

Grönqvist, E., Öckert, B., & Rosenqvist, O. (2025). Does the "Boost for Mathematics" boost mathematics? A large-scale evaluation of the "lesson study" methodology on student performance. *American Economic Journal: Economic Policy*, *17*, 345–372.

Holmlund, H., Häggblom, J., & Lindahl, E. (2024). *The boost for reading* (IFAU Report No. 2024:5). IFAU – Institute for Evaluation of Labour Market and Education Policy.

Holmstrand, L., & Härnsten, G. (2003). *Förutsättningar för forskningscirklar i skolan: En kritisk granskning.* Stockholm: Myndigheten för skolutveckling.

Jacob, B., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *The Journal of Human Resources*, *39*, 50–79.

Jordan, A., Schwartz, E., & McGhie-Richmond, D. (2009). Preparing teachers for inclusive classrooms. *Teaching and Teacher Education*, *25*(4), 535–542. doi: 10.1016/j.tate.2009.02.010

Kahmann, R., Droop, M., & Lazonder, A. W. (2022). Meta-analysis of professional development programs in differentiated instruction. *International Journal of Educational Research*, *116*, 1–13. doi: 10.1016/

j.ijer.2022.102072

Kalinowski, E., Gronostaj, A., & Vock, M. (2019). Effective professional development for teachers to foster students' academic language proficiency across the curriculum: A systematic review. *AERA Open*, *5*(1). doi: 10.1177/2332858419828691

Kirsten, N., & Carlbaum, S. (2020). Kompetensutveckling för professionella lärare? Introduktionen av kollegialt lärande i svensk skola. *Pedagogisk forskning i Sverige*. doi: 10.15626/pfs25.01.01

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. doi: 10.3102/0013189X20912798

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. doi: 10.3102/0034654318759268

List, J. A. (2024). Optimally generate policy-based evidence before scaling. *Nature*, *626*, 491–499.

Lynch, K., Gonzalez, K., Hill, H., & Merritt, R. (2025). A meta-analysis of the experimental evidence linking mathematics and science professional development interventions to teacher knowledge, classroom instruction and student achievement. *AERA Open*. doi: 10.1177/23328584251335302

Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, *92*, 1441–1462.

Mccrea, P., Goodrich, J., & Barker, J. (2025). *Inclusive teaching – A discussion paper.* Steplab. Retrieved from `https://steplab.co/resources/inclusion-paper/682b7ad7af6f000001b321fd`

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, *57*(1), 149–179.

OECD. (2015a). *Immigrant students at school: Easing the journey towards integration.* Paris: OECD Publishing. doi: 10.1787/9789264249509-en

OECD. (2015b). *Improving schools in Sweden: An OECD perspective.* Paris: OECD Publishing.

OECD. (2021). *Promoting inclusive education for diverse societies: A conceptual framework* (OECD Education Working Papers No. EDU/WKP(2021)17). OECD.

O'Hagan, K. G., & Stiefel, L. (2024). Does special education work? A systematic literature review of evidence from administrative data. *Remedial and Special Education*, *46*. doi: 10.1177/07419325241244485

Pang, M. F., & Ling, L. M. (2012). Learning study: Helping teachers to use theory, develop professionally, and produce new knowledge to be shared. *Instructional Science*, *40*(3), 589–606. doi: 10.1007/s11251-011-9191-4

Regeringen. (2015). *Uppdrag att svara för genomförandet av fortbildning i specialpedagogik för lärare i grundskolan, motsvarande utbildning vid särskilda ungdomshem och sameskolan.* Regeringsbeslut I:3, U2015/05783/S. Utbildningsdepartementet.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244. doi: 10.1016/j.jeconom.2023.03.008

Savolainen, H., Engelbrecht, P., Nel, M., & Malinen, O.-P. (2012). Understanding teachers' attitudes and self-efficacy in inclusive education: Implications for pre-service and in-service teacher education. *European Journal of Special Needs Education*, *27*, 51–68.

Scanlon, D., Gelzheiser, L. M., Vellutino, F., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. *Learning and Individual Differences*, *18*, 346–359. doi: 10.1016/j.lindif.2008.05.002

Schwartz, A. E., Hopkins, B. G., & Stiefel, L. (2021). The effects of special education on the academic performance of students with learning disabilities. *Journal of Policy Analysis and Management*, *40*, 480–520. doi: 10.1002/pam.22282

SFS 2011:185. (2011). *Skolförordning [Education Ordinance].* Utbildningsdepartementet. Retrieved from `https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/skolforordning-2011185_sfs-2011-185/`

Sims, S., Fletcher-Wood, H., Godfrey-Faussett, T., Mccrea, P., & Meliss, M. (2026). Modelling evidence-based practice in initial teacher training: Effects on teachers' skills, knowledge and self-efficacy. *Instructional Science*, *54*. doi: 10.1007/s11251-026-09779-2

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., ... Anders, J. (2023). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research*, *95*(2), 213–254. doi: 10.3102/00346543231217480

Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M., & Maulana, R. (2019). Differentiated instruction in secondary education: A systematic review of research evidence. *Frontiers in Psychology*, *10*, 1–23. doi: 10.3389/fpsyg.2019.02366

Söderlund, G., Thoutenhoofd, E., Westman Andersson, G., Gerrbo, I., & Berhanu, G. (2024). Inclusive education in Sweden: A survey of teachers' attitudes, self-efficacy and collaboration towards better meeting pupils learning. *European Journal of Inclusive Education*, *3*. doi: 10.7146/ejie.v3i1.145185

Swedish National Agency for Education. (2025). *Specialpedagogik för lärande: Slutredovisning av regeringsuppdrag om fortbildning i specialpedagogik 2015–2025* (Skolverket rapport No. 2025:2853). Skolverket.

Thoutenhoofd, E. D., Söderlund, G., Westman Andersson, G., Berhanu, G., & Gerrbo, I. (2020). *Utvärdering av kompetensutvecklingsinsatsen specialpedagogik för lärande (SFL)* (RIPS: Rapporter från Institutionen för pedagogik och specialpedagogik No. 20). Institutionen för pedagogik och specialpedagogik, Göteborgs

universitet.

Vlachos, J. (2019). Trust-based evaluation in a market-oriented school system. In M. Dahlstedt & A. Fejes (Eds.), *Neoliberalism and market forces in education: Lessons from Sweden* (1st ed.). Routledge. doi: 10.4324/9780429470530

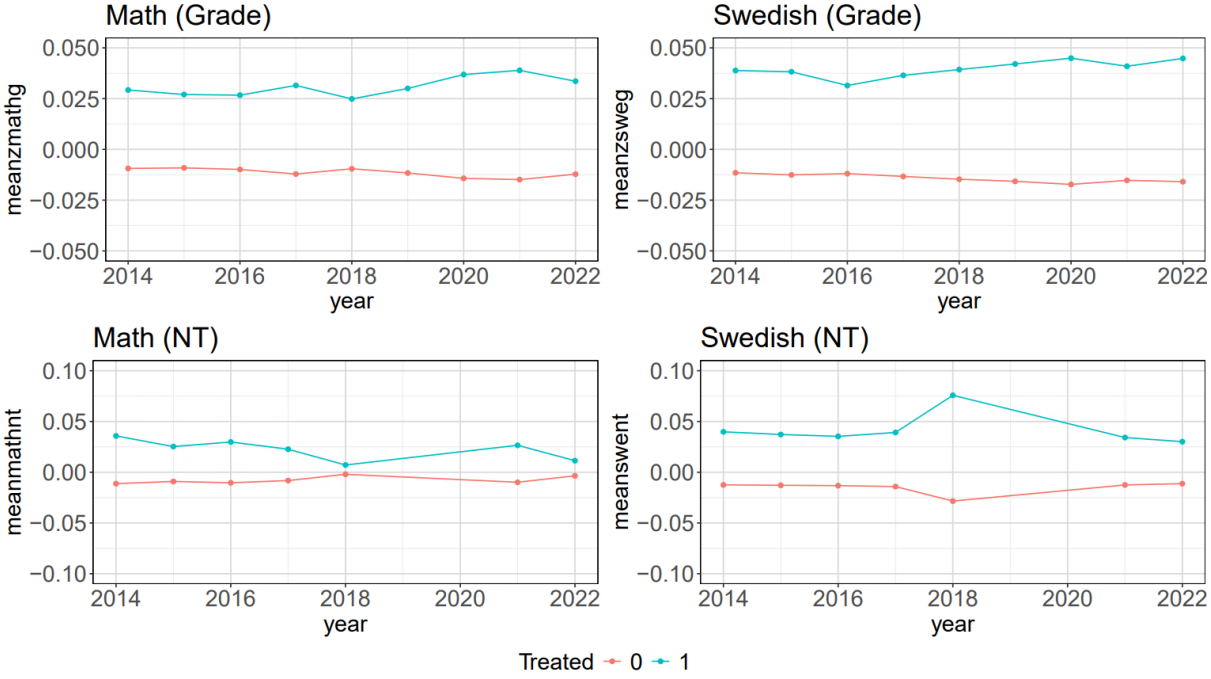
Yoshikawa, H., et al. (2015). Experimental impacts of a teacher professional development program Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, *51*, 309–322.

Zentall, S. S., & Javorsky, J. (2007). Professional development for teachers of students with ADHD and characteristics of ADHD. *Behavioral Disorders*, *32*(2), 78–93. doi: 10.1177/019874290703200202

Supplementary material

Figure A1

Outcome variables over time



Note. No national tests were conducted in 2019 or 2020 due to the coronavirus pandemic.

Supplementary material section B

Figure B1

Students with individualized education programs (IEPs)

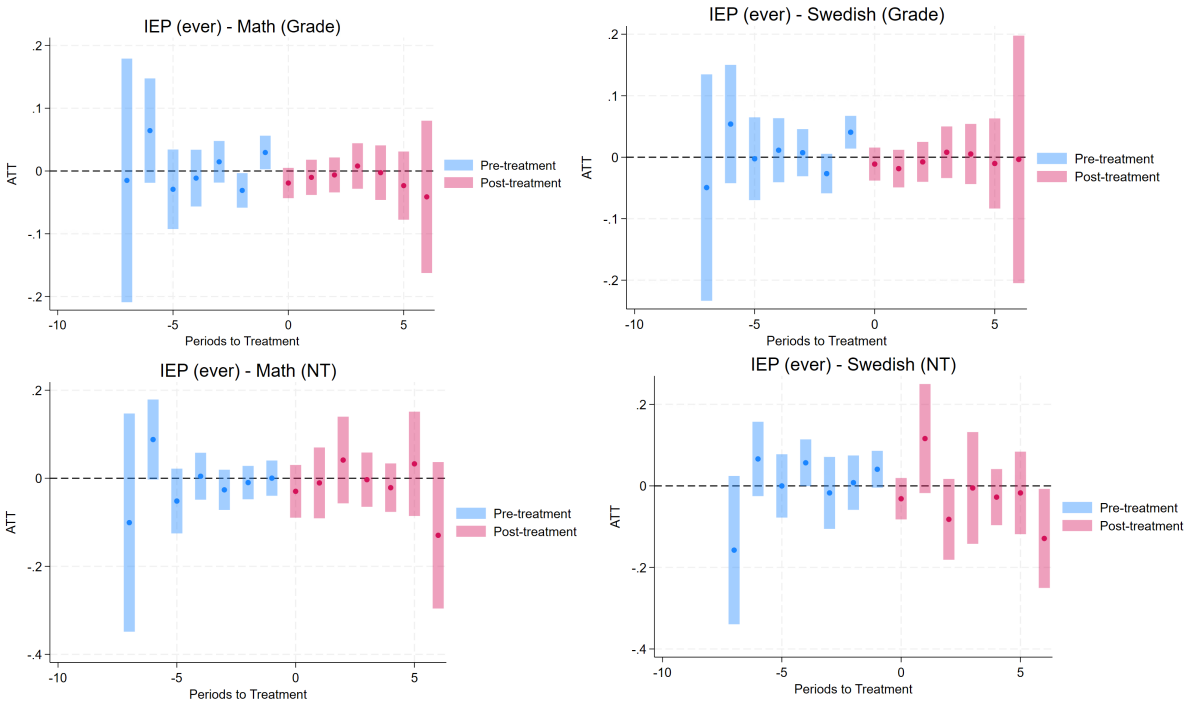


Figure B2

Migrant background

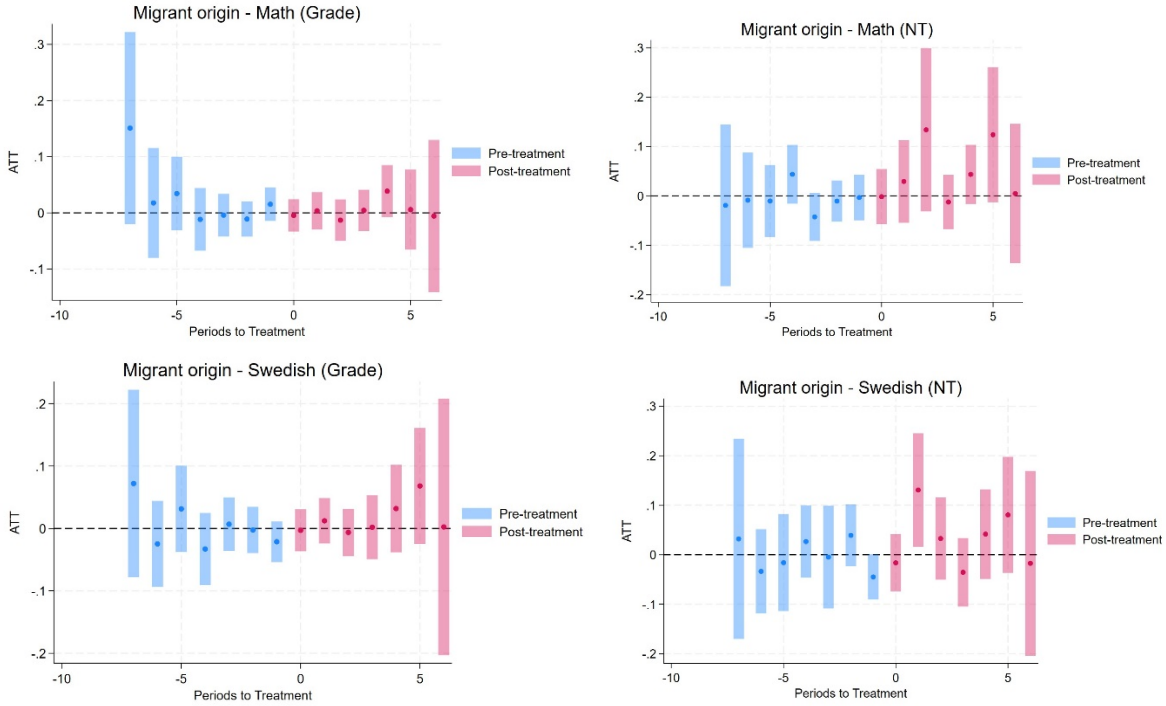


Figure B3

Childhood immigrants

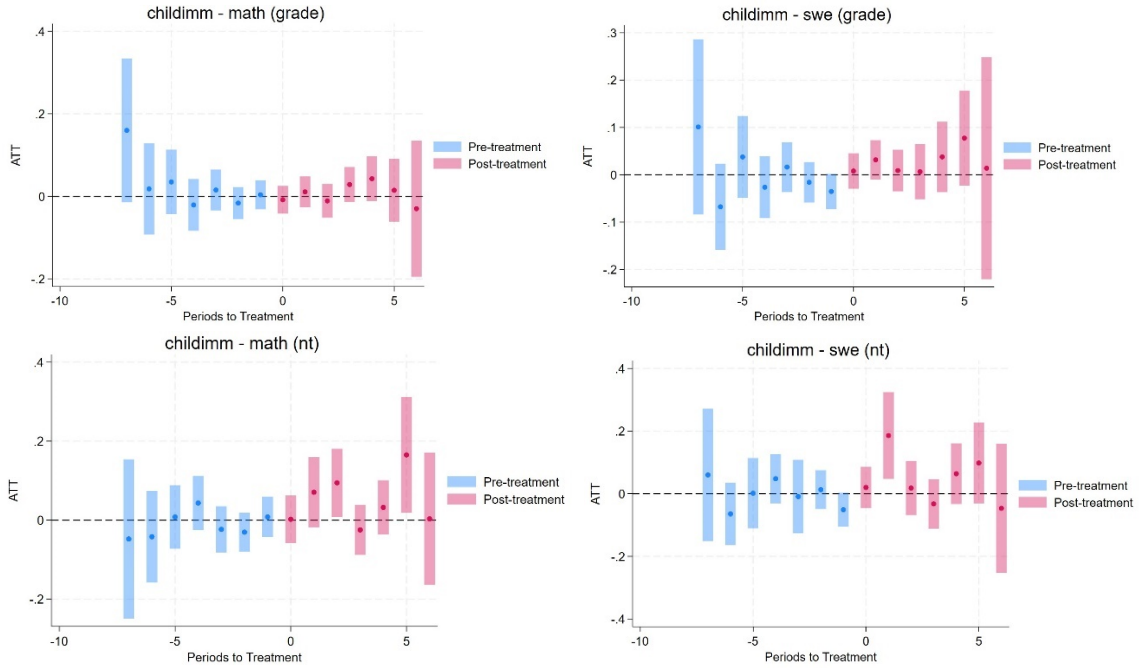


Figure B4

Arrived within 5 years

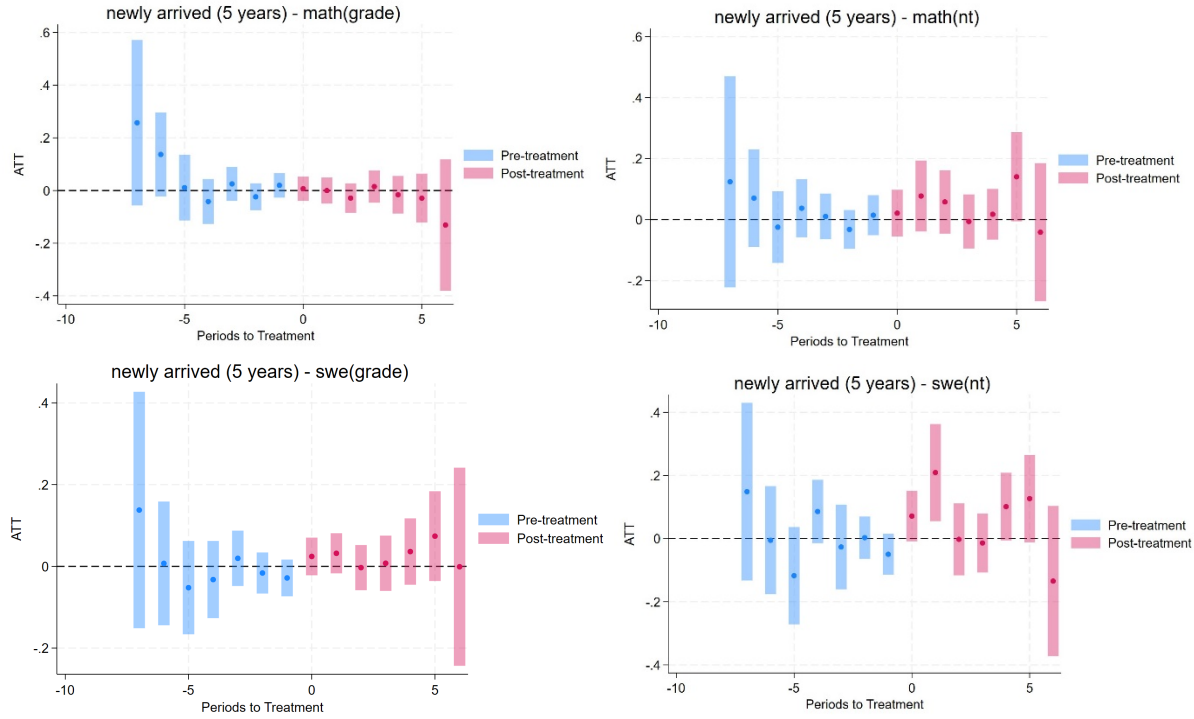


Table B1

Average treatment effects by migration background and Swedish level

Migration background –Swedish level	ATT	SE	*p*-Value (unadjusted)	*p*-Value (RW adjusted)
Newly arrived students – Swedish as a second language				
Grade	-0.0158	0.0292	0.588	0.990
National test score	0.0828	0.0393	0.035	0.465
Newly arrived students – Swedish as a first language				
Grade	0.0031	0.0539	0.953	0.995
National test score	0.3406	0.1434	0.018	0.390
Childhood immigrants – Swedish as a second language				
Grade	-0.0005	0.0295	0.985	0.995
National test score	0.0903	0.0428	0.035	0.435
Childhood immigrants – Swedish as a first language				
Grade	-0.0231	0.0270	0.394	0.975
National test score	0.0537	0.0671	0.423	0.975
2 foreign born parents – Swedish as a second language				
Grade	-0.0347	0.0276	0.209	0.855
National test score	0.0273	0.0565	0.629	0.990
2 foreign born parents – Swedish as a first language				
Grade	-0.0065	0.0234	0.789	0.990
National test score	0.0717	0.0459	0.118	0.745

Note. The table shows the average treatment effects from the difference-in-differences analysis for subgroups of migrant origin students. Scores have been standardized by outcome measure and year. Standard errors clustered by school. RW= Romano-Wolf.

Supplementary material section C

Figure C1

Boys

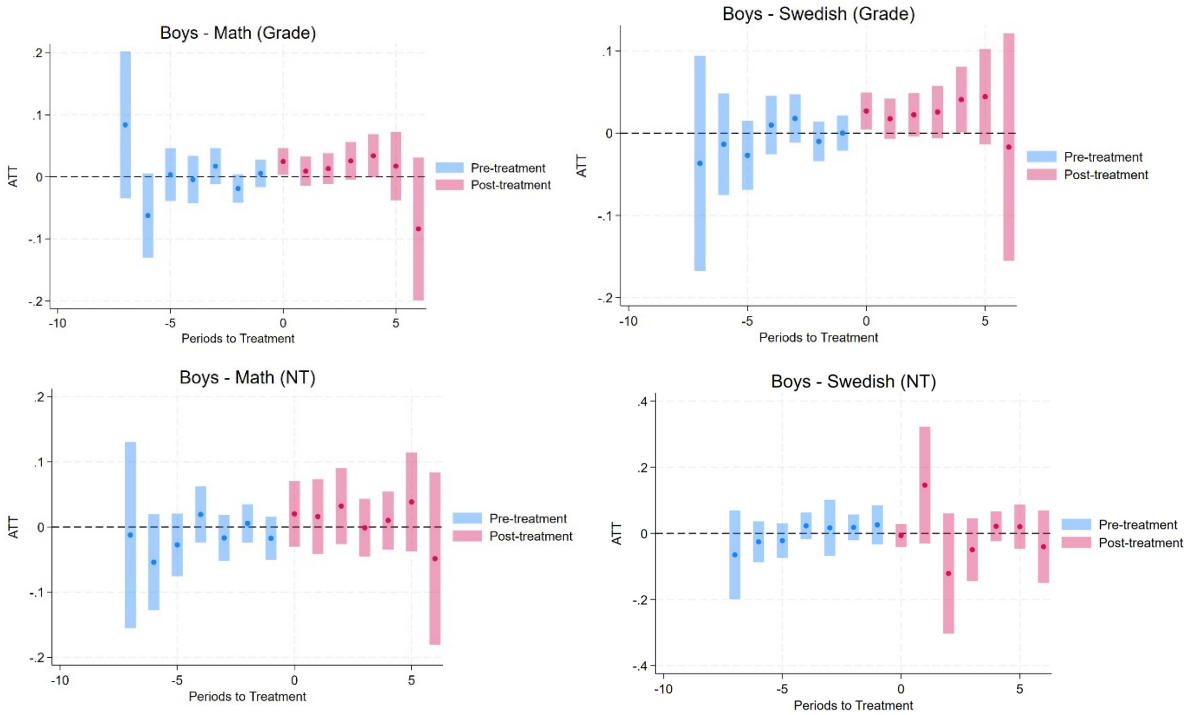


Figure C2

Girls

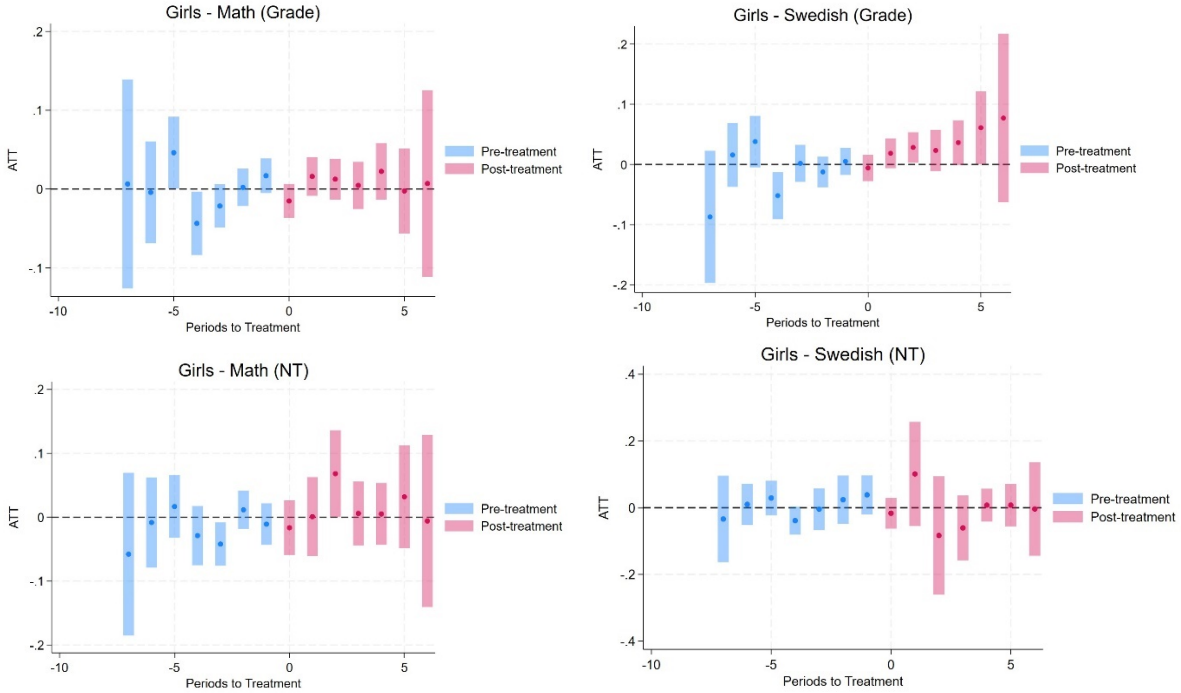


Figure C3

Bottom achievement quartile

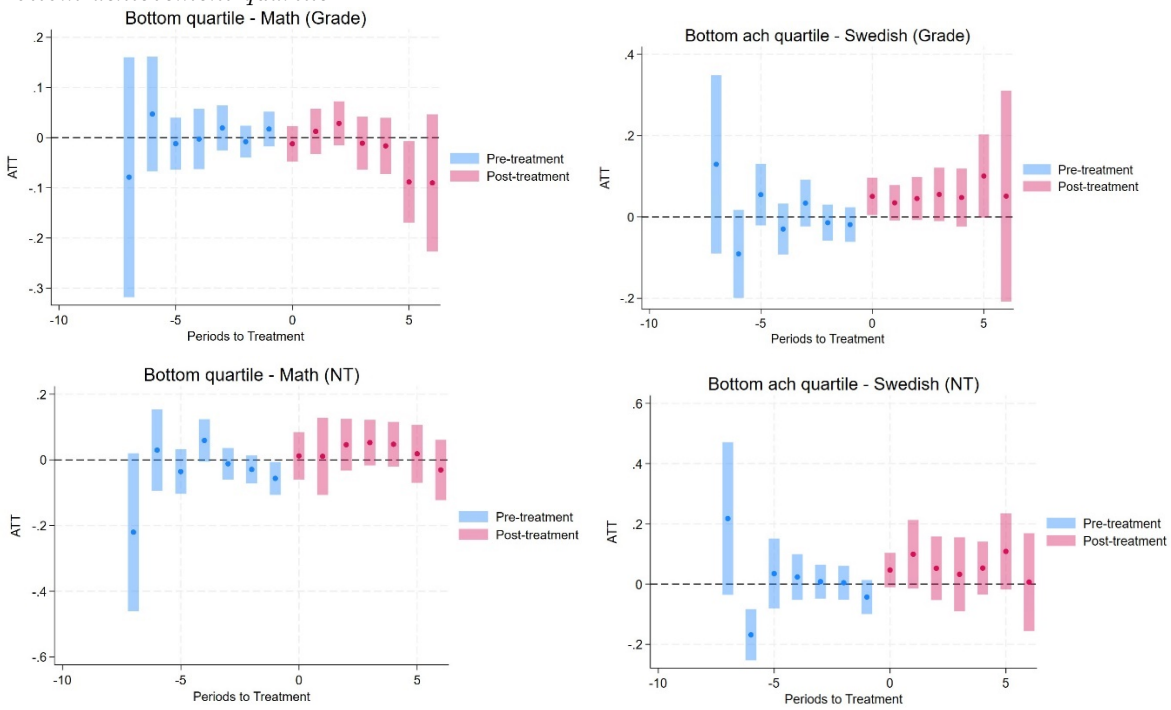


Figure C4

Top achievement quartile

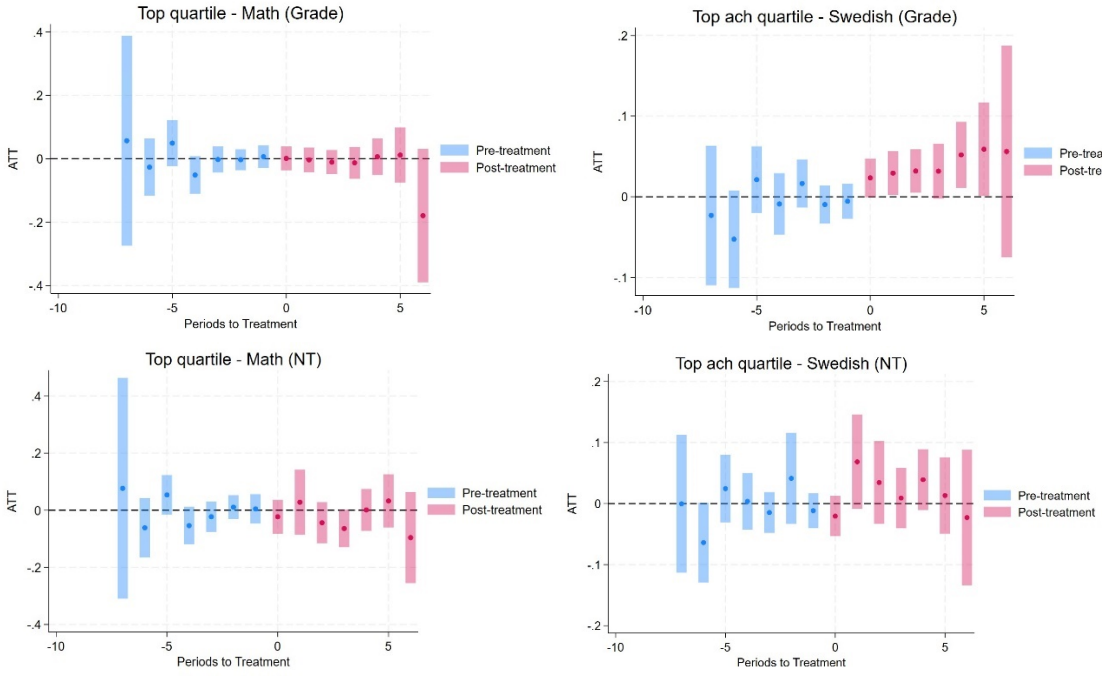


Figure C5

Independent schools

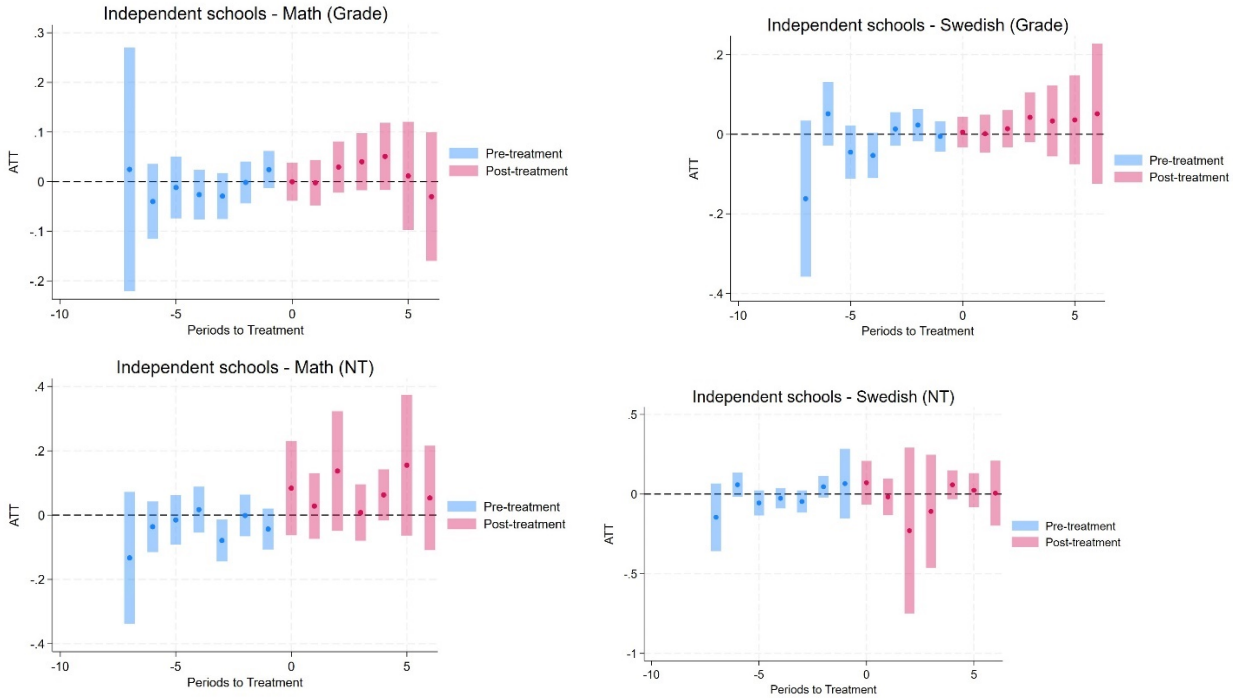


Figure C6

Public schools

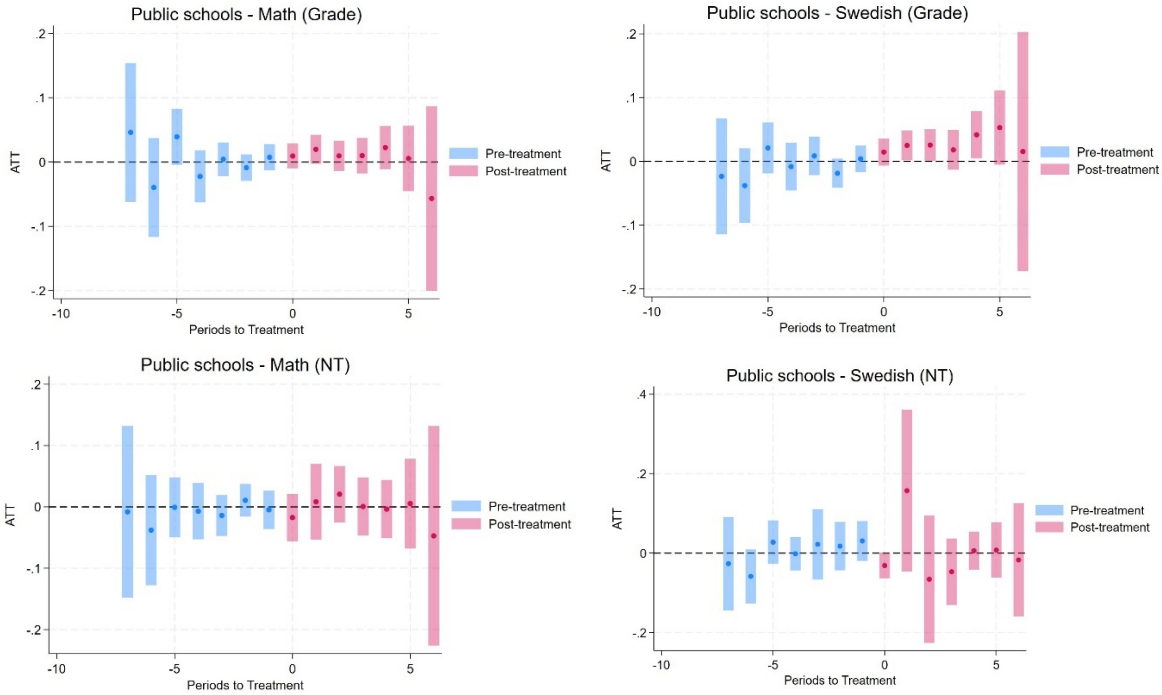


Table C1

Estimated SFL effects by subgroup

	(1) Bottom ach quartile	(2) Top ach quartile	(3) Independent school	(4) Public school	(5) Low teacher certification	(6) Low teacher experience	(7) Boys	(8) Girls
Grades								
Math								
ATT (SE)	−0.0066 (0.018)	−0.0055 (0.017)	0.0161 (0.020)	0.0126 (0.009)	0.0273 (0.018)	0.0035 (0.019)	0.0188 (0.010)	0.0058 (0.010)
Observations	305,455	486,885	353,400	1,613,688	413,397	406,363	1,013,093	954,014
Pre-trend p-Value	0.077	0.563	0.000	0.710	0.555	0.451	0.151	0.054
Swedish								
ATT (SE)	0.0504* (0.021)	0.0342** (0.012)	0.0177 (0.022)	0.0254* (0.011)	0.0188 (0.0249)	−0.0136 (0.024)	0.0262 (0.011)	0.0210 (0.011)
Observations	305,455	486,885	353,400	1,613,688	413,397	406,363	1,013,093	954,014
Pre-trend p-Value	0.463	0.844	0.245	0.715	0.055	0.118	0.486	0.157
National test scores								
Math								
ATT (SE)	0.0278 (0.026)	−0.0172 (0.025)	0.0750* (0.036)	−0.0021 (0.018)	0.0577 (0.029)	0.0396 (0.031)	0.0158 (0.017)	0.0086 (0.018)
Observations	212,630	349,331	247,444	1,132,928	292,630	285,977	705,231	675,145
Pre-trend p-Value	0.019	0.004	0.030	0.475	0.185	0.009	0.272	0.002
Swedish								
ATT (SE)	0.0612 (0.033)	0.0172 (0.016)	−0.0173 (0.070)	−0.0013 (0.024)	−0.0178 (0.063)	−0.0999 (0.088)	0.0017 (0.025)	−0.0080 (0.024)
Observations	212,630	349,331	247,444	1,132,928	292,630	285,977	705,231	675,145
Pre-trend p-Value	0.601	0.037	0.216	0.433	0.420	0.139	0.689	0.057

Note. The table shows the estimated effects of SFL on pooled Grade 6 and Grade 9 math and Swedish grades and national test scores when pooled across all students. Scores have been standardized by outcome measure and year. *** = $p < .001$, ** = $p < .01$, * = $p < .05$

Supplementary material section D

Table D1

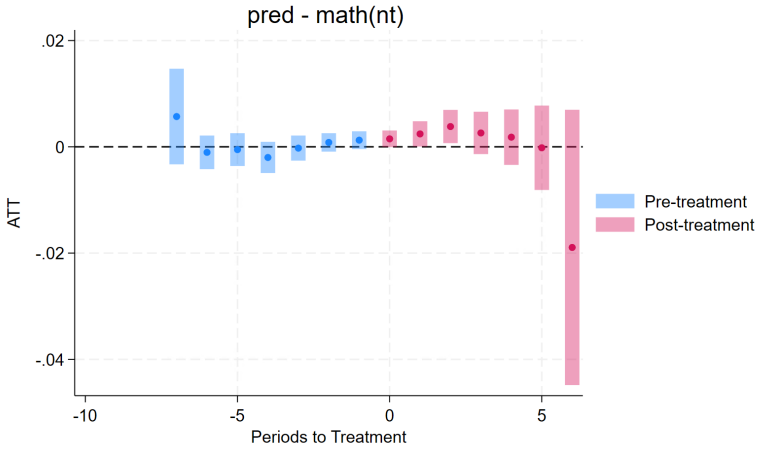
Estimated effects of SFL on predicted mathematics national test scores

	ATT	Standard Error	z	p-Value
Predicted math test score	0.002	0.0014	1.46	0.145

Note. Variables in the linear prediction models include gender, migration background, highest level of parental education, and school fixed effects. $R^2 = 0.16$.

Figure D1

Event study estimates of SFL on predicted test scores



Supplementary material section E

Figure E1

Alternative treatment specification - estimated event time effects

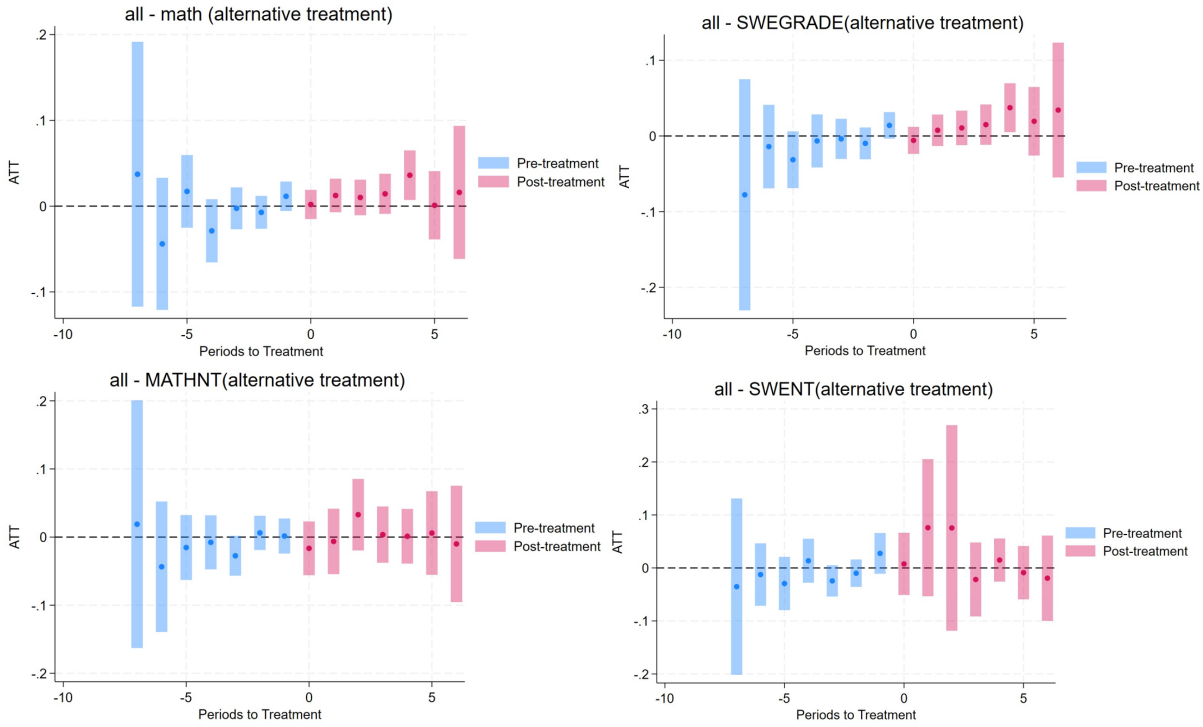


Table E1

Estimated effects of SFL for all students (alternative treatment specification)

	ATT	Standard Error	z	p-Value
Grades				
Math	0.0123	0.008	1.39	0.163
Swedish	0.0119	0.010	1.17	0.241
National test scores				
Math	-0.0001	0.015	-0.01	0.992
Swedish	0.0169	0.021	0.79	0.427

Note. The table shows the estimated effects of SFL on pooled Grade 6 and grade 9 math and Swedish grades and national test scores across all students. Scores have been standardized by outcome measure and year.

Table E2

Estimated effects of SFL for student and school subgroups based on alternative treatment definition (national test scores only)

	ATT	Standard Error	z	*p*-Value
IEP				
Math	-0.001	0.028	-0.04	0.968
Swedish	0.006	0.032	0.21	0.837
Arrived within 5 years				
Math	0.025	0.031	0.81	0.418
Swedish	0.061	0.039	1.57	0.116
Independent schools				
Math	0.029	0.032	0.88	0.379
Swedish	0.042	0.033	1.26	0.207
Low certified teachers				
Math	0.045	0.042	1.13	0.260
Swedish	0.019	0.040	0.49	0.628

Note. The table shows the estimated effects (ATTs) of SFL on grade 6 and grade 9 math and Swedish national test scores. Outcomes have been standardized by grade and yearly cohort. The change in effect column refers to the change in effect as compared to the main treatment definition (conditioned on a coach being assigned to the school). The sensitivity checks are only run for subgroups where statistically significant effects were found in the model specification including main treatment definition.

Supplementary material section F

Table F1

Estimated effects of SFL for students in small schools

	ATT	Standard Error	z	p-Value
Grades				
Math	0.0252	0.0222	1.13	0.258
Swedish	0.0113	0.0272	0.42	0.767
National test scores				
Math	0.023	0.0455	0.52	0.600
Swedish	-0.018	0.0369	-0.49	0.625

Note. Schools are classified as small if they fall within the bottom quartile of student enrollment (a pre-treatment 2-year average).

Supplementary material section G

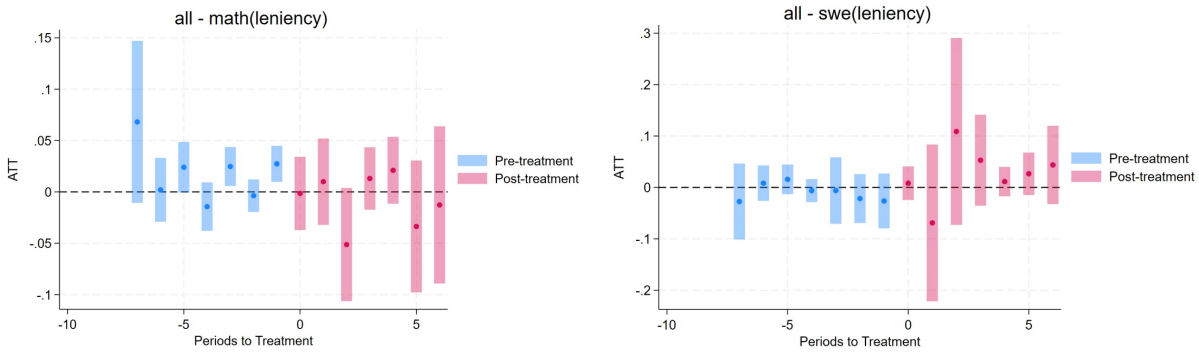
Table G1

Grading leniency

	Coefficient	Standard Error	z	p-Value
Math	-0.0019	0.0123	-0.15	0.877
Swedish	0.0197	0.0209	0.94	0.346

Note. Grading leniency is operationalized as the difference between cohort-year standardized grades and national test scores.

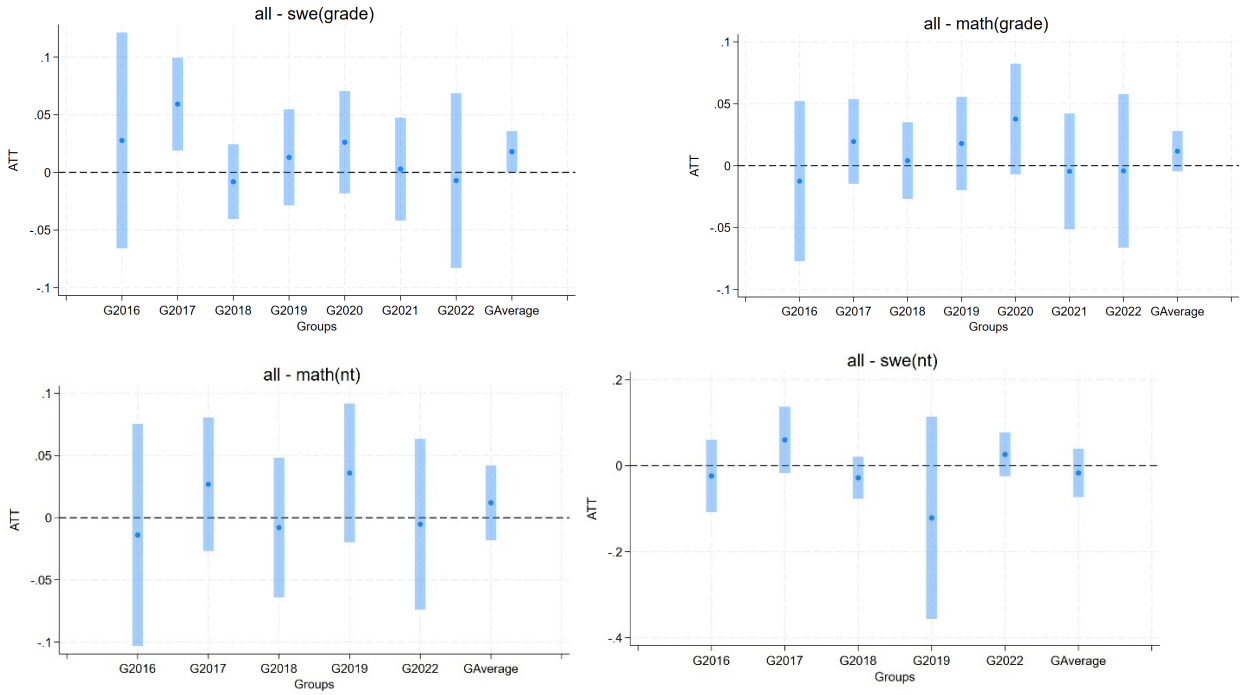
Figure G1



Supplementary material section H

Figure H1

Cohort-specific effects (all students)



Supplementary material section I

Figure I1

All students

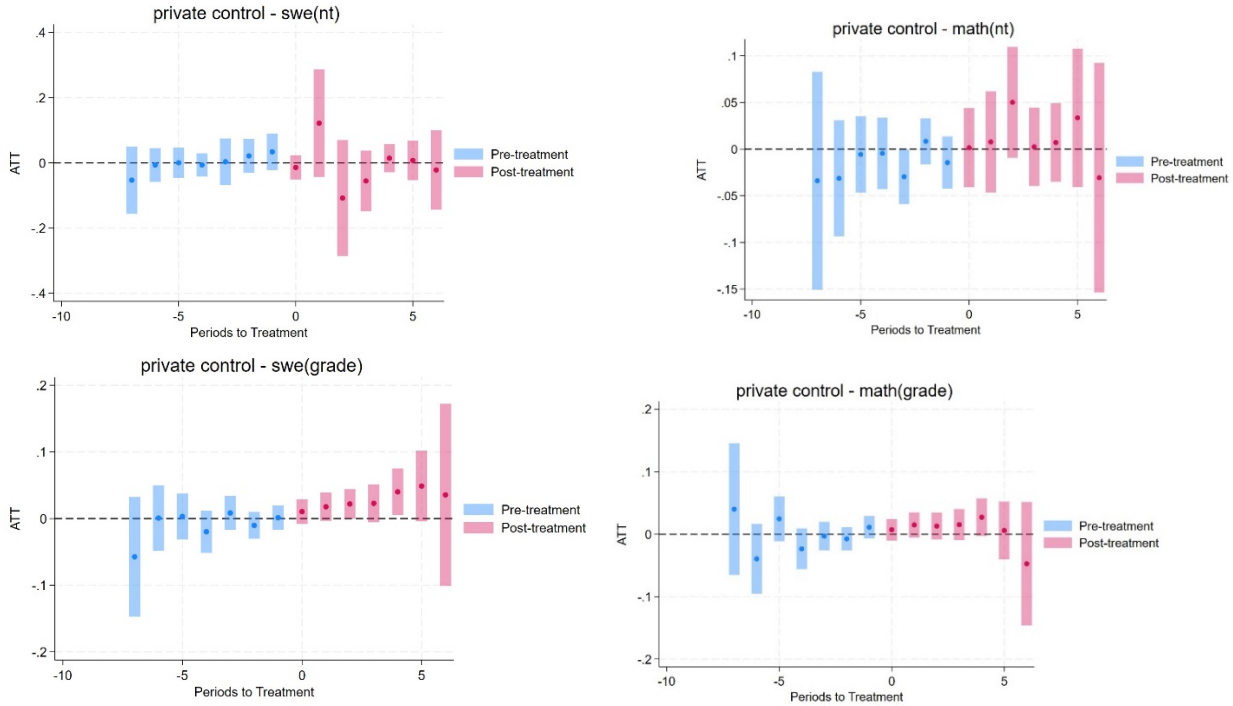


Figure I2

Students with IEPs

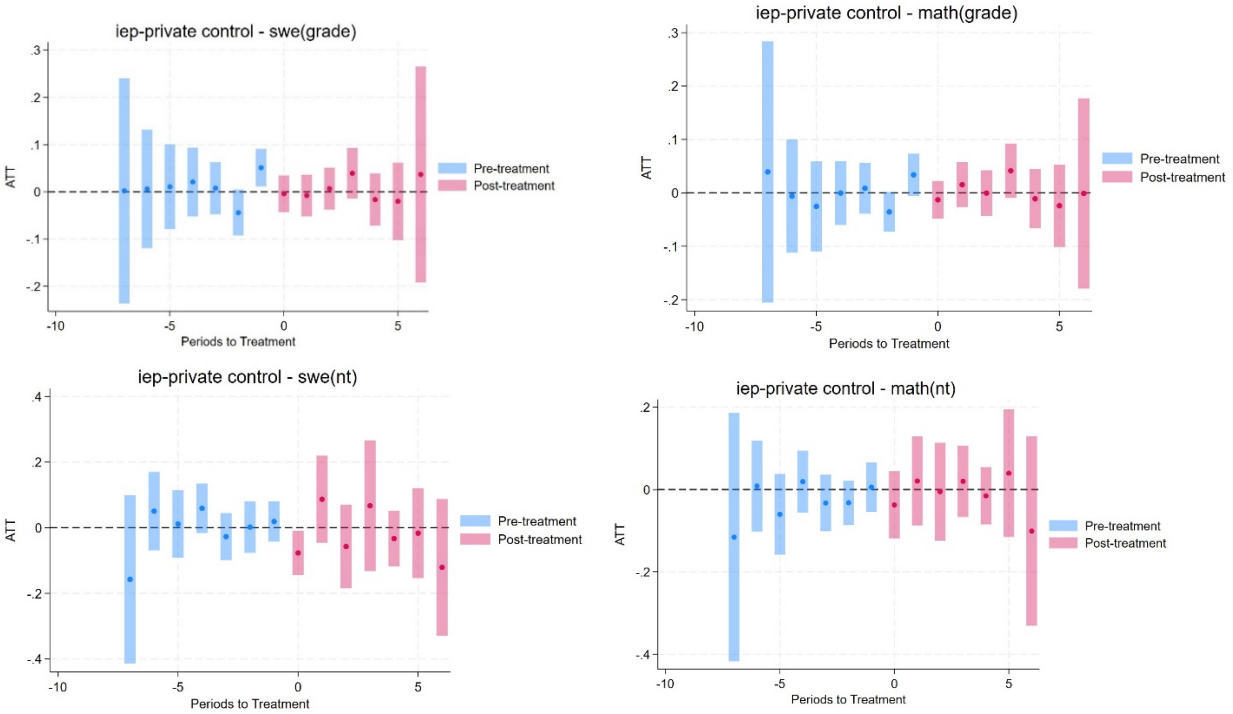


Figure I3 *Students arrived within 5 years*

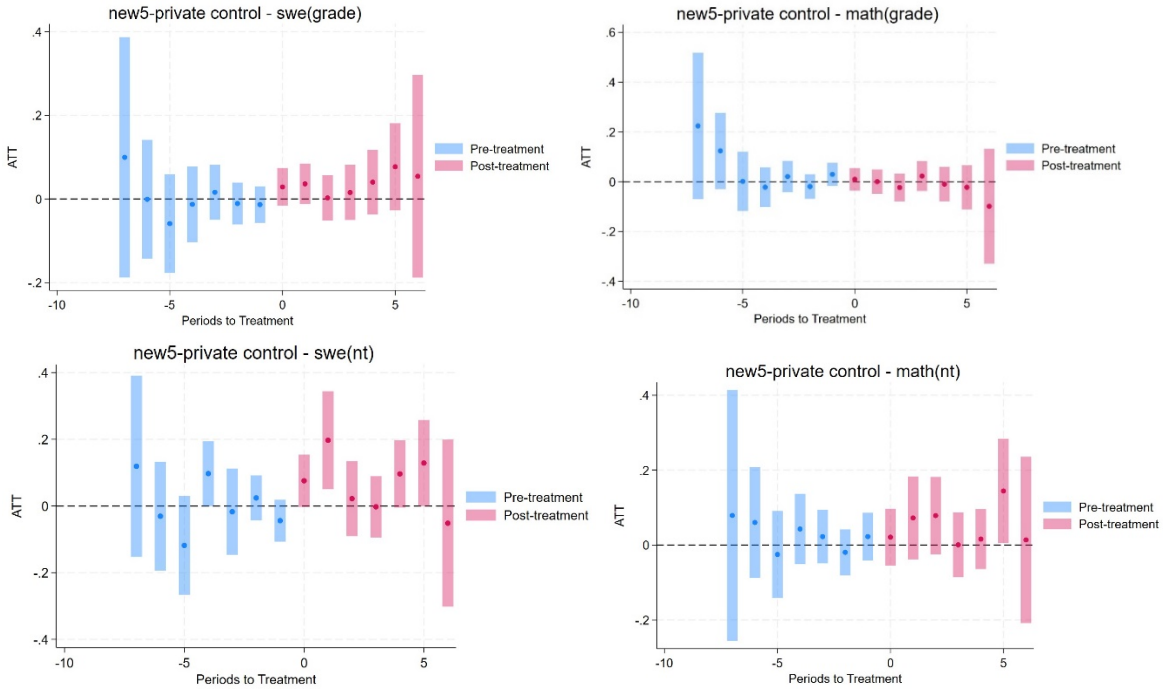


Figure I4

Childhood immigrants

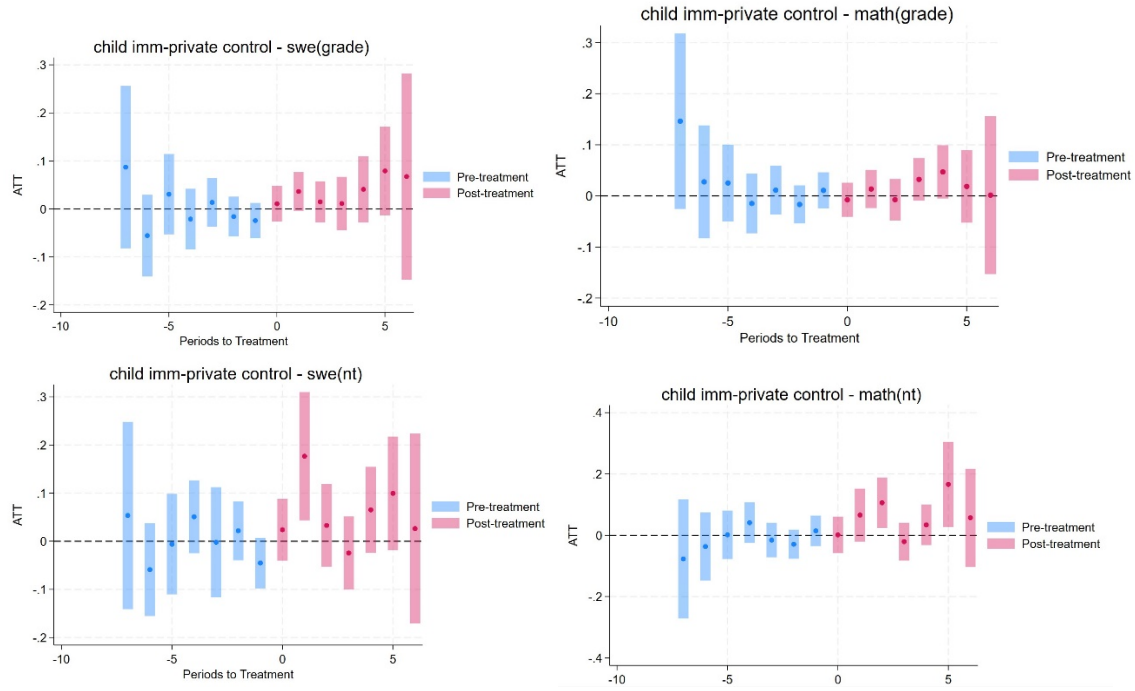
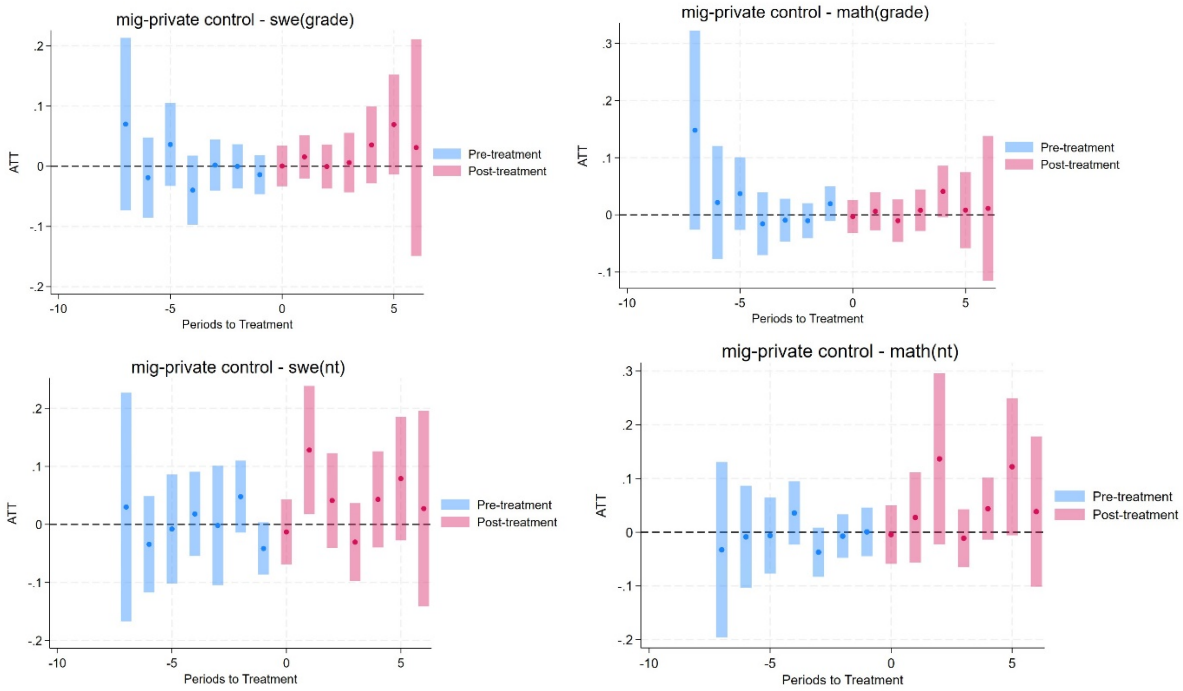


Figure I5

Migration background



Supplementary material section J

Table J1

Estimated effects of SFL for all students by grade

	ATT	Standard Error	p-Value	Lower 95% CI	Upper 95% CI	Observations	Pre-trend p-Value
National test scores							
Math (Grade 6)	0.015	0.016	0.373	−0.017	0.047	734,221	0.132
Math (Grade 9)	−0.002	0.028	0.945	−0.057	0.053	600,830	0.041
Swedish (Grade 6)	0.005	0.015	0.740	−0.025	0.035	728,920	0.684
Swedish (Grade 9)	−0.049	0.057	0.394	−0.160	0.063	728,920	0.096
Grades							
Math (Grade 6)	0.013	0.012	0.265	−0.010	0.037	1,002,438	0.004
Math (Grade 9)	0.010	0.012	0.413	−0.014	0.035	964,683	0.027
Swedish (Grade 6)	0.009	0.014	0.670	−0.018	0.036	1,002,980	0.143
Swedish (Grade 9)	0.037	0.014	0.008	0.009	0.065	966,974	0.091

Note. The table shows the estimated effects of SFL on Grade 6 and Grade 9 math and Swedish grades and national test scores when pooled across all students. Scores have been standardized by outcome measure and year. Standard errors clustered by school.