# School Autonomy and Subject-Specific Timetables*

Iman Dadgar †

August 11, 2021

**Abstract**

This paper studies the effect of a reform that increased school-level autonomy in determining how to allocate time between different subjects – i.e., constructing a subject-specific timetable. Using registry data from all students in Sweden in a Difference-in-Differences framework, we found that students' short-term outcomes were not affected by the reform. However, there is a positive effect on long-term results, especially for students from low socio-economic status households.

*Keywords:* School autonomy, Decentralization, lesson plan, Student outcomes

*Acknowledgement: I thank Karin Edmark, Matthew Lindquist, Thor Norström, Roujman Shahbazian, Dan-Olof Rooth, and Malin Tallås Ahlzén. I have benefited greatly from comments from seminar participants at the Institute for Evaluation of Labor Market and Education Policy (IFAU), the Swedish Institute for Social Research (SOFI) at Stockholm University, and the Department of Economics at Stockholm University. I would also like to thank Mari Eneroth and Sofie Burman for their help in gathering information concerning the timetable reform. Financial support from Handelsbankens forskningstiftelser and from IFAU is gratefully acknowledged.

†The Swedish Institute for Social Research (SOFI), Stockholm University, 106 91 Stockholm.iman.dadgar@sofi.su.se

# 1 Introduction

Because high-quality education is perceived as vital to a country's success, children spend a considerable amount of their time in school. Therefore, how much time is allocated to specific subjects needs careful consideration. Recently, research has focused on how increasing time spent on specific subjects (often focusing on STEM-related topics) affects students' future employment and earnings (Kirkeboen et al. ((2016)). Little attention, however, has been given to who should decide how to allocate time across subjects. That is, should subject-specific timetables be determined and regulated by the central government and therefore homogeneous across schools or should each school decide how to allocate the teaching time between subjects?

A school-level schedule solution has the potential advantage of enabling schools to adjust their subject-specific teaching time to the local student's needs based on the fact that schools have more relevant and detailed information about their students (Hanushek et al. ((2011))). However, a school-based timetable increases each school's workload and administrative activities, as teachers and principals need to determine the timetable (Zabojnik ((2002))). In addition, there is a risk that inequality will increase when the central government provides fewer regulatory guidelines as some schools might not be qualified to decide about the details of the timetable (Lundahl ((2002))).

Few empirical studies have addressed this question. Hanushek et al. ((2011)) use panel data from 42 countries observed annually for ten years and investigates the effect of increased school autonomy on student outcomes. They found that more autonomy positively affects student performance in developed countries, but more autonomy negatively affects performance in developing countries. Fuchs and Wöbmann ((2007)) use the PISA database to examine the effect of school autonomy on student performance. They found that school autonomy concerning managing personnel, teachers' decisions, and textbooks positively affects student outcomes, but school autonomy has adverse effects in areas with a strong potential for opportunistic behavior, such as forming a school

budget.

This study evaluates a Swedish policy that transferred the decision-making authority over time allocation across subjects from the national to the school level. The experiment took place in 900 of Sweden's approximately 3000 primary and lower secondary schools in the early 2000s. The central government implemented the experiment to evaluate the effects of allowing schools to allocate total teaching time across subjects. Several studies on the reform implementation were carried out in connection with the trial; these findings are reviewed in section 2.3. However, no comprehensive evaluation has been performed on the long-term impact of reform using extensive microdata on students, a gap in knowledge this study fills.

This study evaluates the impact the timetable reforms had on students' grades, the probability of entering a STEM field, the probability of finishing upper secondary school, and years of completed education at the age of 25. Here, we study the average effect on the students in the participating schools and estimate the impact separately for students from different socioeconomic statuses (SES) and academic qualifications. The latter is essential for determining if the policy decreased the inequality in the system by primarily benefiting low-achieving students or if it was more beneficial for already high-achieving groups and therefore increased educational inequality.

Because the schools were not randomly selected (i.e., they were self selected), the experiment is assumed to be influenced by school-specific characteristics. Our analysis is based on a Differences-in-Differences design (DiD) that includes cohort- and school-fixed effects. This method controls for the time invariant influence of all unobservable characteristics constant over time. DiD assumes that the outcome variables follow parallel trends in the absence of treatment. We checked the main requirement of DiD model by testing if the trends of the outcome variables are parallel before the policy implementation; we found this to be the case for three of four of the outcomes.

Although the school-based timetable had no impact on short-term outcomes, it did

generate long-term benefits. Compared to pupils who were not exposed to the reform, pupils who were exposed to the reform during primary school had a 2% higher probability of studying STEM subjects and a 1.56% higher probability of completing high school. In addition, females from low SES and both male and female students from large schools gained more from the program. The non-linear impact of the treatment study found a clear dose-response relationship: more exposure was associated with more extensive effects. Moreover, two robustness checks – one divided the study period and one matched the control group to techniques for the choice – produced similar results.

This study makes several contributions to the literature. First, this is the first large-scale micro data study that looks at school autonomy regarding the use of timetables. Second, the policy was implemented as a quasi-experimental scheme where only some municipalities could participate. That is, this study was able to control many potential confounding factors, moving closer to the causal effect of interest. Third, this study delivers novel evidence on the long-term impacts of decentralization policy. Most of the existing evaluations of decentralization policies are short-term and mainly focus on achieved grades. Fourth, this study's focus on how policy impacted all students made it possible to investigate heterogeneity in the school-based timetable's effect across parental income, parental education, gender, and immigration status. Finally, this study's sample sizes are likely large enough to estimate the potential effects of the policy precisely.

The rest of the paper is organized as follows. In the next section, we describe how the timetable reform was implemented. Section 3 presents the data, and section 4 presents our empirical strategy. In section 5, we check the validity of the empirical strategy. The results are presented in section 6. Section 7 contains the robustness check, and section 8 presents the conclusions.

# 2 The Swedish Time Schedule Pilot Project

## 2.1 The content and background of the reform

The timetable reform under study in this paper was implemented in the context of increased decentralization of the Swedish public sector. The reform was preceded by three main decentralization reforms of the education system (see sa Ahlin and Mörk ((2008)) and **?**). In 1991, the responsibility for compulsory schools was shifted from the national level to the municipality level. In 1993, the grant structure changed; municipalities were paid a general grant rather than a target-based grant, so the municipalities had more flexibility in how to spend the grant money. Finally, in 1996, the wage-setting decision was shifted from central negotiation to school managers and municipalities. These reforms meant that Sweden went from being regarded as having one of the most centralized education systems among the OECD countries in 1990 to having one of the most decentralized education systems in 1999 (OECD [1998]).

At the end of the 1990s, the Swedish government decided to go one step further and implement a pilot project in which some schools would be given the authority to shape the timetable. Under the existing system, the timetable was determined centrally by the Ministry of Education, although schools were free to decide about some dimensions. The schools participating in the pilot project were given expanded decision power over the timetable and were free to schedule the time allocated to different subjects. The sole restriction was that the total instruction time needed to reach the minimum specified hours (SOU 2005:101).

The project was based on a government report published in 1997 (SOU 1997:121), which proposed running a pilot project where a set of schools would be allowed to depart from the national schedule. The project's primary motivation was that a strict schedule was not compatible with the goal-based grading system implemented in 1998 for the compulsory education system nor with the goal-based grading system implemented in

1994 for the upper secondary education system. It was argued that merely participating in a class for a particular duration does not necessarily mean that students learn the subject deeply or broadly [REF ]. It was also argued that schools needed more freedom to decide how the total teaching time was to be allocated between subjects to encourage the new, more goal-oriented education system. The hope was that the pilot project would yield valuable information and experiences that could help determine whether a decentralized timetable system should be implemented for all schools in the country.

## 2.2 Selection of participating municipalities and schools

The government proposition to run the policy experiment was approved by the Parliament on November 3, 1999. The policy was implemented in two steps. First, the Ministry of Education invited all municipalities to apply for participation in the experiment. Of the country's 289 municipalities, 79 applied; of these, 70 were selected to participate. Second, selected municipalities were allowed to decide which of their schools would participate in the experiment. It was basically up to the municipality to choose which schools were to be involved. Some municipalities selected all schools to participate, and some municipalities selected a limited number of schools to participate (Figure A.1 ).

Although it is not known how the school-level decisions to participate were made in the 70 participant municipalities, a study by Rönnberg (2007) sheds light on this process in 16 of the municipalities. The absolute majority of school leaders surveyed in the study emphasized that schools were selected based on their willingness; that is, interested schools that volunteered were chosen to participate. However, there are three exceptions to the voluntary notification procedure for schools. In these three municipalities, either all schools participated or the municipality specifically selected schools for participation without sending a general request to every school. In total, 183 schools (20% of the students) participated in the pilot experiment in these 16 municipalities.

In the municipalities that allowed the schools to decide whether or not to participate in the experiment, several reasons for participating were documented – e.g., to increase individualized learning, to increase interdisciplinary work, to support a more comprehensive and overall view of learning, to develop new methods of instruction and learning, and to increase students' interests and responsibility. The main aim was to increase students' opportunities to plan their own work and let the content of the work control the planning of the day. For these municipalities and schools, the timetable represented a time constraint that was inconsistent with a goal-driven and results-driven system.

## 2.3 How did the participating schools implement the pilot policy?

Unfortunately, there is no comprehensive, centrally available information about how all the schools that participated in the pilot project changed the timetable and other related aspects. However, a relatively large number of studies were carried out in relation to the implementation of the policy experiment, and this section reports what we know about the implementation in some of the participating schools based on these studies. These were primarily written between 2000 and 2007, and some of them were part of the project Schools Without a Nationaly Set Timetable (Skola Utan National Timplan, SKU).

In some schools, teachers would meet with one student at a time to discuss what activities would best meet the student's needs and interests [REF]. These meetings were intended to help students work in a more goal-oriented and independent manner. Often, students had a mentor or supervisor and the schools planned and recorded the students' weekly work in logbooks or planning books that a parent was required to sign. Planning time became a natural part of the schedule, and students, with the help of their teachers, often created their own individual development plans. In addition, some schools also allocated more time to interdisciplinary sessions.

Most schools stated that they developed their work teams extensively, that the work teams gained greater authority and independence, and that the work teams engaged in further pedagogical discussions. Another clear tendency was to increase physical activity. Several schools created profile classes where some subjects or subject areas were given increased time, such as Sports, Culture, Media, Languages, and Mathematics. Several schools stated that they used a flexible starting time in the morning and ending time in the afternoon and that this contributed to a calmer environment. Other elements mentioned more often in the reports are age-integrated teaching, more extended coherent teaching sessions without interruptions, level groups, and help with homework in the afternoons (Skolornas arbetssätt; SOU 2005:101).

Of the participating schools in 16 municipalities, just over half (57%) of the 55 principals said that participation in the pilot program meant that the time distribution changed to a large or considerable extent at their school, but about 40% said that the changes were small in both these aspects [REF]. Unlike the principals, the ten surveyed teachers' association representatives believed the pilot policy did not result in any significant changes: eight representatives marked the changes as marginal and noted that the changes mainly concerned redistribution of subject-specific time to student's own work sessions (stugtid). A longitudinal study of three schools complemented the description that changes mainly related to modified working methods and only slightly to rearrangements of subjects and students (Lundahl ((2005)); SOU 2005, 102).

Based on interviews with students and teachers,Elmeroth et al. ((2005)), looking at how participant schools reallocated the time, found that the timetable pilot may have had a different impact on different subjects. In several of the studied experimental schools, the school day was divided into subject time, individual time, and common time. Both students and teachers believed that Social Studies, Swedish, and Swedish as a Second Language had benefited from the experimental set up. In these subjects, the knowledge goals were perceived as evident – i.e., it was easy for students to continue with

the tasks in these subjects during their individual study time. The teachers in science subjects, however, were concerned that the skills learned risked being superficial as the time for these subjects were perceived as scarce and insufficient for conducting demonstrations, experiments, and laboratory work. For the language subjects, the difficulties depended partly on the limited ability to perform verbal exercises during the specific time and other teachers' lack of knowledge of different languages. The latter becomes a problem if students need help with languages during sessions led by teachers who do not know these languages (Elmeroth et al. ((2005))). Alm Alm ((2003)) points out that mentor time has established itself in the trial schools for grades 7–9, and students generally call for more trusting relationships with their teachers (Skolornas arbetssätt; SOU 2005:101,)

In sum, the pilot project seems to have induced schools to implement more individualized and student-led sessions, where students work individually according to their plans and progressions, but less time was devoted to specific subjects. However, it is difficult to know if these changes will be viewed as substantive or marginal and to what extent the experiences of the schools surveyed in the above studies are generalizable to all 900 participating schools.

# 3 Data

The study is based on panel data covering all grade 9 students in Sweden from 1990 to 2010. These data encompass students born between 1975 and 1998 as students start grade 9 at age 15 and include almost 100,000 students yearly[1]. All observations with missing information on student background characteristics (about 4%) were dropped from the analysis. This section describes the three categories of variables used in the regression analysis: the treatment variables, the outcome variables, and the control vari-

[1] Some students may start school early or late and therefore be a bit younger or older

ables.

## 3.1 The treatment variables

We define the main treatment variable and two alternative treatment variables. The first is defined as a simple dummy variable for cohorts who attended grade 9 in a treated school and in post-treatment years. The two alternative variables are added to take into account the treatment intensity. The first alternative variable captures the treatment's linear effects and is defined as the number of years each student can be assumed to have studied in the treated schools and post-treatment. The second alternative variable captures the nonlinear effect of the treatment and is defined as a three-level indicator. The indicator takes a value of one if a student can be assumed to have been treated between one to three years after her compulsory school career (grade 1–9), the indicator takes a value of two if the student can be assumed to have been treated between 4 and 6 years after her compulsory school career, and the indicator takes a value of three if the student can be assumed to have been treated between 7 and 9 after her compulsory school career.

Since the data used for this study only provide links between schools and students measured at the end of grade 9, the end of compulsory education, the treatment intensity variables are based on the school the student attended grade 9 in combination with school-level information on what grades are offered by the school. For example, a student who finished grade 9 in school A in 2009 is assumed to have been subject to six years of treatment if school A offered grades 4–9 but not grades 1–3. In other words, for such cases, we assume that the school the student attended in grades 1–3 was not a treated school. That is, some error in this variable is likely because the student may not have attended school A throughout grades 4–8 and because the 1–3 grade school the student attended may have been treated. Therefore, the treatment effect will either be downward or upward biased, depending on what is more prevalent.

Although the intensive treatment variable has this flaw, we still find it valuable to use this variable to approximate the impact of treatment duration on the effects. Columns 4, 5, and 6 of Table B.1 in the Appendix shows the number of students each year who belong to these treatment groups. Almost 18% of the students were in the treated group according to the simple treatment dummy variable. Of these, 50% were exposed to the treatment for less than three years, 26% for four to six years, and 23% for more than seven years. The distribution of the affected students across municipalities is shown in Figure A.2, where municipalities are shaded according to the share of students who attended schools that participated in the pilot project. There is a high variation between municipalities in this group. For example, in Enköping and Sundbyberg, almost all students were involved in the experiments (dark blue), whereas in Västerås no more than 10% of the students were exposed to the reform.

## 3.2 Dependent variables

A school-specific timetable could affect students' academic outcomes in the short term since schools could allocate more time to certain subjects or help students in private sessions. This might affect student grades and graduation rates, the study track students attend in upper secondary education, and how long students continue studying. To capture these effects, four dependent variables were used. First, the immediate effect on students' school results is measured by the final grades awarded when students finish grade 9. We study the impact on the final grades in different subjects and the grade point average (GPA)[2]. Since the grading system changed in 1998[3], which resulted in a substantial change in the GPA distribution, the outcome measure is defined as the

[2]GPA is measured as the mean value of 16 subjects.

[3]Before 1998, five levels in each subject from 1 to 5 (1 is the lowest and five is the highest). After 1998, the grading system became "goal-based" – i.e., the evaluation was based on pre-specified goals defined for each topic, and students were graded on a 4-level scale, with the lowest being a "Fail" grade.

student's percentile rank in the cohort to generate an arguably more consistent measure over time than a standardized measure. For the subject-specific grades, we use two outcomes: getting a passing grade and the probability of getting a high grade. For these two variables, the sample was restricted to 1998 and onwards, when the new grading system was introduced. Getting a passing grade means students did not fail the course. Getting a high grade means getting the grade VG (Pass with distinction) or MVG (Pass with special distinction). The second outcome variable is a dummy variable that indicates whether students attended an academic STEM track[4] in the first year of upper secondary school. In Sweden, students select their track after finishing compulsory school (i.e., after grade 9). The selection is based on their GPA and interests. This variable, however, only shows a student was accepted, not whether they finished the track.

The third and fourth dependent variables are the long-term outcomes that include years of education and a dummy variable to indicate if a student finished high school. These two outcomes are measured when students are 25 years old and includes all students who attended grade 9 between 1990 and 2007.

## 3.3 Control variables

Several control variables that are likely relevant for the outcomes are included in the model. These controls are gender, age, immigration status, parents' income, and parents' education. All student's background characteristics were measured when students finished grade 9 [5]. Empirical and theoretical studies have shown that these variables are highly correlated with educational outcomes in Sweden. We also use a set of variables

[4]This is defined as attending the Science or the Technical Science tracks in the fall of grade 10

[5]This means that we assume that the treatment itself did not affect these covariates. Given the nature of the covariates (measuring either students' characteristics such as age and gender or the parental education level and income), this seems like a fairly innocuous assumption.

capturing municipality characteristics: average income in the municipality, the share of households under social assistance, the unemployment rate, the population size, the land area, the percentage of immigrants, and the share of left party (Social Democratic and the Left Party) votes on city council in 1998.

# 4 Empirical Strategy

To identify the policy's causal effect separately from potential confounding factors such as the quality of schools and unobserved parental and student background variables, we use a Differences-in-Differences (DiD) estimation strategy. This model allows us to control for fixed unobserved variables that could be correlated with the policy and educational outcomes. The DiD equation for estimating the average treatment effect of the reform is

$$y_{ist} = \alpha + \beta_b Tr_{sb} + \beta_0 Treat_{st} + \beta_1 X_i + T_t + S_s + \varepsilon_{ist} \tag{1}$$

$y_{ist}$ are the outcomes of student $i$ in school $s$ at time t. $Treat_{st}$ is a treatment variable, defined as one for students who attend a treated school in a post-treatment year and zero otherwise. $S_s$ is a school-fixed effect that controls for permanent school characteristics that can affect educational outcomes and $T_t$ is year of fixed effects handle time-changing factors that influence outcomes in the whole country at the same time, for example, nation-wide education policies. $X_i$ are student's characteristics such as gender, immigration background, parents' education and parent's income, and $\varepsilon_{ist}$ is an error term.

The parameter of interest is $\beta_0$, which captures the average effect of the policy on the treated group. The $\beta_0$ estimator can be seen as an intention to treat the effect. Although the $\beta_0$ estimator captures the impact of attending a school that participated in the timetable reform in a year when the reform was in place, we lack comprehensive and precise information about how the participating schools implemented the reform

(see section 2.3). Thus, we will not be able to tell if any estimated impact is due to any particular component of the reform, for example, if some particular change to the schedule was more or less beneficial to students. We will, however, discuss the results concerning the partial information on the reform available from the studies carried out in relation to its implementation (see section 2.3).

To get a quick indication of whether the parallel trends' assumption behind the DiD estimation holds for each outcome variable, a placebo pretreatment dummy ($Tr_{sb}$) was added to the regression equation. This variable is defined as one for students graduating from a treated school (i.e., a school that will eventually be treated) in the first half of the pretreatment period, 1990–94, and is zero otherwise[6][7]. Non-zero $\beta_b$ estimates suggest that the outcomes of the treated and untreated groups of students develop differently before the reform and therefore suggest that the assumptions of the DID specification are not fulfilled. Non-significant estimates, on the other hand, indicate that the specification is valid. The placebo coefficient is reported in all baseline results in the result tables. We will return to the issue of pretreatment trends below.

As explained in the previous sections, we use both the simple treatment dummy variable of a regression equation 1 and the two alternative treatment variables that were described in section 3.1. In addition to the average treatment equation above, several heterogeneity regressions are estimated by running the model separately for subsets of students to test whether the effect differs across the groups of girls/boys, low-income parents, highly educated parents, and immigrant students. We also run this model for students with different qualifications determined by the students' GPA (i.e., rank) to see how students with diverse educational ranks react to the treatment.

[6]This means that the reference period is 1995–1999, and both the placebo coefficient and the treatment effect are estimated in relation to this period.

[7]One of the studied outcomes – i.e., the likelihood to enroll in a STEM track in the first grade of upper secondary school – had no observations prior to 1995, which means that the placebo estimate is omitted from the estimation for this outcome.

# 5 Validity of the Empirical Strategy

The DiD model assumes the outcome variables in the control and treatment groups follow parallel trends in the absence of the treatment. This assumption cannot be tested formally as the counterfactual state of "no treatment" for the treated group in the treatment period cannot be observed. However, an informal way to evaluate the likelihood that this assumption holds is to study whether the pretreatment period trends are parallel – i.e., determining whether the differences in school outcomes between the control and treated groups are constant over time before the introduction of the reform. In addition, the DiD model assumes that no other changes were implemented at the time of policy implementation that would confound the estimation of the reform effects. This section provides a detailed evaluation of the assumptions underlying the DID equation.

## 5.1 Reform selection process

The treatment selection process for this pilot project needs to be considered when evaluating the more likely threats to the validity of the DiD model. As described in section 2, a selection was made at the municipality level before the school level. As a result, there are two types of non-participant schools: non-participant schools in participating municipalities (the left branch of the flowchart) and schools in the non-participating municipalities (right branch of the flowcharts) (Figure A.1). Judging from the studies referenced in section 2, it is likely that many of the schools in the former group could have participated in the experiment if they chose to; we call this Group A. We call Group B the schools that did not have a chance to participate in the experiment due to decisions made at the municipality level.

Using either of these control groups or the combination of the two comes with advantages as well as disadvantages. Control group A helps control for the schools' municipality-related unobservable characteristics since these groups are in the same

municipalities as the treated schools. This is beneficial since the municipalities in Sweden are responsible for compulsory school provisions. These schools have the same local school funding system, are subject to the same municipality-level decisions, and share the same demographic and economic environment. The disadvantage is that these schools potentially voluntarily chose to opt-out of the treatment. This increases the likelihood of selection bias at the school level. Schools in control group B, on the other hand, are not in the same municipalities, which means that they may be subject to different municipality-level trends. However, they were not allowed to participate in the experiment even if they wanted to, which means that we might have found good matches in terms of characteristics correlated with school-level selection in treatment at the school level within this group.

We will use the following estimation strategy with respect to the definition of the control group. First, we evaluate the appropriateness of both of these potential control groups as well as of the combination of the two based on how well they balance with the treatment group with respect to pretreatment characteristics and trends. This evaluation will determine what control group is used for the main estimations of the paper. Second, we will present some results for the groups that are not chosen as the main strategy to evaluate the robustness of the main results. Third, we will further investigate if the results are sensitive to the composition of the control group by combining the DiD with a matching strategy to increase balance in a robustness section.

## 5.2 Descriptive statistics and balance check

This section first presents descriptive statistics for the municipalities and schools in treated and (potential alternative) control groups. It then evaluates whether the assumptions underlying the DID specification of equation 1 holds. The following information is depicted in Table 1. The municipalities selected for treatment had a higher population density, more students per school, and more immigrant students. In addition, the

standard deviation of most variables is much higher in the participating municipalities, which means the variation within this group is higher than within the non-participating group. Finally, based on the 1998 election, the selected municipalities have a slightly higher vote share of right leaning political parties.

The school-level characteristics for the three alternative control groups are depicted in Table 2. The schools in control group B are on average more similar to the treated schools, in particular when it comes to the share of immigrants, type of municipality (urban or rural)[8], GPA, and the number of students in each school. For a better sense of the differences, we calculate the normalized values for the differences of the average values of the characteristics for the set of alternative control groups. In other words, we scale the differences with the pooled standard deviation of the two groups. (See Imbens and Rubin (2007) for these normalized differences)[9]. In Table 2, $\Delta_i$ indicates the normalized differences – i.e., the smaller the value of $\delta_i$, the more similar the control group is to the treatment group regarding that variable. For our study, this indicator also has another advantage: we can determine which group gives the smallest index to find the most comparable control groups. The bold font indicates the lowest value of $\delta_i$ in each row. Groups ALL (all control schools) and B are more similar to the treatment group than control group A.

What is more important for the validity of the DiD model is the balance in terms of changes over time in the treated and non-treated units. To understand whether the covariates of the students in the treated and control groups are balanced in terms of the type of time and cross-sectional variation used in the DiD analysis, we replace $Y_{ist}$ in equation 1 with the predetermined characteristics before the policy period (1990–1999).

[8] 1 is more urban and 9 is more rural.

[9] The normalized difference is calculated by $\delta_i = \frac{\bar{X}_t - \bar{X}_c}{\frac{(s_t{}^2 - s_c{}^2)}{2}^{0.5}}$, where $\bar{X}_t$ and $\bar{X}_c$ are the mean value of treated and control groups and $s_c$ and $s_t$ are the standard deviation of these two groups. This indicator's advantages are that it does not depend on the sample size (in contrast to the t-test) and is more informative in groups with a high number of observations.

That is, we regressed each covariant on the pretreatment placebo dummy (1995–1999)[10] and used only the years before the treatment period. Therefore, a coefficient value close to zero indicated that the treatment and control groups are more balanced in terms of the pretreatment period changes in the outcome variable. Figure 1 shows the results for this balancing test for the following variables: immigrant students, females, parents' incomes, and parents' education levels. Panel a shows all control schools compared with the treated schools, panel b shows sample A estimates, and panel c shows sample B estimates. The result is similar to the cross-sectional analysis above: the sample is more balanced relative to the treated sample when either group B or the total population is used as the control group, respectively, and the balance is a bit worse when group A is used as a control group. Statistically significant non-zero coefficients are not estimated for group B and the group ALL (i.e., all control schools). This finding provides additional support for the DiD estimation strategy.

## 5.3 Parallel Trends

The DiD model assumes that the outcome variables in the treatment and control groups would follow parallel trends in the absence of treatment. A straightforward way of evaluating the parallel trend assumption is to plot pretreatment period trends in the outcome variables for both treatment and control groups. That is, if the trends are parallel in the pretreatment period, we expect that they would have also been parallel in the post-treatment period in the absence of treatment. Figure 2 shows the trend of the four outcomes of interest for the treated schools, and control groups A, B, and ALL (control

[10]Note that the pretreatment dummy variable here takes value 1 for the years 1995–99, whereas in the regression specification (1) takes value 1 for 1990–94. The reason for the latter is that in the main specifications we wanted to use were the years before the treatment period 1995–99 as the reference period, so the treatment effect is estimated in relation to these years.

A+B) for 1995–1999 [11].

Control groups B and ALL reveal that schools in these groups follow almost the same trend in probability of low education and years of schooling. For the short-term outcome, the average final grade percentile rank in grade 9 for the treated schools is quite different from the control groups; in particular, the trends diverge in 1998, when the new grading system was introduced[12]. For the probability of attending a STEM track in the first year of upper secondary school, all control groups follow almost the same trend and are similar to treated schools. This finding suggests that care must be taken when interpreting the average grade percentile rank outcome results. To further evaluate the pretreatment trends and how they compare to the changes after the reform, event-study analyses are used (see the Results section).

Overall, the control groups "all schools" and group B are more balanced concerning pretreatment characteristics and trends than control group A. For the main results, we chose "all schools" to avoid dropping observations. We show results based on control subgroups A and B separately in the Results section. As a further sensitivity test, the Robustness section presents results after making the control group more similar to the treated group (i.e., matching the control group with the study group).

# 6 The effect of the timetable reform on the outcome variables

The empirical analysis begins by estimating the short- and long-term educational effects of the timetable reform. The section continues with the analysis of the reform's impact

[11] In order to facilitate the comparison, Figure A.3 shows the trend in covariant when the starting points of the trend are normalized to zero

[12] In section 7, we run the main model only using 1998 and 1999 as pretreatment periods for all the outcomes

in samples A and B. Section 6.2 checks whether the reform has heterogeneous effects for students with different background characteristics. This section ends with a discussion of the robustness checks used.

## 6.1 The Main Results

Equation 1 is used to conduct an event-study analysis but replaces the simple post-reform dummy with separate dummy variables for each year. If the assumption of parallel trends is valid, the estimated coefficients for these dummies should be zero for all pretreatment years ($\leq$ 1999). The estimated coefficients for the post-reform years ($>$ 1999), on the other hand, capture the development of the treatment effect over time. Figure 3 shows the event study analysis for percentile rank GPA, the probability of enrolling in a STEM field in upper secondary school,[13] the probability of graduating from upper secondary school within 12 years, and the years of education at the age of 25.

Figure 3 suggests that there are no statistically significantly different trends of the outcome variables in the treatment and control group pretreatment, but there is some evidence of divergence post-treatment for STEM, years of schooling, and upper secondary graduation (although the trend returns to the pretreatment levels after a few years for the latter outcome) . We take this finding as further support of the validity of the DiD and as a suggestion of a treatment effect of the timetables reform. However, we did not consider whether the students potentially received different treatment doses during their full compulsory school trajectory. Table 3 accounts for this and shows the effects of a school-based timetable on different short- and long-term outcomes for the first of the alternative treatment variables – i.e., when the treatment variable captures the number

[13]Note that years of schooling are computed based on information on the highest level of completed education at age 25. This means, for example, that a student who attended the first grade of high school but did not complete high school will be recorded as having the same number of years of education as an individual who attended zero years of high school.

of the years that students were (plausibly[14] ) exposed to the treatment. The placebo shows the dummy variable's estimates for all treated schools for 1990–1994. The full model (equation 1) estimated when year-fixed and school-fixed effects and control variables are added in all regressions. For the long-term outcomes, the estimated period is restricted to 1990–2007 to measure student outcomes at the age of 25.

One additional year of exposure to the school-based timetable does not affect the percentile rank. The second column shows that one additional exposure year to a school-based timetable increases the probability of attending an upper secondary STEM program by 0.0013. When calculated over the nine years of exposure (i.e., the whole of primary school education), the gain of the effect is approximately 5.6% and the probability of graduating from high school increases by 0.0012, compared to the mean value of the graduation share, which affects almost 1.56% of students exposed for nine years. The last column shows that the reform has a small positive effect on a student's years of schooling. This number was almost 1% when a student was in the same school for nine years.

The next step included three sets of dummy variables in the models. The treatment effect was divided into three levels: when students were plausibly treated schools for less than three years, 4–6 years, and 6–9 years (Figure 4). This table shows a similar pattern as the linear variable: the dummy variable indicating a high level of exposure is statistically significant and of the same sign as the linear effect. The point estimates are zero for students exposed to the treatment for fewer than three years (level 1). The effects start increasing after that, and the most considerable effects are seen for those exposed for seven to nine years (level 3). This finding suggests that if students are exposed to the treatment for more than six years, the effect comes close to the effects for nine years of exposure (cf. Table 3). For most of the outcomes between one and

[14] As explained in section 3.1, since we can only link students to schools in grade 9, these alternative treatment variables are based on assumptions of the students' school trajectories.

three years, the estimates are insignificant and much smaller.

How students were graded in various subjects was also analyzed by looking at two outcomes: getting a passing grade and the probability of having a high grade for several subjects[15]. Table 4 shows the reform's effect on the likelihood of pass and above pass grades and achieving a high grade (pass with distinction and pass with excellence) for three core courses – Mathematics, English, and Swedish. These subjects are important as students need to pass them to qualify for the regular upper secondary school educational tracks. The reform seems to correlate with earning a high grade in two of the core subjects. When a student was exposed to nine years of treatment, the effect size for receiving a high grade in Mathematics and English is 5.12% and 3.56%, respectively (Table 4). However, nine years of exposure to treatment did not affect treatment grades in Swedish.

## 6.2 Results for Sample A and Sample B

This section presents, for the sake of completeness, the results using two alternative control groups (A and B), which are described in section 5.1, rather than the full control sample. Figure A.4 and show the event study of sample A (panel a ) and sample B (panel b). Before the reform, in almost all outcomes, there is no relation between outcomes and the treatment variable, so these cases do not support the parallel trend assumption.

The treatment effect, however, looks different depending on whether sample A or B is used as the control group (Table 5). Using sample A, the results suggest no effect on the outcomes. On the other hand, using sample B results in a positive impact on the treatment for all four outcomes.

Given these divergent results, it is difficult to determine what effect to trust. Based

[15]The period of analysis is restricted to 1998–2010 because the grading system changed in 1998 and the passing and high grade definition changed. In the new system, students could get 4 grades – fail (f), pass, pass with distinction, and pass with excellence.

on the discussion in section 5, sample B's average school performance is more similar to the treated school compared to sample A before the reform. Because this finding suggests that sample B is a better comparison group, we had more confidence in these results. However, the fact that the result is sensitive to the choice of control group calls for caution. In section 7, we further evaluate how the results from using the alternative samples compare after combining the DiD with a matching strategy to improve balance.

## 6.3 Heterogeneity Effect

In this section, we examine whether some subgroups were more affected by the reform than others. According to the government, one of the experiment's motivations was to give schools the freedom to offer more help to students with a foreign background. The government also emphasized that students with low SES needed more individual instruction, which can be provided with the school-based timetable. Furthermore, according to the balance test in section 5.2, large schools were more likely to be treated. Perhaps, large schools need more flexibility since they have more diverse students, although this is only speculation.

To evaluate if the treatment effect differs between these groups of students, two heterogeneity analyses were performed: one based on the background characteristics of students and the school size and one based on the student's ability. The effect of the treatment on the two outcomes – probability of participating in the STEM field and the likelihood of graduating from high school – is estimated on separate samples for gender, immigration status, parent education (in four levels), family income (in four groups), and school size (in four-level) (Figure 6). Females, students from low-income families, larger schools, and graduating from high school were more affected by the treatment (Figure 6; Panel a and Panel b). The effect size for parents with low education and males was slightly larger than the other variables .

As the reform did not seem to affect students' percentile rank, it is unclear how

different students in the rank distribution reacted to the reform. The students were divided into quartile ability groups, from rank 1 (the lowest GPA) to rank 4 (the highest GPA) (Figure 7). This students with a higher rank were more likely to choose STEM fields and students with a lower rank were more likely to graduate from high school.

### 6.4 The effect of reform on schools' composition of students

We used the main model (equation 1) rather than the outcome as the dependent variable for the different characteristics of the students. The treatment estimates show whether the share of students with different characteristics changed due to the reform. In all cases, the treatment coefficient was not significant (Figure 8). That is, the composition of the students in the treated schools did not change after the reform.

## 7 Robustness Check

In this section, we present the results of two sensitivity checks. First, we analyzed the different matched groups to see how results changed when using a more similar control group to the treatment group. Second, we tested the sensitivity of the results with respect to the study period 1998–2010, when the grading system changed.

### 7.1 Matching

One way to increase the chance that the parallel trend assumption holds is to limit the sample of analysis to more similar schools – i.e., to find the schools in the non-participating group that are more comparable with the participating schools. This section aims to minimize unobservable variable bias due to selection by matching pre-reform characteristics. We used a standard matching strategy to match school-level data from the year before the reform was implemented. The following variable is used for

matching schools: immigration share, the average parental income, the average level of parents' education, the school size, and the municipality size. We matched three control groups: all controls (i.e., matching schools in all municipalities), control A, and control B.

We first checked whether matching improves both balance test (Table B.2) and parallel trend assumptions (Figure A.5). The absolute value of each estimate was higher in most cases than the non-matched sample . In the matched sample, the percentile rank received the significant coefficients (Table 6). That is, in a more similar control group, the effect is more substantial. With the exception of STEM, all other effect sizes increased in this sample.

For sample A – a match group in Panel b (Table 6) – the coefficient for STEM outcomes is significant, but others remain non-significant compared to non-matched group. For sample B, the match group for all coefficients except STEM outcomes gets the higher estimates. That is, based on the strategy used, the estimated effect is positive using the more credible control groups, but the results are sensitive to the choice of the control group.

## 7.2 New grading system

The new grading reform was introduced in Sweden in 1998 when the goal-based systems replaced the norm-based system. In the old system, the scale was 1–5. The new system uses four scales: fail, pass, pass with distinction, and pass with excellence. A critical aspect of the new system was that the level of competition decreased since students were graded based on their own ability rather than relative to their cohort. The grading reform was also one of the main reasons that the timetable experiment was implemented as it was argued that a more flexible timetable is needed to gain more from the goal-based system. By limiting the study period to 1998 and onward, we isolated the reform effect to when the new grading system was implemented.

The results of this check are presented in Table 7. Comparing these results with results in Table 3, the estimates for percentile rank begin to turn significant in this sample. The probability of attending the STEM field increased from 0.0013 to 0.0016. The other two coefficients are similar to the main results.

# 8 Concluding

This paper analyzes the impact of a school-based timetable on several student outcomes by investigating a program that gave some schools in the country the opportunity to allocate time to different topics as they deemed appropriate. The reform was implemented as a pilot project, and municipalities and schools could choose to be part of the treatment. We use the whole population registry data of all students who graduated from grade 9 from 1990 to 2010 and implemented a DiD model. The identifying assumption is that exposure to the school-based timetable is as good as random and restricted by school-fixed and cohort-fixed effects and the included covariates. A set of validation checks supports this assumption.

Although there was no effect on the overall GPA of the students in the targeted schools, there was statistically significant effect on the probability that students choose the STEM field and a negligible impact on the years of education and the probability that students graduate from high school by the age of 25. We also found that more prolonged exposure to the school-based timetable led to higher impacts. The results suggest that timetable reform does not affect students in the short term, but it helps them in long-term.

The paper results also show a heterogeneous effect of the reform on students with different backgrounds. We checked this for two outcomes: the probability that students chose a STEM track and the probability that students finished high school by the age of 25. In terms of the STEM field, females and students with low-income parents gain

more from the reform. For the high school graduation outcome, males and all students with low educated parents benefited from the reform. We also looked at students' ability distribution: more students with high performance chose the STEM track and the probability of graduating from high school increased for students with low performance. This paper also found that students in large schools benefited more from the reform than students from smaller schools. Students from larger schools had the twice the probability to choose a STEM track. Large schools potentially can benefit from this type of reform since they have more diversity in terms of students' abilities and backgrounds. They could use timetable freedom as a tool to respond to pupils' individual needs.

The size of the estimates is negligible for most of the outcomes. However, the reform is cost-efficient since these types of interventions are relatively inexpensive to implement, especially compared with other reforms such as reducing class size and increasing teacher's salaries. For the external validity of the effect of this reform in another context, we should consider that at the time of the reform the Swedish education system was among the most decentralized education systems in the world, so the schools and municipalities were very familiar with these kinds of decisions.

One of the limitations of this study is that we did not know how the schools in the treated groups used the timetable's freedom. In other words, we did not know exactly how many hours they allocated to STEM-related subjects or to individual sessions. Further research is needed to develop a more complete picture.

# References

F. Alm. *Skolämnen och alternativen: Schemat som indikator på vad som händer i skolor utan timplan*. Linköping University Electronic Press, 2003.

B. Bimber. *School Decentralization: Lessons from the Study of Bureaucracy*. Rand Corporation, 1993.

B. Bruns, D. Filmer, and H. A. Patrinos. *Making Schools Work : New Evidence on Accountability Reforms*. Number 2270 in World Bank Publications. The World Bank, January 2011. URL http://ideas.repec.org/b/wbk/wbpubs/2270.html.

E. Elmeroth, L. Eek-Karlsson, R. Olsson, and L.-O. Valve. Tid för målstyrning: utvärdering av kvalitetsarbete i försöksverksamheten med utbildning utan timplan i grundskolan, 2005.

T. Fasih, F. Barrera-Osorio, and H. A. Patrinos. *Decentralized Decision-Making in Schools*. The World Bank, 2009. doi: 10.1596/978-0-8213-7969-1. URL http://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-7969-1.

T. Fuchs and L. Wöbmann. What accounts for international differences in student performance? a re-examination using pisa data. *Empirical Economics*, 32(2-3): 433–464, 2007. ISSN 0377-7332. doi: 10.1007/s00181-006-0087-0. URL http://dx.doi.org/10.1007/s00181-006-0087-0.

E. A. Hanushek, S. Link, and L. Woessmann. Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA. CESifo Working Paper Series 3648, CESifo Group Munich, 2011. URL http://ideas.repec.org/p/ces/ceswps/_3648.html.

L. J. Kirkeboen, E. Leuven, and M. Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

L. Lundahl. Sweden: decentralization, deregulation, quasi-markets-and then what? *Journal of education policy*, 17(6):687–697, 2002.

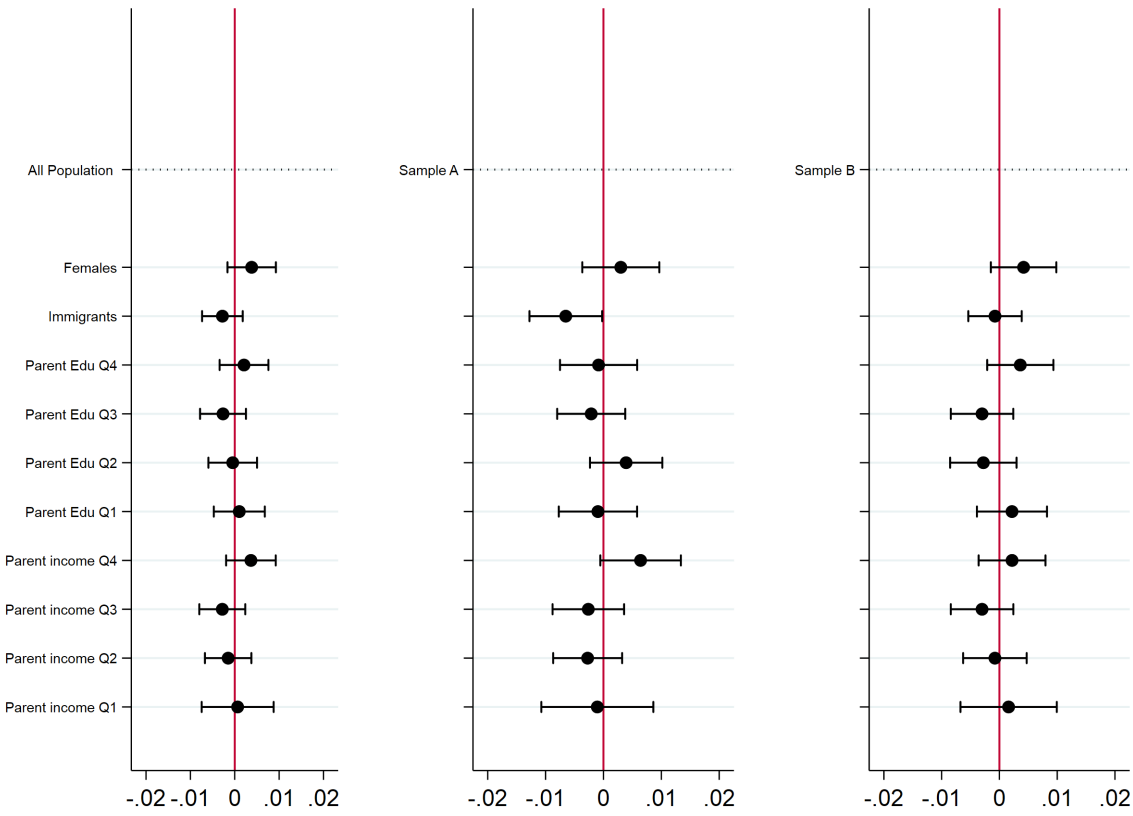
L. Lundahl. A matter of self-governance and control the reconstruction of swedish education policy: 1980-2003. *European education*, 37(1):10–25, 2005.

A. sa Ahlin and E. Mörk. Effects of decentralization on school resources. *Economics of Education Review*, 27(3):276 – 284, 2008. ISSN 0272-7757. doi: http://dx.doi.org/10.1016/j.econedurev.2007.01.002. URL `http://www.sciencedirect.com/science/article/pii/S0272775707000301`.

E. Wallberg, R. Eliasson, L. Fernstedt, I. Gullstam, M. Lundstrom, M. Molander, and M. W. å. Utan timplan. Technical report, Slutbetä nkande från Timplanedelegationen, 2005.

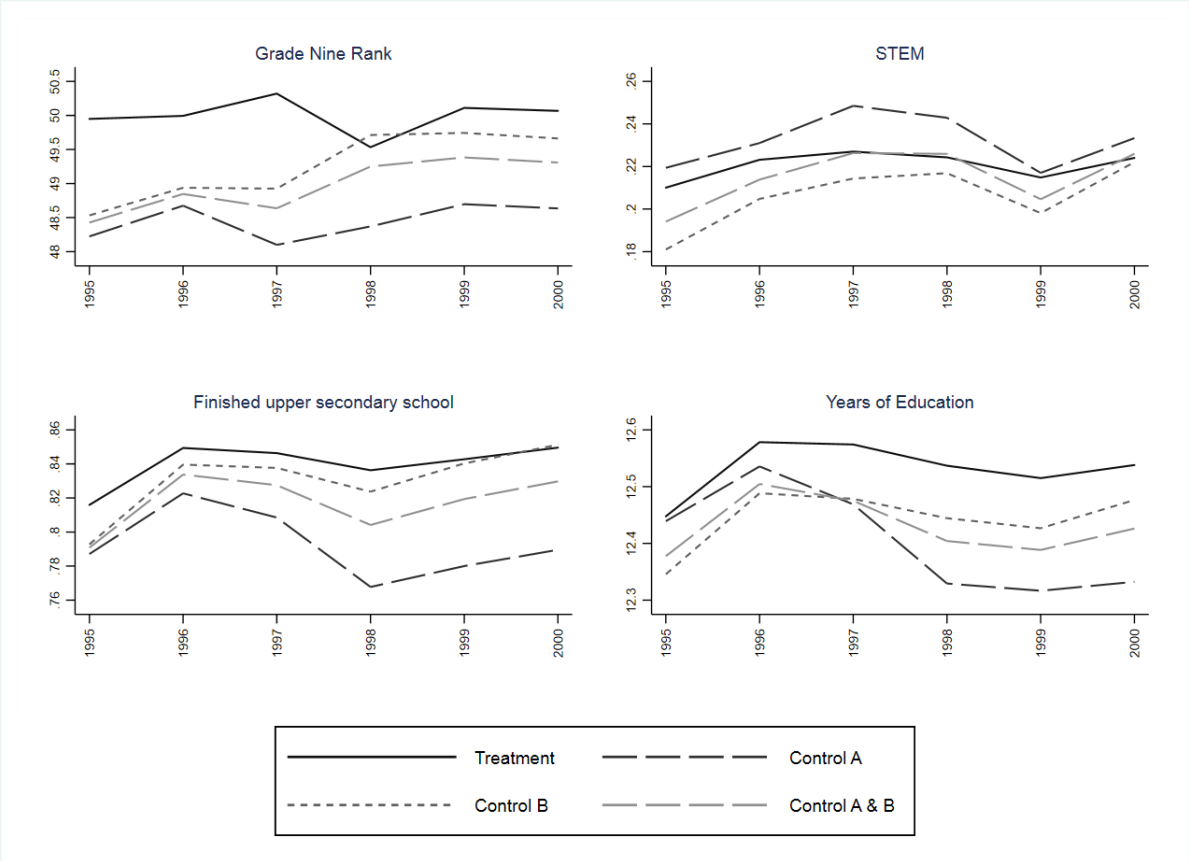
J. Zabojnik. Centralized and decentralized decision making in organizations. *Journal of Labor Economics*, 20(1):1–22, 2002.

Figure (1) Balance of covariates for sample A, sample B, and all populations.




Note: this figure shows the estimates and 95% confidence interval of equation 1, replacing $Y_{ist}$ with the predetermined characteristics before the policy period (1990–1999). This means that we regressed each covariant on the pre-treatment placebo dummy (1995–1999) and used only the years before the treatment period (1990–1999).

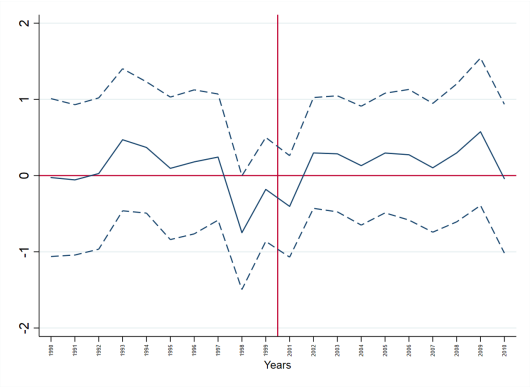
Figure (2) Trends in outcomes, pre-reform period.



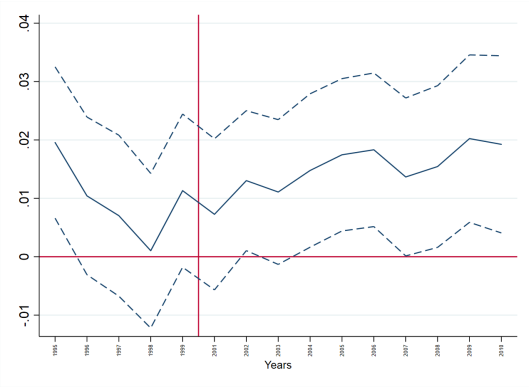

Notes: The figure shows the pre-reform trend (1990–1999) of grade 9 rank, probability of choosing STEM, years of schooling, and the likelihood of high school graduation. The different lines differentiate between schools in treatment, control A, control B, and all control schools.

Figure (3) Event study graph of the effect of school timetable on different outcomes.

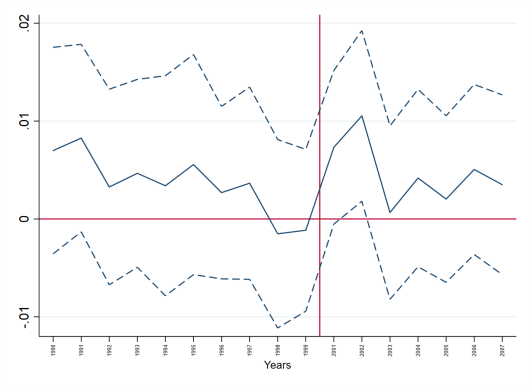
(a) Grade Nine Rank



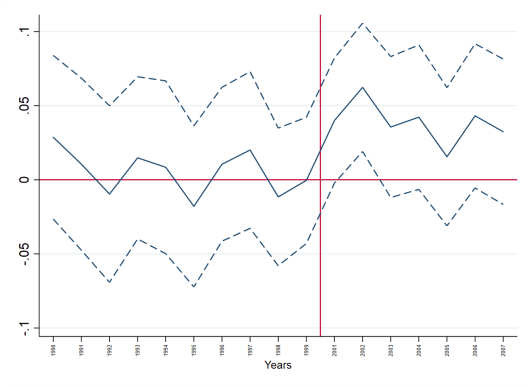

(b) STEM




(c) Finished upper secondary school



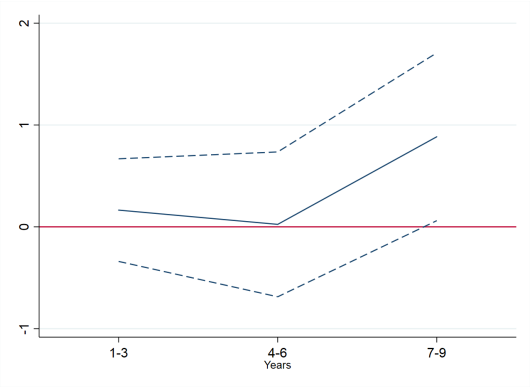

(d) Years of Schooling age 25



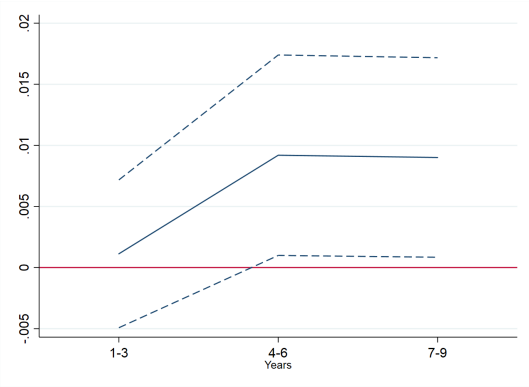

Notes: The figure shows the pre-reform trend (1990–1999) of grade 9 rank, probability of choosing STEM, years of schooling, and the likelihood of high school graduation. The different lines differentiate between schools in treatment, control A, control B, and all control schools.

Figure (4) Effect of years of exposure to school-based timetable on different outcomes

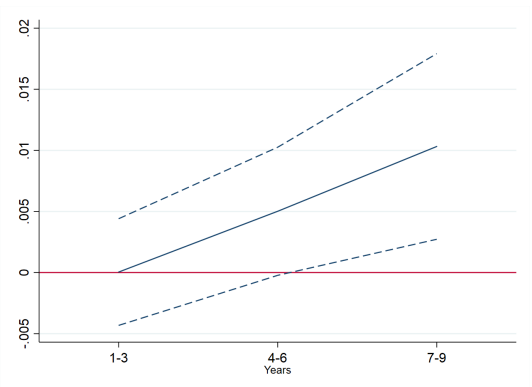
(a) Grade Nine Rank



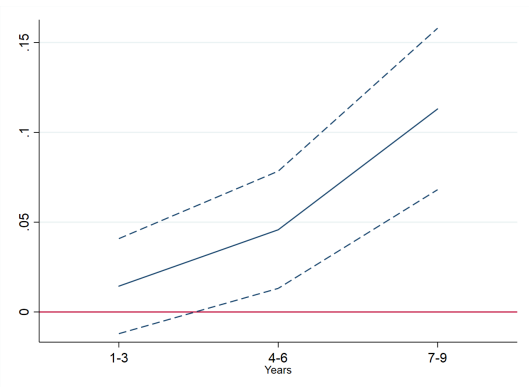

(b) STEM




(c) Finished upper secondary school




(d) Years of Schooling age 25

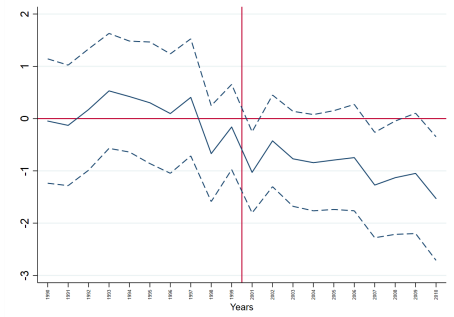



Note: The figure shows estimates and 95% confidence intervals from a regression on the effect of years of exposure to the school-based timetable on percentile rank in grade 9, the probability enrolling in a STEM track, probability of high school graduation, and years of education. The x-axis shows the level of exposure to the treatment: 1–3 years, 4–6 years, and 6–9 years. Control variables include year, school-fixed effect, immigration status, parent income (4 levels), and parent education (4 levels). Standard errors are clustered at the school level.

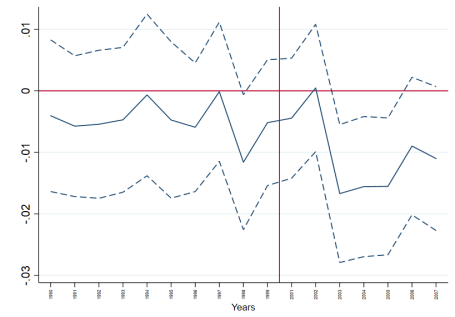
Figure (5) Event study graph of the effect of school timetable on different outcomes, sample A and B

Sample A

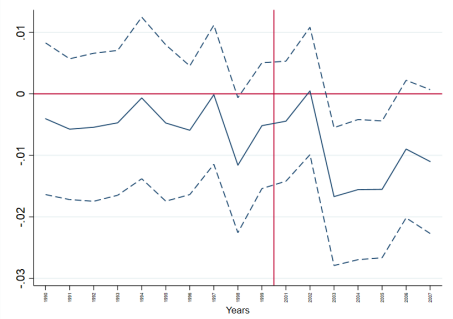
(a) Grade Nine Rank



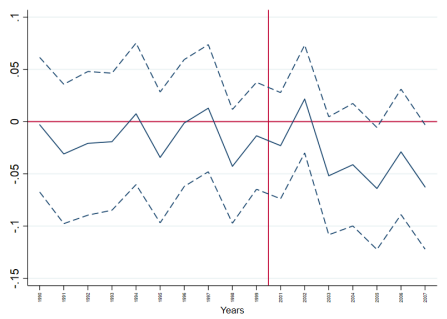

(b) STEM




(c) Finished upper secondary school

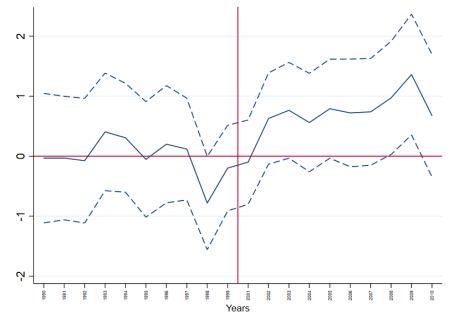



(d) Years of Schooling age 25

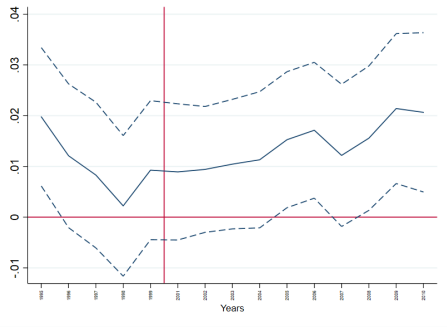



Sample B

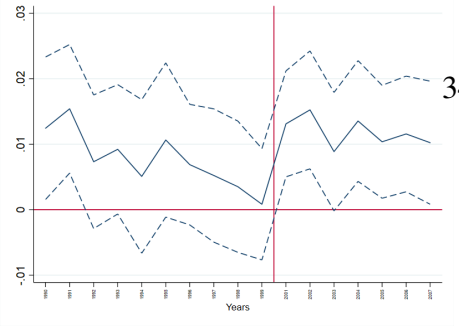
(e) Grade Nine Rank



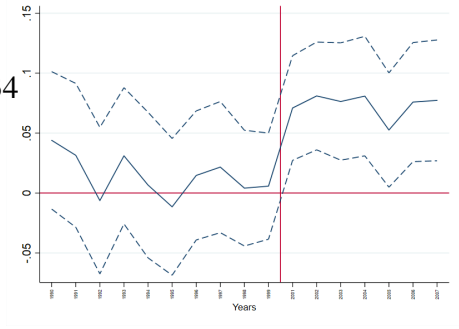

(f) STEM




(g) Finished upper secondary school

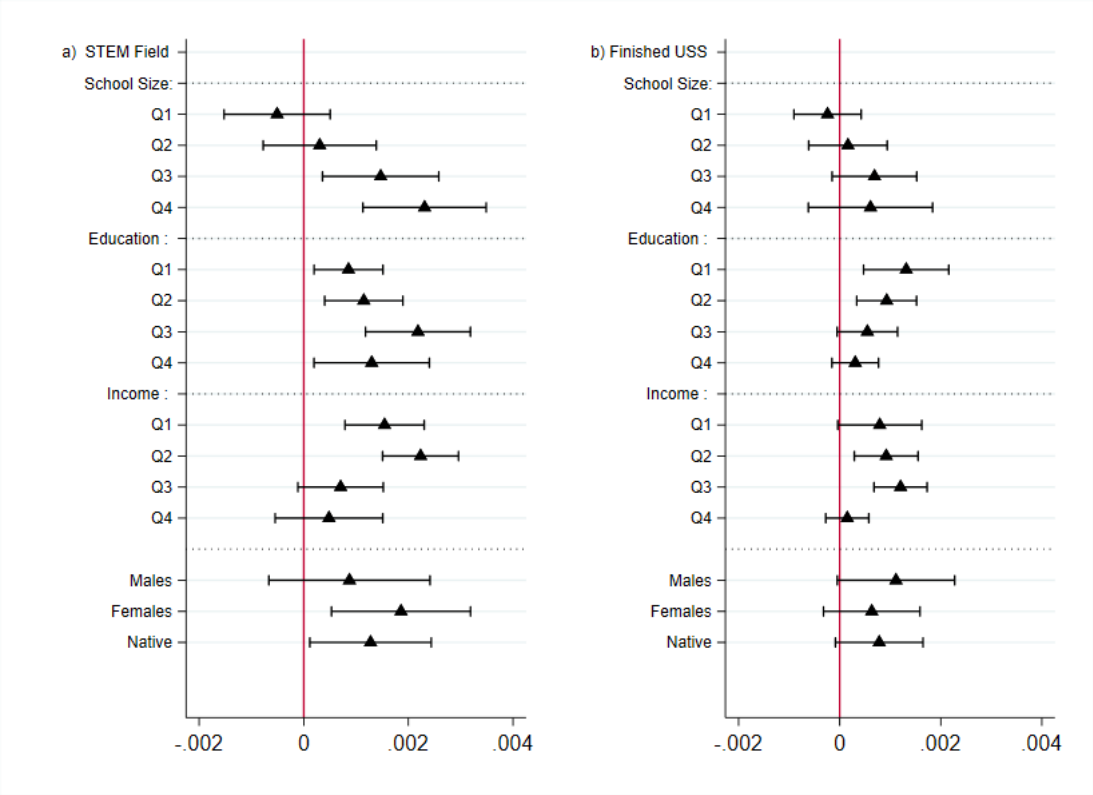



(h) Years of Schooling age 25



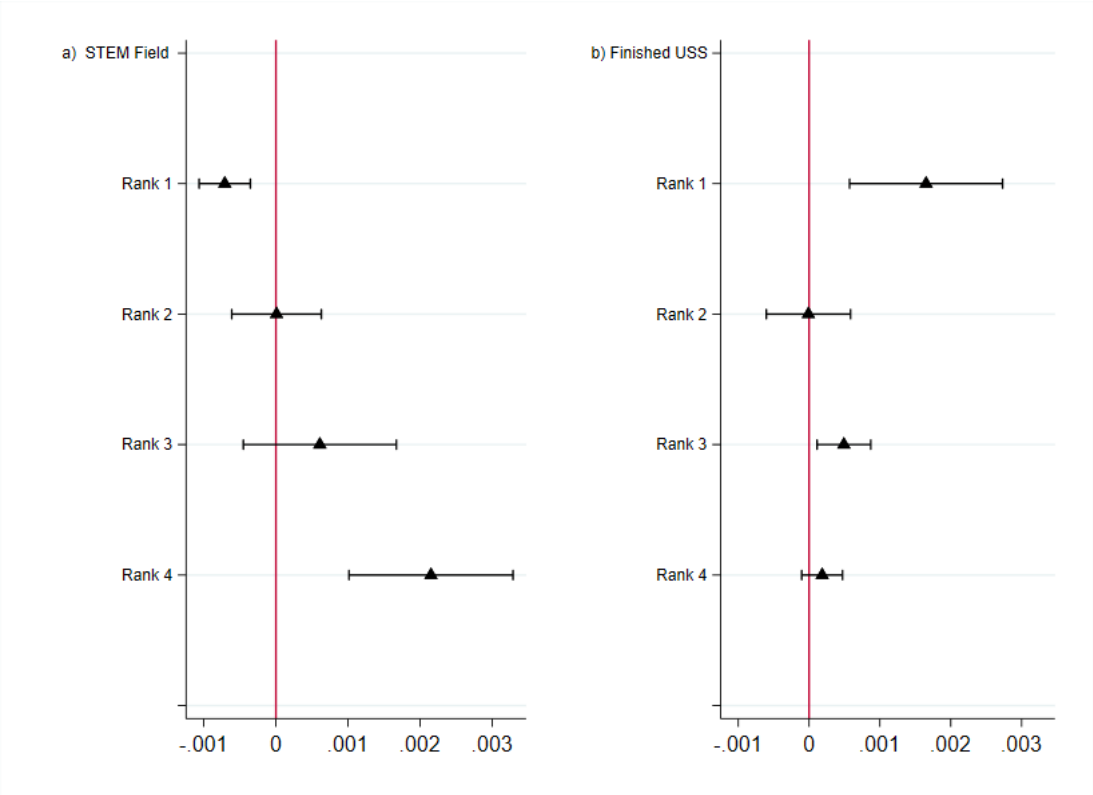

Note: The figure plots estimates and 95% confidence intervals from an event study regression for sample A (Panel a-d ) and sample B (Panel e-h).

Figure (6) Heterogeneity analyses, linear effect of school-based timetable on choosing STEM track and high school graduation



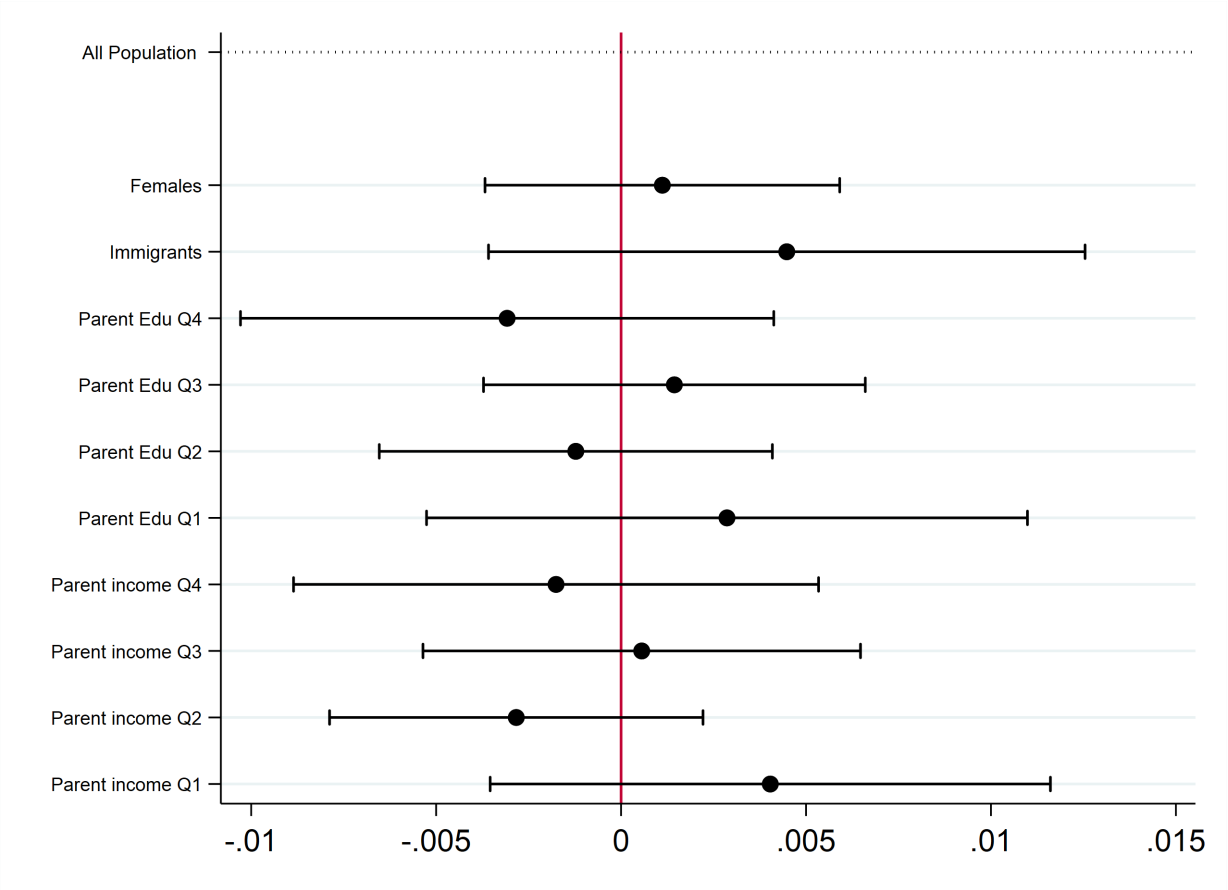

Note:This figure shows DiD estimates and 95% confident interval of the effect of school-based timetable exposure on choosing STEM track in Panel a and low education (less than 11 years) at the age of 25 on Panel b. Each dot represents the effect of a sub-sample of the population. For the school size, the education of father and mother were added and ranked. The low-educated parent is the first quantile of family educational rank, and the highly educated is the fourth quantile of class. Income quantile is calculated as the sum of parent income. Parents' income and education were used when students were in grade 9. In all models, school- and cohort-fixed effects were added.

Figure (7) Heterogeneity analyses: linear effect of school-based timetable on choosing STEM field (Panel a) and finished upper secondary school at the age of 25 by student quartile rank




Note:This figure shows DiD estimates and 95% confident interval of the effect of school-based timetable exposure on choosing STEM track (Panel a) and the probability of graduating from high school at the age of 25 (Panel b). Each dot represents the effect of a sub-sample of the population. Each rank represents the student's national quantile rank. In all models, school- and cohort-fixed effects were added.

Figure (8) The effect of the reform on predetermined characteristics




Note: This figure shows the estimates and 95% confidence interval of equation 1, replacing $Y_{ist}$ with the predetermined characteristics. This means that we regressed each covariant on the treatment dummy. Control variables include cohort- and school-fixed effects. Standard errors are clustered at the school level.

Table (1) Participating and non-participating municipality characteristics

| | Participant | Non-Participant |
|---|---|---|
| Number of Students | 683.8 (841.3) | 243.9 (357.5) |
| Female (%) | 48.37 (2.461) | 48.62 (3.941) |
| Immigrants (%) | 13.43 (7.656) | 9.297 (5.861) |
| Born Out of Sweden (%) | 9.433 (4.503) | 6.358 (3.632) |
| Average Income | 4.436 (0.624) | 4.181 (0.720) |
| Median Income | 3.976 (0.261) | 3.870 (0.259) |
| Employment | 76.82 (3.986) | 76.28 (4.181) |
| Inhabitants Per Square Kilometer | 342 (788.6) | 59 (144.7) |
| Total Cost per student | 56288.4 (7688.9) | 54740.6 (4848.4) |
| Right-wing share (%) | 0.41 (0.49) | 0.33 (0.47) |
| Number of municipalities | 70 | 219 |

Table (2) Descriptive statistics for treated and control group one year before the treatment (1999)

| | Treatment | Control All | Δ All | Control A | Δ A | Control B | Δ B |
|---|---|---|---|---|---|---|---|
| Immigrant(%) | 3.38 | 3.73 | -0.03 | 5.89 | -0.20 | 2.59 | 0.09 |
| | (8.20) | (11.80) | | (15.47) | | (9.13) | |
| Females(%) | 48.66 | 47.38 | 0.12 | 47.19 | 0.12 | 47.48 | 0.12 |
| | (5.84) | (13.98) | | (15.81) | | (12.91) | |
| Fathers YoE | 11.49 | 11.47 | 0.03 | 11.75 | -0.24 | 11.31 | 0.19 |
| | (0.91) | (1.11) | | (1.23) | | (1.01) | |
| Mother YoE | 11.53 | 11.51 | 0.02 | 11.67 | -0.13 | 11.43 | 0.12 |
| | (0.81) | (1.00) | | (1.24) | | (0.83) | |
| Father income (log) | 7.17 | 7.12 | 0.12 | 7.13 | 0.09 | 7.12 | 0.15 |
| | (0.37) | (0.47) | | (0.56) | | (0.41) | |
| Mother income (log) | 6.86 | 6.81 | 0.13 | 6.86 | 0.01 | 6.78 | 0.21 |
| | (0.33) | (0.43) | | (0.47) | | (0.41) | |
| Municipality size (1–10) | 3.89 | 4.15 | -0.12 | 3.15 | 0.36 | 4.69 | -0.36 |
| | (2.09) | (2.39) | | (1.98) | | (2.42) | |
| School size | 100.52 | 77.44 | 0.51 | 75.61 | 0.54 | 78.41 | 0.50 |
| | (41.64) | (48.18) | | (50.00) | | (47.20) | |
| GPA (0–320) | 200.67 | 198.45 | 0.10 | 195.62 | 0.17 | 199.93 | 0.04 |
| | (13.98) | (29.89) | | (38.57) | | (24.01) | |
| Percentile Rank (1–100) | 50.11 | 49.38 | 0.07 | 48.70 | 0.12 | 49.74 | 0.04 |
| | (6.81) | (12.32) | | (15.18) | | (10.51) | |
| Choosing STEM (%) | 21.49 | 20.46 | 0.09 | 21.70 | -0.02 | 19.81 | 0.16 |
| | (8.66) | (12.61) | | (14.12) | | (11.69) | |
| Years of education | 12.51 | 12.39 | 0.19 | 12.32 | 0.26 | 12.43 | 0.15 |
| | (0.43) | (0.84) | | (1.00) | | (0.73) | |
| Low education (%) | 16.98 | 19.53 | -0.18 | 21.86 | -0.30 | 18.30 | -0.11 |
| | (8.50) | (17.85) | | (21.35) | | (15.55) | |
| Finished USS(%) | 84.28 | 81.94 | 0.17 | 78.00 | 0.38 | 84.03 | 0.02 |
| | (7.87) | (17.66) | | (22.09) | | (14.36) | |
| Observations | 183 | 1009 | | 350 | 1 | 659 | |

Note: This table shows mean value and standard deviation for the treated group, All control, control A, and control B. The normalized difference is indicated next to each control group. In each line, the smallest normalized difference is illustrated with the bold font.

Table (3) The Effect of school-based timetable on different outcomes

| | Percentile Rank | STEM | Finished USS | Years of Education |
|---|---|---|---|---|
| Reform 0-9 | 0.091 | 0.0013** | 0.0012** | 0.012*** |
| | (0.059) | (0.00059) | (0.00057) | (0.0034) |
| Placebo | 0.32 | | 0.0045 | 0.018 |
| | (0.28) | | (0.0028) | (0.016) |
| Observations | 2,212,332 | 1,544,042 | 1,702,762 | 1,659,152 |
| Mean of outcome | 50.7 | 0.21 | 0.86 | 12.6 |
| Effect size (% of mean*9) | 1.62 | 5.63 | 1.21 | 0.89 |

Note:This table shows DiD estimates and 95% confidence interval of the effect of school-based timetable exposure on grade 9 percentile rank ( column 1), the probability that students enroll in a STEM track (column 2), the probability of graduating from upper secondary school (column 3), and the year of education (column 4). Placebo is a dummy for the years 1990-1994. The variable STEM is available from 1995. Control variables include cohort- and school-fixed effects, immigration status, parent income (4 levels), and parent education (4 levels). Standard errors are clustered at the school level. Robust standard errors in parentheses. *** P¡0.01, ** P¡0.05, * P¡0.1.

Table (4) The linear effect of the school-based timetable on Mathematics, English and Swedish grades.

| | (1) | (2) |
|---|---|---|
| | Pass | High |
| Mathematics | 0.00056 | 0.0023*** |
| | (0.00044) | (0.00086) |
| Mean of outcome | 0.92 | 0.40 |
| Effect size (% of mean *9) | 0.55 | 5.12 |
| English | 0.0012*** | 0.0022** |
| | (0.00032) | (0.00096) |
| Mean of outcome | 0.93 | 0.54 |
| Effect size (% of mean*9) | 1.19 | 3.56 |
| Swedish | -0.000053 | 0.000089 |
| | (0.00017) | (0.00016) |
| Mean of outcome | 0.039 | 0.014 |
| Effect size (% of mean*9 ) | -1.23 | 5.62 |
| Observations | 1,501,277 | 1,501,277 |

Note: The estimates in each panel show the effect of the treatment on the probability of passing the course and getting a high grade in the corresponding subjects. The period of the study is 1998–2012, i.e., during the new grading system. Control variables include cohort- and school-fixed effects, immigration status, parent income (4 levels), and parent education (4 levels). Standard errors are clustered at the school level. Robust standard errors in parentheses. *** P¡0.01, ** P¡0.05, * P¡0.1.

Table (5) The Effect of school-based timetable on different outcomes in sample A and B

| | Percentile Rank | STEM | Finished USS | Years of Education |
|---|---|---|---|---|
| | | *a. Sample A* | | |
| Reform 0-9 | -0.091 | 0.0011 | -0.00014 | 0.0031 |
| | (0.071) | (0.00072) | (0.00072) | (0.0043) |
| Placebo | 0.54* | | 0.0030 | 0.018 |
| | (0.32) | | (0.0033) | (0.019) |
| Observations | 1,023,726 | 710,666 | 778,990 | 758,239 |
| Mean of outcome | 51.3 | 0.22 | 0.86 | 12.6 |
| Effect size (% of mean) | -1.60 | 4.46 | -0.15 | 0.22 |
| | | *b. Sample B* | | |
| Reform 0-9 | 0.19*** | 0.0013** | 0.0020*** | 0.018*** |
| | (0.059) | (0.00061) | (0.00058) | (0.0034) |
| Placebo | 0.23 | | 0.0056* | 0.020 |
| | (0.30) | | (0.0029) | (0.016) |
| Observations | 1,565,562 | 1,113,086 | 1,213,658 | 1,183,295 |
| Mean of outcome | 50.2 | 0.21 | 0.86 | 12.5 |
| Effect size (% of mean) | 3.45 | 5.88 | 2.05 | 1.31 |

This table shows DiD estimates and 95% confidence interval of the effect of school-based timetable exposure on grade 9 percentile rank ( column 1), the probability that students enroll in a STEM track (column 2), the probability of graduating from upper secondary school (column 3), and the year of education (column 4). Panel a shows the estimates for Sample A, and Panel b shows that Sample B. Placebo is a dummy for the years 1990–1994. The variable STEM is available from 1995. Control variables include cohort and school fixed effects, immigration status, parent income (4 levels), and parent education (4 levels). Standard errors are clustered at the school level. Robust standard errors in parentheses. *** P¡0.01, ** P¡0.05, * P¡0.1.

Table (6) The Effect of school-based timetable on different outcomes in sample A and B

| | Percentile Rank | STEM | Finished USS | Years of Education |
|---|---|---|---|---|
| | | *a.Match All* | | |
| Reform 0-9 | 0.13** | 0.0014** | 0.0013** | 0.013*** |
| | (0.061) | (0.00063) | (0.00061) | (0.0037) |
| Placebo | 0.69** | | 0.0042 | 0.022 |
| | (0.31) | | (0.0030) | (0.017) |
| Observations | 1,232,497 | 901,786 | 975,026 | 950,412 |
| Mean of outcome | 50.6 | 0.21 | 0.86 | 12.6 |
| Effect size (% of mean) | 2.27 | 5.73 | 1.37 | 0.95 |
| | | *b.Match A* | | |
| Reform 0-9 | -0.033 | 0.0012 | 0.00029 | 0.0073* |
| | (0.072) | (0.00075) | (0.00074) | (0.0044) |
| Placebo | 0.57* | | 0.0046 | 0.026 |
| | (0.34) | | (0.0036) | (0.021) |
| Observations | 770,539 | 557,730 | 606,519 | 590,928 |
| Mean of outcome | 51 | 0.22 | 0.86 | 12.6 |
| Effect size (% of mean) | -0.58 | 4.82 | 0.31 | 0.52 |
| | | *c.Match B* | | |
| Reform 0-9 | 0.20*** | 0.0012* | 0.0023*** | 0.019*** |
| | (0.061) | (0.00063) | (0.00059) | (0.0036) |
| Placebo | 0.24 | | 0.0066** | 0.023 |
| | (0.32) | | (0.0030) | (0.017) |
| Observations | 1,067,143 | 788,088 | 841,627 | 820,584 |
| Mean of outcome | 50.2 | 0.21 | 0.86 | 12.6 |
| Effect size (% of mean) | 3.63 | 5.24 | 2.37 | 1.33 |

This table shows DiD estimates and 95% confidence interval of the effect of school-based timetable exposure on grade 9 percentile rank ( column 1), the probability that students enroll in a STEM track (column 2), the probability of graduating from upper secondary school (column 3), and the year of education (column 4). Panel a shows the estimates for Sample All matched, Panel b shows that Sample A match and Panel c Sample B match. Placebo is a dummy for the

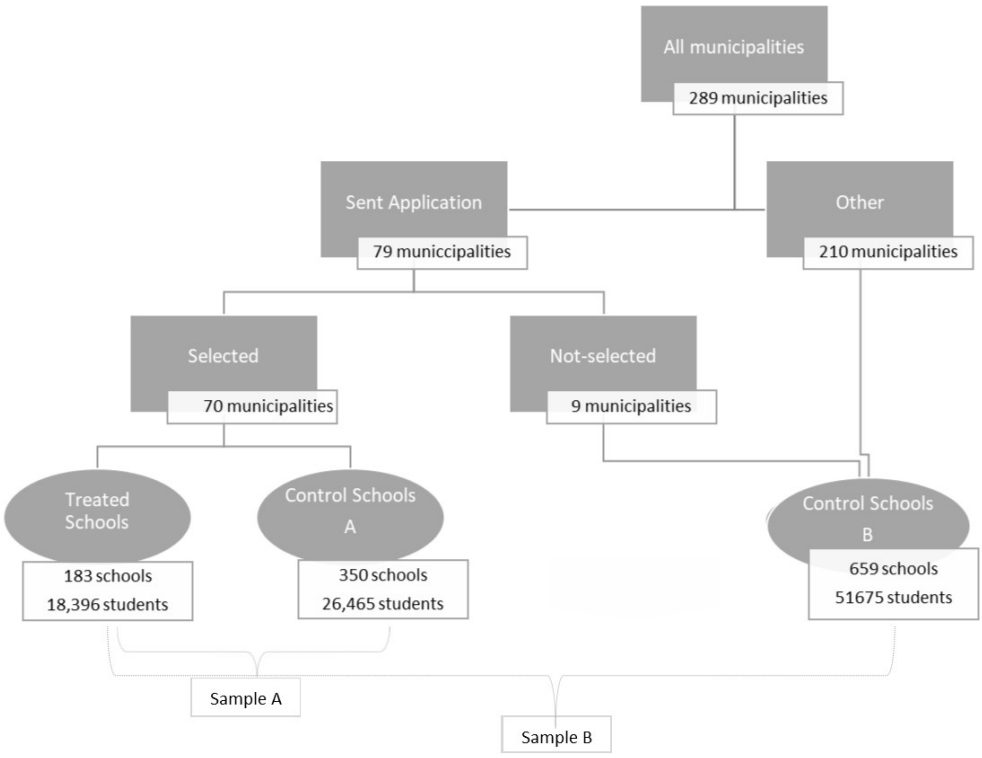
Table (7) The Effect of school-based timetable on different outcomes for 1998–2010

| | Percentile Rank | STEM | Finished USS | Years of Education |
|---|---|---|---|---|
| Reform 0-9 | 0.12** | 0.0016** | 0.0012* | 0.011*** |
| | (0.057) | (0.00062) | (0.00065) | (0.0034) |
| Observations | 1,501,277 | 1,292,444 | 991,707 | 965,689 |
| Mean of outcome | 50.9 | 0.21 | 0.86 | 12.6 |
| Effect size (% of mean) | 2.07 | 6.61 | 1.28 | 0.77 |

# Appendices

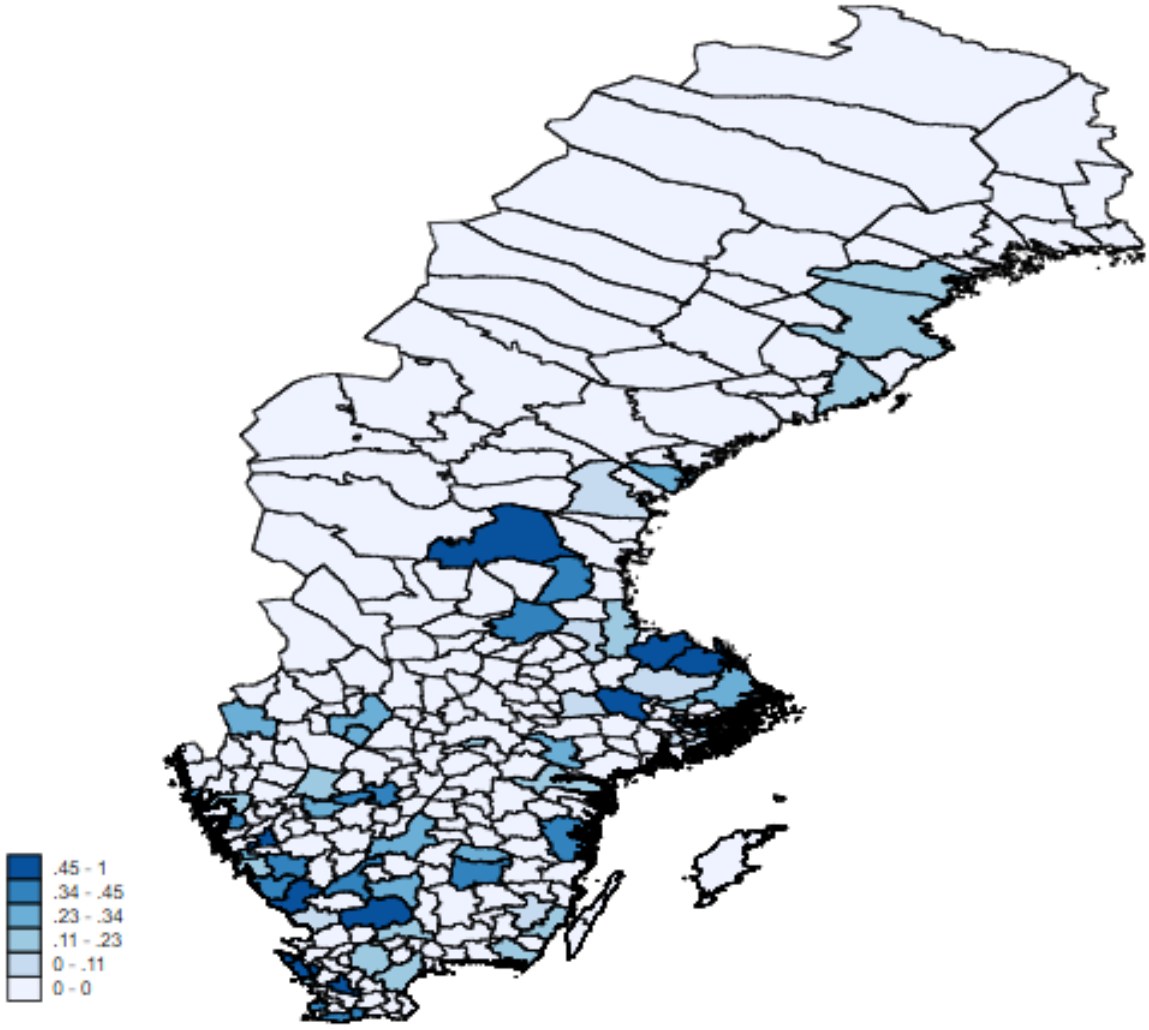
## A Figures

Figure (A.1) The experiment implementation process




Note:This figure shows how the reform was implemented in 2000. First, the Ministry of Education invited all municipalities to apply for participating in the experiment. In the second stage, selected municipalities were allowed to decide which of their schools would participate in the experiment.

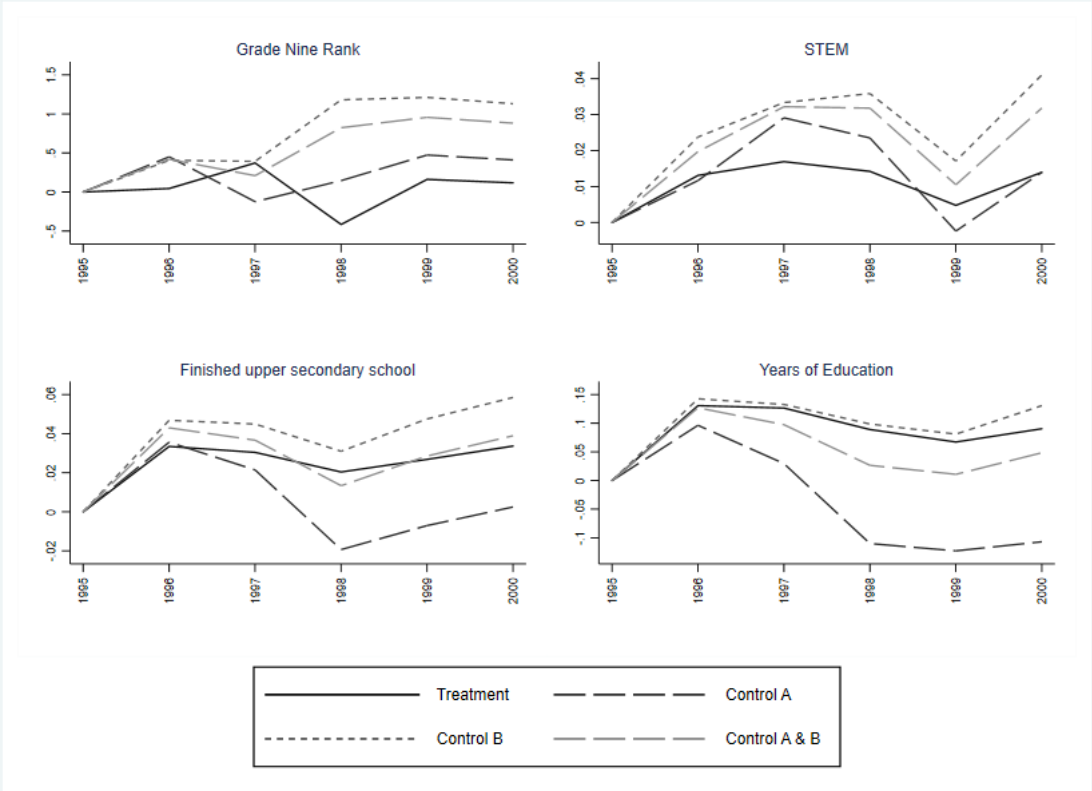
Figure (A.2) . The distribution of the affected students across municipalities in Sweden




Note: This figure shows the share of students affected by the reform in 2000. White municipalities are the control group.

s

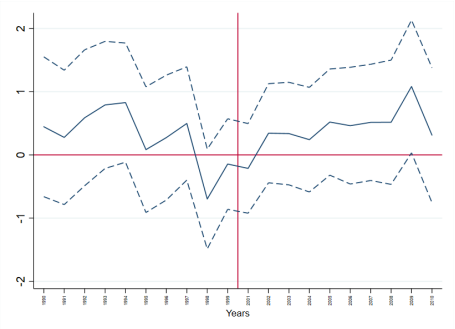
Figure (A.3) Pre-reform trend



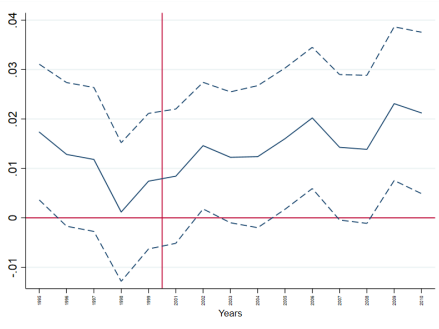

Notes: The figure shows the pre-reform trend (1990–1999) of grade 9 rank, probability of choosing STEM, years of schooling, and the likelihood of high school graduation. The different lines differentiate between schools in treatment, control A, control B, and all control schools.

Figure (A.4) Event study graph of the effect of school timetable on different outcomes, matched

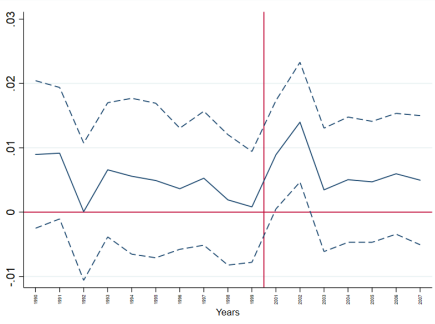
(a) Grade Nine Rank



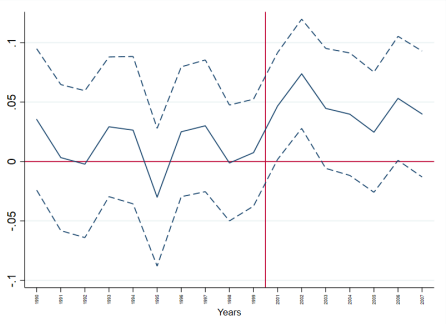

(b) STEM




(c) Finished upper secondary school




(d) Years of Schooling age 25

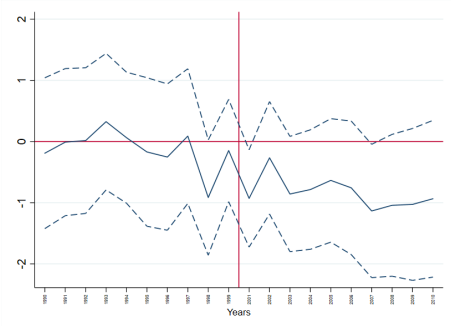



Note: The figure plots estimates and 95% confidence intervals from an event study regression for matched group

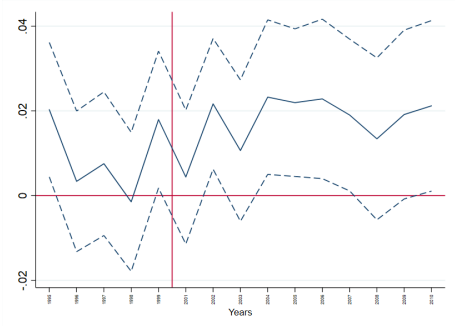
Figure (A.5) Event study graph of the effect of school timetable on different outcomes, matched sample A and B

Sample A, Matched

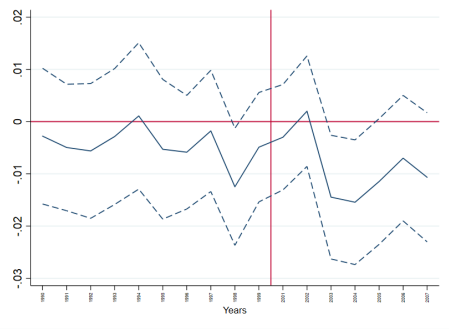
(a) Grade Nine Rank



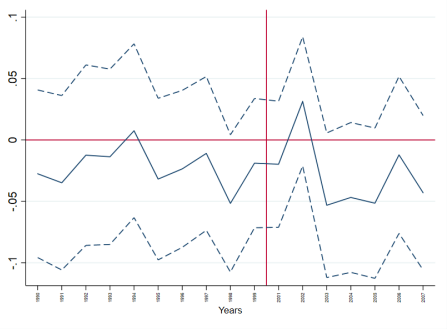

(b) STEM




(c) Finished upper secondary school

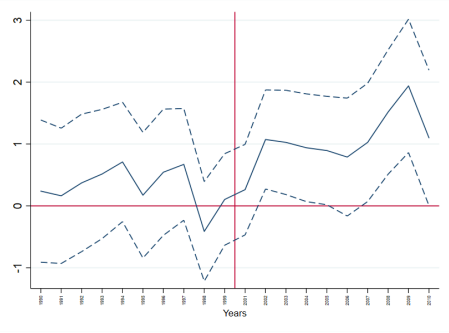



(d) Years of Schooling age 25

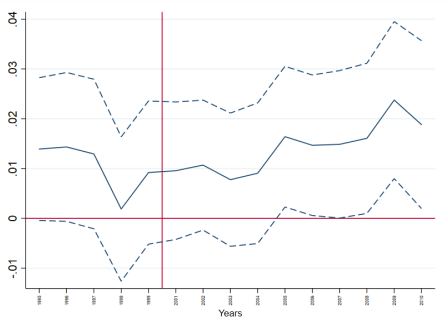



Sample B, Matched

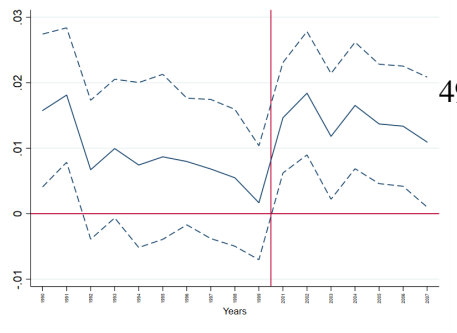
(e) Grade Nine Rank



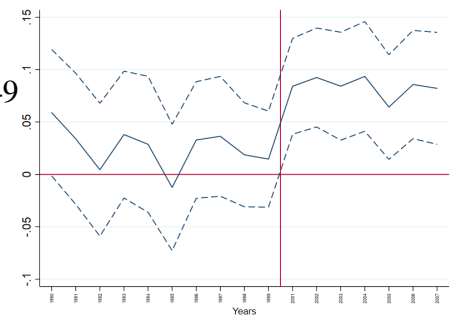

(f) STEM




(g) Finished upper secondary school




(h) Years of Schooling age 25






Note: The figure plots estimates and 95% confidence intervals from an event study regression for matched sample A (Panel a-d ) and matched sample B (Panel e-h).

# B Tables

Table (B.1) Participating and non-participating municipality characteristics

| Year | Control Group | | | Level of Treatment | | | Total |
|---|---|---|---|---|---|---|---|
| | All | A | B | 1–3 years | 4–6 years | 7–9 years | |
| 2000 | 79,967 | 26,876 | 53,091 | 19,551 | 0 | 0 | 99,518 |
| 2001 | 83,660 | 28,126 | 55,534 | 20,355 | 0 | 0 | 104,015 |
| 2002 | 85,560 | 28,912 | 56,648 | 21,050 | 0 | 0 | 106,610 |
| 2003 | 87,817 | 29,517 | 58,300 | 7,305 | 13,684 | 0 | 108,806 |
| 2004 | 93,811 | 31,915 | 61,896 | 7,943 | 13,972 | 0 | 115,726 |
| 2005 | 96,822 | 33,605 | 63,217 | 7,782 | 14,650 | 0 | 119,254 |
| 2006 | 102,843 | 36,098 | 66,745 | 8,199 | 4,626 | 10,305 | 125,973 |
| 2007 | 102,430 | 36,622 | 65,808 | 8,122 | 4,614 | 9,931 | 125,097 |
| 2008 | 100,996 | 36,697 | 64,299 | 7,590 | 4,538 | 9,701 | 122,825 |
| 2009 | 98,185 | 36,464 | 61,721 | 7,063 | 4,214 | 9,046 | 118,508 |
| 2010 | 94,753 | 35,703 | 59,050 | 6,795 | 3,914 | 8,565 | 114,027 |
| 2011 | 88,667 | 34,015 | 54,652 | 5,913 | 3,670 | 8,225 | 106,475 |
| 2012 | 83,456 | 32,574 | 50,882 | 5,539 | 3,360 | 7,339 | 99,694 |
| Total | 1,198,967 | 427,124 | 771,843 | 133,207 | 71,242 | 63,112 | 1,466,528 |

Table (B.2) Descriptive statistics for treated and control group one year before the treatment (1999)

| | Treatment | Control All | Δ All | Control A | Δ A | Control B | Δ B |
|---|---|---|---|---|---|---|---|
| Immigrant(%) | 3.38 | 3.51 | -0.01 | 4.15 | -0.08 | 2.68 | 0.09 |
| | (8.20) | (10.40) | | (10.08) | | (8.25) | |
| Females(%) | 48.66 | 47.84 | 0.11 | 47.72 | 0.09 | 48.08 | 0.07 |
| | (5.84) | (9.14) | | (12.97) | | (9.17) | |
| Fathers Years of education | 11.49 | 11.47 | 0.02 | 11.63 | -0.14 | 11.40 | 0.10 |
| | (0.91) | (1.00) | | (1.03) | | (0.97) | |
| Mother Years of education | 11.53 | 11.50 | 0.04 | 11.60 | -0.07 | 11.49 | 0.05 |
| | (0.81) | (0.89) | | (1.04) | | (0.78) | |
| Father income (log) | 7.17 | 7.18 | -0.01 | 7.16 | 0.04 | 7.18 | -0.00 |
| | (0.37) | (0.36) | | (0.49) | | (0.33) | |
| Mother income (log) | 6.86 | 6.85 | 0.04 | 6.84 | 0.04 | 6.85 | 0.03 |
| | (0.33) | (0.37) | | (0.45) | | (0.29) | |
| Municipality size (1–10) | 3.89 | 3.95 | -0.03 | 3.32 | 0.28 | 4.45 | -0.25 |
| | (2.09) | (2.35) | | (1.99) | | (2.36) | |
| School size | 100.52 | 92.60 | 0.19 | 86.68 | 0.31 | 90.83 | 0.23 |
| | (41.64) | (42.96) | | (47.74) | | (43.60) | |
| GPA (0–320) | 200.67 | 200.69 | -0.00 | 196.51 | 0.15 | 201.50 | -0.05 |
| | (13.98) | (23.98) | | (35.69) | | (20.51) | |
| Percentile Rank (1–100) | 50.11 | 50.29 | -0.02 | 49.02 | 0.10 | 50.46 | -0.04 |
| | (6.81) | (10.21) | | (13.63) | | (8.82) | |
| Choosing STEM (%) | 21.49 | 21.33 | 0.02 | 21.52 | -0.00 | 20.71 | 0.08 |
| | (8.66) | (10.10) | | (12.76) | | (9.67) | |
| Years of education | 12.51 | 12.49 | 0.04 | 12.41 | 0.14 | 12.54 | -0.05 |
| | (0.43) | (0.64) | | (0.88) | | (0.52) | |
| Low education (%) | 16.98 | 17.43 | -0.04 | 19.75 | -0.20 | 16.13 | 0.09 |
| | (8.50) | (12.84) | | (17.80) | | (10.32) | |
| Finished Upper Secondarly (%) | 84.28 | 83.95 | 0.03 | 80.69 | 0.25 | 85.91 | -0.19 |
| | (7.87) | (13.01) | | (18.51) | | (9.18) | |
| Observations | 183 | 539 | | 273 | | 435 | |

Note: This table shows descriptive statistics for the treated group, All control, control A, and control B after matching. Besides each control group, the normalized difference is calculated. In each line, the minor normalized difference is illustrated with the bold font.