CESifo
AREA
CONFERENCES 2021

Economics of Education
Munich, 3–4 September 2021

Do funds for more teachers improve student performance?

Nicolai T. Borgen, Lars J. Kirkebøen, Andreas Kotsadam, and Oddbjørn Raaum

CESifo

Do funds for more teachers improve student performance?

Nicolai T. Borgen, Lars J. Kirkebøen, Andreas Kotsadam and Oddbjørn Raaum

August, 2021

Abstract

We investigate the effects of a large Norwegian reform that provided extra four-year funding of 600 teachers (1.5 billion NOK in 2012, or 258 million USD) to 166 lower secondary schools. The reform was targeted to schools with a relatively high student-teacher ratio and low average grades. We exploit these two margins in a regression discontinuity setup and find that the reform reduced the average class size by around 10 percent (from 21 to 19). Yet, students' test scores did not improve, and we can reject even a small effect size. In addition, we find no effects of the reform on other academic outcomes such as grades and high school progression. We do find effects on the school environment from the students´ perspective, but evidently not large enough to induce any meaningful effects on academic outcomes.

1 Introduction

There is a long-standing debate as to whether increasing financial resources to schools improves student outcomes or not. Coleman (1968) famously concluded that US school heterogeneity did not explain differences in achievement once family background is controlled for. In an early review of both the US and the international literature, Hanushek (2003) argues that resources do not affect performance. Based on more recent findings, Jackson (2020) and Jackson and Mackevicius (2021) reach the opposite conclusion. Jackson (2020) even concludes about the US literature that "By and large, the question of whether money matters is essentially settled (p. 15)." While the more recent studies use more credible identification strategies, it is not clear whether US results generalize to other places, nor is there to our knowledge any evidence on whether public policies to fund hiring of more teachers will improve learning.[1]

A taxpayer would like to know what will happen to students if the politicians took a part of her contribution to increase funding that enabled schools to hire more teachers. Evidence on this exact question can be provided using data from a Norwegian central government intervention in 2012 with an extra four-year funding of 600 teachers per year (1.5 billion NOK in 2012, or 258 million USD). The extra teachers were targeted at 166 lower secondary schools (grades 8 to 10, ages 14 to 16) with a student-teacher ratio above 20 and grade point average below the national mean in spring 2012. The two sharp margins of the eligibility criteria enable us to credibly identify the effects of more resources using a regression discontinuity framework. We find that the average class size is reduced by two (around 10 percent, from 21 to 19), but no effect on 9th-grade test scores. In fact, the effect is "precisely zero" since we can reject an effect as small as 2.2% of a standard deviation. Furthermore, the null effect does not reflect inadvertent averaging across heterogeneous effects, as we find no treatment effect heterogeneity by gender, entry test scores, parental education, or immigrant background.

Test scores do not capture all of what students learn in school. This is most evidently seen in that successful interventions often have larger effects on longer run outcomes, such as the probability of higher education enrollment, than they do on immediate test scores (e.g., Jackson 2018; Jackson et al. 2020). Therefore, an increasing number of recent studies aim to identify effects on non-test-score outcomes, which could be caused by a wide variety of skills, including what is often, and somewhat imprecisely, called non-cognitive skills. Non-cognitive skills can be affected in school and often have long term impacts on educational and labor market outcomes (Chetty et al., 2011; Cornelissen and Dustmann, 2019; Heckman et al., 2013; Fredriksson et al., 2013). Jackson et al. (2020) find that schools that improve socioemotional development (measured as interpersonal skills, school connectedness, academic engagement, grit, and academic effort) in ninth grade also increase college enrollment and high school graduation rates. Jackson (2018) finds that teachers affect non-cognitive skills, as proxied by absences, suspensions, course grades, and on-time grade progression in 9th grade. These teacher-induced non-cognitive skills predict longer run academic outcomes better than teacher–induced test scores.

We investigate the policy's effects on non-test outcomes using other individual student outcomes as well as school environment measures. Since the policy did not influence short-run test scores, any

[1] However, there are evaluations of specific in-school interventions, for example showing positive effects of intense tutoring (Guryan et al., 2021; Kirkebøen, 2021; Fryer Jr, 2017).

effects on teacher and exam grades (grade 10) or completing the first year of high school (grade 11) would likely be caused by effects on non-cognitive skills. We can reject meaningful effects on these long-run outcomes, suggesting that the extra teachers did not substantially improve non-cognitive skills relevant for academic outcomes. Moreover, using a national student survey conducted among all 10-graders each year we can study effects on the school environment. When constructing an index of the 12 sub-indices that are publicly reported in Norway, including well-being, teacher support, and bullying, we find an effect of the reform of about 5 percent of a student-level SD. While a better school environment likely cause better academic outcomes, which we show in auxiliary analyzes (to be added), the effect of the extra teachers on the environment was likely to small for any effects of the reform on academic outcomes to materialize.

Our results contribute to the literature on the effects of resources on educational outcomes. There is an older literature arguing that increased resources do not lead to better outcomes (Hanushek, 2003). However, recent US studies using school financing reforms or changes in the components of the school financing formula show positive effects (e.g., Baron 2021; Brunner et al. 2020; Jackson et al. 2016; Jackson 2018; Jackson et al. 2020; Lafortune et al. 2018, see Jackson (2020) for a review). In particular, Jackson and Mackevicius (2021) conduct a meta-analysis of 31 US studies with credible identification and conclude that school spending affects test scores and, even more so, longer run educational attainment.[2]

Most of these studies use non-tailored spending changes, but two recent studies are able to investigate what type of spending matters or in which context the effects are largest. Baron (2021) cleverly exploits elections in Wisconsin, where there is a legal requirement to have referenda before changing school expenditures. In particular, there are different referenda for operational and capital expenditures. Using a regression discontinuity framework, he finds that operational expenditures positively affect test scores and post-secondary enrollment, but there are no effects of increased capital expenditures. Brunner et al. (2020) instead exploit district variation in union strength and find larger effects of school spending on students' performance in areas with stronger unions, where the funds where spent on teacher compensation (rather than hiring new teachers, as in districts with weak unions). Our study provides additional evidence that hiring more teachers does not impact student learning.

Our study also adds to the literature on the effects of class sizes on performance. As in the case of spending, this literature yields mixed results. In a landmark study, Krueger (1999) investigated the effects of Project STAR (Student-Teacher Achievement Ratio), which randomly assigned students to smaller classes from kindergarten through 3rd grade, and found a large increase in performance due to class size reductions. Later work investigated the longer term effects of the STAR experiment and found positive effects on college attendance (Chetty et al., 2011). Contrary to these results, Hoxby (2000) exploits as if random variation across cohorts within schools in Connecticut and can rule out even small effects of class size on performance in math, reading, and writing in 4th and 6th grade. Woessmann and West (2006) also use variation across grades within schools to identify class size effects in 11 countries using TIMSS data. They only find effects in 2 out of the 11 countries, in Greece and Iceland.[3]

[2]Outside of the US, there are fewer studies on the effects of non-targeted resources with credible research designs. One exception is Hægeland et al. (2012), which find that hydro power induced revenues to schooling in Norway had a positive effect on learning outcomes.

[3]Neither the US nor Norway was included in their sample.

Most studies investigating the effects of class size use rule-induced class size reductions, where classes need to split if they reach a certain threshold. A classic example of this method is Angrist and Lavy (1999), which has been interpreted as finding positive effects of smaller classes in Israel, although the results were actually mixed. They found consistent effects for fifth graders, mixed results for fourth graders, and zero effects for a sample of third graders. In a follow up study, Angrist et al. (2019) find no effects using later cohorts. Results from other contexts are also mixed. Positive effects of class size reductions have been found in Sweden (Fredriksson et al., 2013), Denmark (Browning and Heinesen, 2007), and Bolivia (Urquiola, 2006). Argaw and Puhani (2018) find no effect of class size in elementary school on choosing a more academic track in Germany and Angrist et al. (2017) find no effects on gains in learning in Italy. Earlier Norwegian studies have found mixed results. An early study by Bonesrønning (2003) found some weak evidence that larger classes in lower secondary school lead to less favorable short run outcomes, but no subsequent Norwegian study has confirmed it. Leuven et al. (2008) find no effect on short run test scores in lower secondary school. Leuven and Løkken (2020) can reject small effects on long run outcomes from class size reductions in primary and lower secondary schools in Norway and Falch et al. (2017) find no long run effects from size reductions in Norwegian lower secondary schools.

We find no effects on student outcomes of an increased funding of more teachers that reduced class sizes. Evidence on class size effects is often based on natural experiments that induce variation in student teacher ratios at margins far from what is often relevant for policies. In contrast, we contribute to this literature by using an actual policy for reducing class sizes. Our evaluation is policy relevant as we identify the parameter of interest from marginal educational investments.

2 Institutions, reform, and data

Up to grade 10 (age 16), Norwegian municipalities administer the public schools. The municipalities get revenues from income tax, user charges on services, and transfers from the central government. They spend money on schools as well as child care, health care, and other services like water supply and renovation. Schools are free of charge and compulsory throughout grade 10, and the share of lower secondary private schools is small (about 3-4%[4]).

The allocation of input in terms of personell for teaching is highly compensatory in Norway. To reduce achievement gaps, school administrations allocate more resources to schools with disadvantaged student populations and students from low income families with less educated parents typically attend smaller classes (Leuven et al., 2008). We illustrate compensatory policy by means of entry test scores rather than family background. As can be seen in panel A of Appendix Table A.1, one standard deviation increase in students' entry test scores is associated with 3.5 more students per teacher in lower secondary schools, relative to a mean of 15.9.

In the fall of 2012, the Norwegian Parliament decided an extra four-year funding of 600 teachers per year to municipal school administrations (1.5 billion NOK, or 258 million USD). The purpose of the intervention was to increase the student-teacher ratio in regular teaching, allowing schools to tailor the teaching to the individual students better. The funding was intended to increase basic skills, improve the learning environment, and reduce the frequency of special needs education. In line with a compensating resource allocation principle, the extra resources were targeted at 166

[4]Statistics Norway, table 05232 (https://www.ssb.no/statbank/table/05232).

lower secondary schools (grades 8 to 10, ages 14 to 16) with an average student-teacher ratio above 20 *and* average grades below the national mean (in 2012). For schools that met the two conditions, the number of extra teachers varied by school size. Schools received funds to pay for one, two, three, four, or five additional teachers depending on whether the number of students was 0-99, 100-199, 200-299, 300-399, or more than 400.[5] The schools with extra funding were located in 98 different municipalities, and all counties in Norway were represented.

The extra resources were given to increase the number of qualified teachers in regular teaching and were not intended to be used for special needs education. Apart from these requirements, schools were free to organize the regular teaching as they wanted (two teachers in regular classes, divide classes into smaller groups, etc.). A survey administered to the principals at all treatment schools (with 65% response rate) provides information on how schools used the extra funds (Kirkebøen et al., 2017). About three of four principals (77%) report that the extra funds were used to have two teachers available for the class, while 66% report that the funds were used to divide the class into smaller groups. Very few principals report that the extra resources were always used to divide groups by students' skill levels. Still, nearly 60% report that groups were sometimes divided in this way (with 40% reporting that it never was divided by skill level). The principals report that the resources were primarily used in math (95%), Norwegian (90%), and English (78%). Finally, 70% report that resources were divided across all lower secondary grades. Given our precise null findings documented below, it is also interesting to note that 95% of principals believed that the funding improved students' learning outcomes, suggesting that interventions' perceived effectiveness may differ considerably from the evidence based on a credible treatment effects analysis.

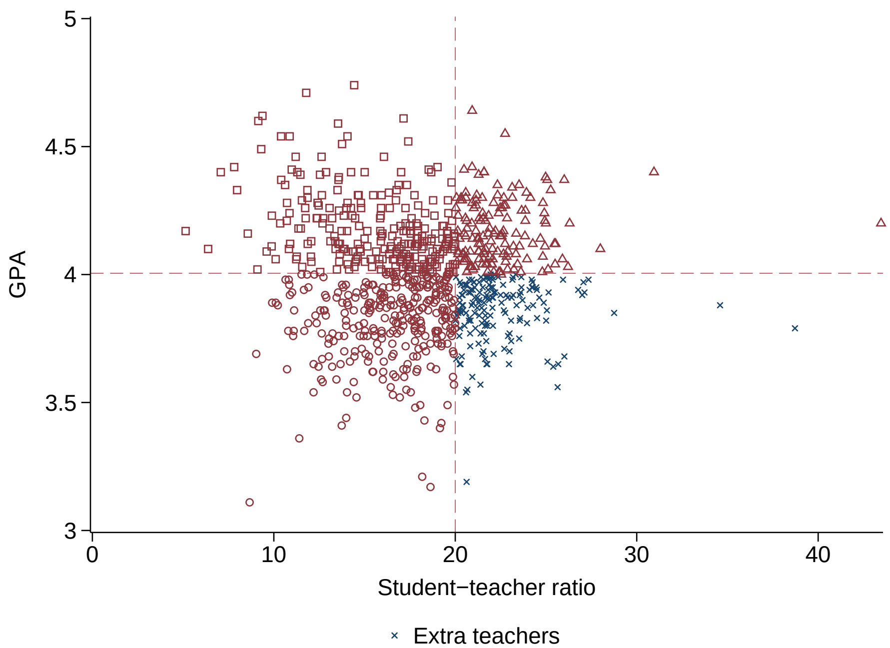
The starting point for our data set is the population of eligible lower secondary schools. The extra funding was allocated based on the 2011/2012 regular student-teacher ratios and grade point average (GPA) based on teacher assessed subject-specific performance and externally-graded exams scores by the end of grade 10. The GPA is dominated by teacher grades, but the written and oral exams also count (weight of about 0.1). Student-teacher ratios and GPA are well defined for 859 out of 1 089 public lower secondary schools, containing 97.5 percent of the students. The remaining 230 schools are small, and many of these schools cater for special needs students.

Figure 1 shows how schools are distributed along the student-teacher (x-axis) and the GPA (y-axis) dimensions in the pre-reform year (2011/2012). The treated schools, marked with blue x, have GPA below the average and above-average group size. The two sharp margins enable us to use a regression discontinuity framework, as discussed below.

We merge our data set of school treatment status with four different data sources, that together allow us to study a variety of outcomes: (i) The compulsory school register ("GSI"), with school-level data enabling us to study resource use in schools, including student-teacher ratios, the share of qualified teachers, and how teacher resources are used (e.g., on regular teaching and special needs teaching); (ii) individual teacher data from matched employer-employee data, which enable us to study teacher characteristics; (iii) student-level outcome data from administrative registers, including standardized tests, end-of-compulsory-school exam scores and teacher grades, and an

[5]The student-teacher ratio is measured as the number of regular-teaching student hours divided by the number of regular-instruction teacher hours. Thus, it disregards student and teacher hours spent on for example special needs teaching and special services for Norwegian learners, and is a measure of the typical group size in regular instruction settings.

Figure 1. Pre-reform GPA and student-teacher ratios of treated and control schools



Note: The figure shows the schools by student-teacher ratio and GPA. These forcing variables are constructed from student and school registers. The red dotted lines mark the two cut-offs. The markers indicate whether the schools received extra teachers (from a separate data source).

early measure of progression in upper secondary school; and finally, (iv) school-level responses from a yearly national student survey, which allow us to study attributes of the learning environment in the schools.

In Table 1, we show key descriptive statistics for our main student-level estimation sample, separately for the pre-funding (2009-2012) and the funding (2013-2016) periods. Here year refer to the first semester of the school year (e.g., 2009 means school year 2009/2010). We include the 2009-2019 outcome cohorts in the analysis, in total about 618,000 students (about 56,000 students per cohort). 169,000 students, or 27 percent, attend treatment schools that receive extra funding during 2013-2016. The extra resources were targeted at schools with below-average GPA and above-average student-teacher ratios. Corresponding with this, the schools that later receive extra funding have about one standard deviation lower GPA than other schools and about one more student per teacher in the pre-treatment year of 2011, as shown by panel A in Table 1. Furthermore, the treated schools are on average significantly larger, with about 341 students vs 279 students for the non-treated schools. GPA levels, student-teacher ratios, and school size all vary more among control schools than among the treatment schools. As we define treatment from the forcing variables in our data, there are a couple of schools we have classified as non-treated that receive extra teachers, this difference of classification affects very few schools and students.

In panel B of Table 1, we report the characteristics of the students. While sex composition is balanced, the treated schools have fewer students with a (or two) college educated parent, more fathers with below-median earnings, and a larger fraction of students with two immigrant parents. However, the differences are mostly modest, with six percentage points more students having at least one highly educated parent and the same for above-median earnings in the untreated schools. Concerning students' outcomes in Panel C, the average entry test score of students in the treated schools is 0.13 SD lower in the pre-treatment years and the differential drops slightly in the post-treatment years. For all outcomes, there is a modest improvement in outcomes for students in the treatment school from pre- to post-treatment years. This is to be expected, as the schools are selected based on poor performance in 2011 (Chay et al., 2005). In the untreated schools, there is a mix of changes in both directions from pre- to post-treatment cohorts.

3 Empirical strategy

Schools were treated by increased funding if they had a student-teacher ratio (STR) above 20 and average grades (GPA) below the national mean in 2011 (i.e., the school year 2011/2012, following the notation from the previous section). These two necessary conditions (cutoffs) place each school in one of the four quadrants of Figure 1. The funding assignment offers two different margins to evaluate treatment effect

$$\begin{aligned} |STR-20| &\rightarrow 0, \quad GPA < Mean \quad (Student-teacher\,ratio\,margin) \\ |GPA-Mean| &\rightarrow 0, \quad \quad STR > 20 \quad \quad (GPA\,margin) \end{aligned} \tag{1}$$

and two placebo margins that can be used to validate the design

$$\begin{aligned} |STR-20| &\rightarrow 0, \quad GPA > Mean \\ |GPA-Mean| &\rightarrow 0, \quad \quad STR < 20 \end{aligned} \tag{2}$$

The reform lends itself to an RD analysis and we start out by estimating the specifications above, using parametric RD estimation and local regressions. We will use the rdrobust and rdplot packages (Calonico et al., 2014) to estimate local linear regressions and "bias-corrected confidence intervals", which are based on higher order polynomials, and to plot the results. Effects are estimated seperately for the two different cutoffs, and the sample is restricted to +/- 1 standard deviation along the specific dimension in these RD analyses.

In RD estimation, there is generally a trade-off between unbiasedness (improved by a small bandwidth) and precision (better by a wider bandwidth and more data). There are several choices to be made in how to summarize the effects and how to analyze the data when we have more than one forcing variable. For instance, one can run separate analyses for the two different margins, and this can be done either parametrically or non-parametrically. Alternatively, one can use both cutoffs simultaneously (Cattaneo et al., 2020). There are also choices to be made with respect to optimal bandwidths and the functional form of the running variables.[6] We will show that our results are robust with respect to the choice of empirical model.

Our preferred model is a parametric specification where both cutoffs are used to identify the effects of the policy. This choice is guided by transparency (as it is easy to understand), credibility

[6]Rather than using the RMSE optimal bandwidth selection of the rdrobust and rdplot packages, we will use a fixed bandwidth of 0.5 SD for the RD estimation. We found that the RMSE optimal bandwidths (typically 0.2-0.3 SD) produced implausibly strong gradients around the cut-offs and correspondingly implausibly large, although imprecise, estimates, cf. Figures 2 and A.1.

Table 1. Descriptive statistics

	Untreated schools		Treated schools	
	2009-2012	2013-2016	2009-2012	2013-2016
A. Pre-treatment school characteristics (2011)				
GPA (de-meaned and standardized)	0.202		-0.743	
	(1.013)		(0.544)	
Student-teacher ratio (de-meaned and standardized)	-0.245		0.734	
	(0.999)		(0.570)	
Students	279		341	
	(139)		(102)	
B. Student characteristics				
Female students (share)	0.488	0.491	0.487	0.487
	(0.500)	(0.500)	(0.500)	(0.500)
No parent with higher education (share)	0.482	0.429	0.543	0.488
	(0.500)	(0.495)	(0.498)	(0.500)
Father has below median income (share)	0.488	0.485	0.533	0.540
	(0.500)	(0.500)	(0.499)	(0.498)
Two foreign-born parents (share)	0.084	0.109	0.147	0.177
	(0.278)	(0.312)	(0.354)	(0.382)
C. Student outcomes				
Average 8th grade test score	0.022	0.016	-0.114	-0.082
	(0.920)	(0.924)	(0.920)	(0.919)
Average 9th grade test score	0.024	0.021	-0.100	-0.074
	(0.917)	(0.922)	(0.928)	(0.928)
Written exam score	0.053	0.071	-0.050	-0.009
	(0.989)	(0.973)	(0.985)	(0.972)
Average teacher grade 10th	0.044	0.031	-0.100	-0.067
	(0.985)	(0.984)	(1.012)	(1.011)
Completed first year high school	0.856	0.880	0.833	0.860
	(0.351)	(0.325)	(0.373)	(0.347)
Number of students	168 305	160 132	63 135	60 712

Note: Years refer to year of school register data and 8th grade test score, and the first semester of the school year. The pre-treatment school characteristics are those used to assign schools to treatment, and are based on 2011 data. While first year high school completion is defined for cohorts with 8th grade tests from 2015 and 2016 only, other outcomes are observed for 93-99 percent of students (most missing for written exam score and teacher grades). Test and exam scores as well as average teacher grades are standardized within year in the total student population.

(by having "in-built" placebo checks), and precision gains (by exploiting more of the data). To simplify the presentation of the model, let z_S be the running variable on the student-teacher ratio margin measuring the distance to the cutoff, and z_G be the running variable on the GPA margin, such that treated schools have $z_S > 0$ and $z_G > 0$. We estimate

$$y = \delta(Z_G Z_S) + \gamma_1 Z_G + \gamma_2 Z_S + \eta_1 z_G + \eta_2 z_S + \eta_{11} z_G^2 + \eta_{22} z_S^2 + \eta_{12} z_G z_S + x\beta + \epsilon, \qquad (3)$$

where y denotes the outcome of interest, z are our running variables and Z are dummies for $Z_a := z_a > 0$. The control variables in x include gender, age, year fixed effects, a cubic in the 8th grade test score, and parental education. Control variables are included to increase precision but results without controls are very similar (as we show in the Appendix). As the treatment is at the school level, standard errors will be clustered at the school-level when we study individual-level outcomes. Our main coefficient of interest is δ, which shows the effects of crossing both cutoffs. The γ coefficients provide placebo effect estimates as they show the effects of crossing only one of the cut-offs. Evidence that $\gamma_1 = \gamma_2 = 0$ and $\delta = 0$ for pre-reform cohorts without treatment and that $\gamma_1 = \gamma_2 = 0$ for treated cohorts suggest that the parametric specifcation of z_S, z_G and x is sufficient to control for confounding differences, and support our identification strategy. The results will be summarized in coefficient plots with confidence intervals.

4 Results

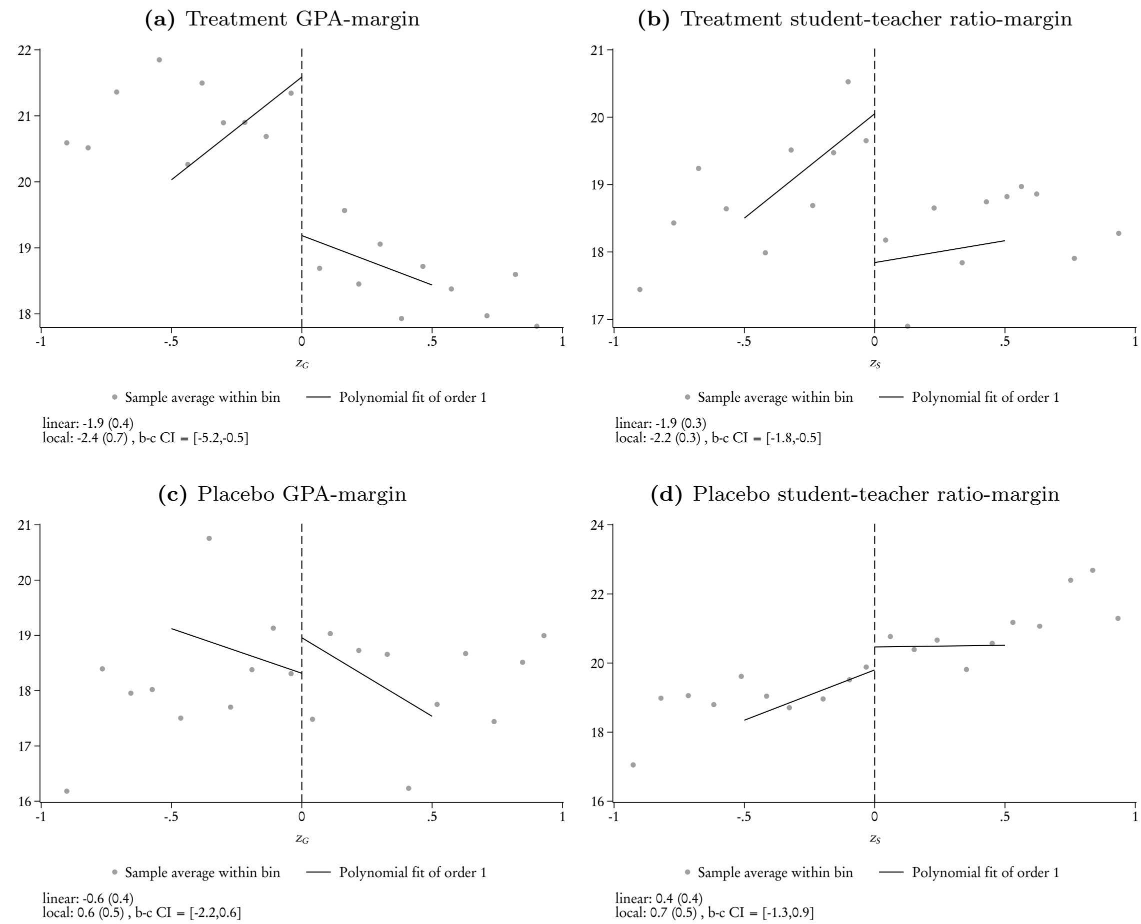
4.1 Effects on school resources

The policy funded new teaching positions in targeted schools over four years (2013-2016). We start by examining how the policy affected the student-teacher ratio. We first present the results for the two treatment margins and the two placebo margins. Panel (a) of Figure 2 shows the RD-estimate for the treatment GPA-margin (with $z_S > 0$). There is a clear drop in the student-teacher ratio around the cut-off from about 21.5 to 19. The linear, local linear, and bias-corrected local linear estimates are all similar and statistically significant with a reduction of about two students per teacher. Panel (b) shows similar results for the treatment student-teacher ratio margin (with $z_G > 0$). As for the GPA-margin, there is a drop of about two students per teacher, from 20 to 18, and all estimates are of the same magnitude and statistically significant.

In sub-figures (c) and (d) we show similar results for the GPA and student-teacher ratio placebo margin, i.e. with $z_S > 0$ and $z_G > 0$ respectively. Here, we expect zero effects on the student-teacher ratio for each of the two margins. This is indeed what we find. All estimates are statistically insignificant, and much smaller than the estimates in sub-figures (a) and (b).

In Figure 3, we combine the two necessary conditions for funding in a single parametric specification as explained when presenting eq. (3). Figure 3 shows the point estimates and 95% confidence intervals for separate school cohorts in treatment schools (blue circle symbol), GPA placebo schools (i.e., schools with $z_G > 0$ and $z_S < 0$; green diamond symbol), and student-teacher ratio placebo schools (i.e., schools with $z_S > 0$ and $z_G < 0$; red triangle symbol). All estimates are within-year, relative to the high-GPA low-student-teacher ratio schools, and conditional on the parametric specification in eq. (3). The treatment coefficients during the treatment period,

Figure 2. RD-estimates of the effect on student-teacher ratio



Note: The graphs show RD-estimates. Data are from years 2013-2016 and the outcome is the student-teacher ratio for regular teaching. Figure notes show coefficients and standard errors from linear and local regressions. All analyses uses student weights. The lines show the local linear regressions and are estimated using Calonico et al. (2014), with triangular weights and a fixed bandwidth of 0.5. Bias-corrected confidence interval (estimated using higher-order polynomial) in brackets. Bins are quantile-based.

shaded in the graph (2013-2016), can be interpreted as effects of the funding. Year refers to the year of the 8th grade test at the entry in lower secondary schools.

We see that the policy reduced the class size during the treatment years (blue circles in the shaded area of the figure). The average effect over the four treatment years is a reduction of 2.3 in the student-teacher ratio, which is substantial compared to the sample average of 18. In Figure A.3 in the Appendix, we show the same results with logged outcomes, and we note a reduced student-teacher ratio by about 12% compared to the pre-policy baseline. These results show that the funding policy had the intended effect on teacher input. In Appendix Figure A.4, we show that the policy increased the number of teachers by about 3.9 teachers (17 percent) in treated schools, which closely corresponds to the number of teachers funded. Thus, there is no evidence of municipalities' funding being crowding-out or substitution to other schools.

We check the validity of our design using three distinct placebo dimensions (pre-treatment cohorts, GPA margin, and student-teacher ratio margin) along with other balancing tests discussed in Section 4.2, all of which support our identification strategy. The first placebo dimension is treatment school coefficients before the funding period (2009-2012), which should be zero under a parallel trends assumption. Changes in the student-teacher ratio (or other outcomes) before the funding period would suggest trends within treatment schools that may bias our effect estimates. We see no indications of significant cohort differentials in schools that later become treatment schools.[7]

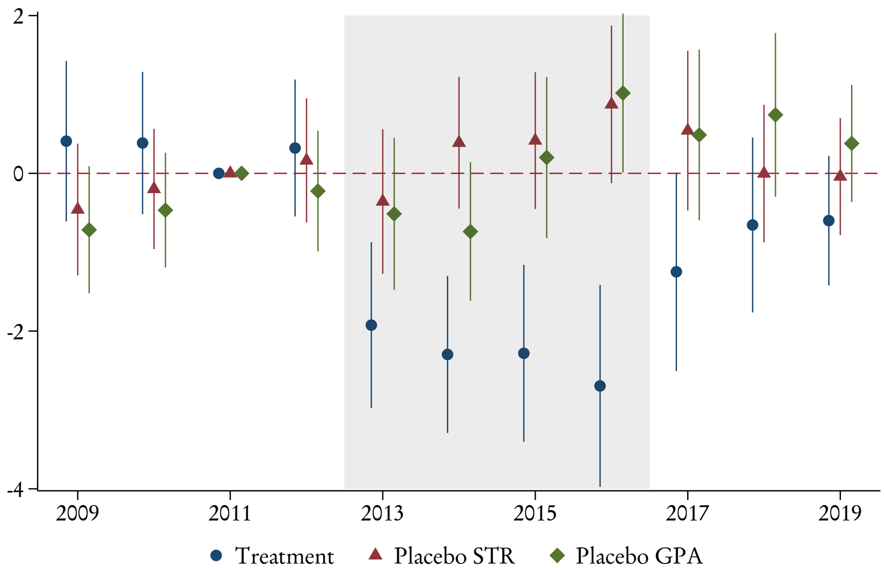
The second and third placebo dimensions are the non-funded schools with low GPA and high student-teacher ratio, respectively. Since none of these placebo schools received funding during the period, these schools can be used to rule out other changes concurrent with the funding policy that presumably would have affected all schools. In Figure 3, there are no significant placebo effects in the period 2013-2015. However, both placebo effects are positive in 2016 and either close to being (STR) or just statistically significant (GPA). The joint test of all 2013-2016 placebo effects is not statistically significant (p-value of 0.135). In all, we conclude that there is strong evidence of a substantial change in the student-teacher ratio.

While the policy aimed to lower the student-teacher ratio in regular 8th to 10th grade instruction, school administrators may reallocate resources within or between schools as a response to improved funding (see discussion in Hoxby (2000)). From a central policy perspective, extra resources could potentially crowd out other input like teaching assistants. Furthermore, the schools may face constraints in which teacher(s) they are able to employ, which may influence the effect of the extra funding. To check whether such adoptions or circumstances influence our results, we further investigate several margins for school resource use. In Appendix Figure A.5, we show effects on total teacher hours and different types of teacher hours.[8] We find an effect on regular teacher hours which is very similar to the effect on total hours as intended by the policy. There is no effect, however, on the proportion of teacher hours used for special needs teaching or special needs language instruction (which the funding specifically was not for). In Appendix Figure A.6, we investigate the qualifications of the schools' teacher stock. We find an increase in teacher hours

[7]The treatment is assigned based on the student-teacher in 2011, thus the placebo effect for this year is zero by construction.

[8]In Figures A.5 and A.6, we study teacher-student ratios, rather than student-teacher ratios, in order to have a fixed denominator and be able to decompose the teacher hours in the nominator. The denominator is total student hours, including e.g. special needs teaching, and thus slightly different from the nominator of hour measure of student-teacher ratio. However, this difference in minor for student hours.

Figure 3. Effects on the student-teacher ratio



sample mean (SD) = 18.0 (3.8)
pooled 2013-2016 estimate = -2.3** (0.3)
p-value joint test of 2013-2016 placebos = 0.1350

Note: The graph shows estimates and confidence intervals from estimating eq. (3). Outcome is regular student-teacher ratio. The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treatment period is shaded. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years, and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses school-level data with student weights and robust standard errors.

taught by qualified teachers, which is very similar to the total effect, and no effect on the number of hours taught by teachers without formal qualifications. Another possible margin of adjustment for the schools is teaching assistants, which is mostly used to support special needs teaching (about 75% of assistant hours are used for this purpose). We do not find any effect on the ratio of assistant hours to student hours, nor any effect on assistant hours used for regular teaching. Finally, we study average characteristics of the teachers using matched employer-employee data (Appendix Figure A.7).[9] We find neither statistically significant nor quantitatively substantial effects on average time since completed education, tenure at the school, the share of teachers with a teaching degree, or average sickness absence of the teachers.

Overall we find no evidence of crowding out of other school input nor any indications that the treatment schools met any restrictions in the hiring process. The funding of extra teachers reduces the student teacher ratio by about two, from around 21 students per teacher.

4.2 Effects on test scores

We now focus on the main objective of the reform, which was to improve student outcomes. We start by studying the effects on standardized test scores in grade 9, when students have been exposed to the treatment for about one year. Figure 4 shows RD-results for the two margins, similar to what we did for school resources in section 4.1. In Figure 2, we found a substantial effect on student-teacher ratios on both treatment margins and none on the placebo margins. Thus, if the extra teachers were effective at improving student outcomes we should see a similar pattern here. However, we do not see any indication of differences in test scores. The estimated effects are small in all of the sub-figures and never statistically significant. However, the RD-estimates are not very precise, and in several cases we cannot rule out effects of 5% of a standard deviation.

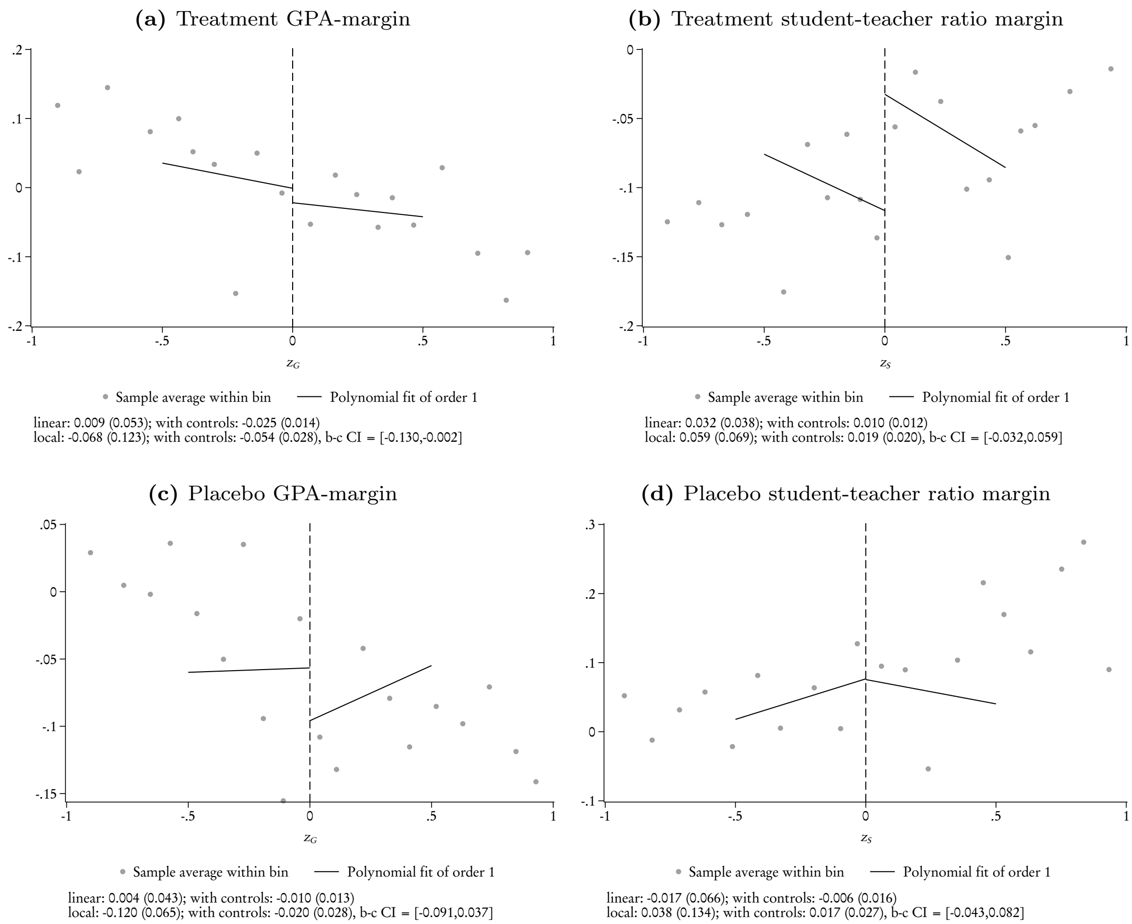
Figure 5 shows the effects of our main specification on students' 9th-grade standardized test scores. One year with a lower student-teacher ratio does not improve the treated students' score on the national standardized test. The average estimated effect across the four treated cohorts is 0.6% of a standard deviation, with the lower and upper band of the confidence intervals ranging from -1.6% and 2.8% of a standard deviation. Any effects of the school funding on 9th-grade test scores are accordingly minor at best. The results are similar if we do not include control variables (see Appendix Figure A.2).

Compositional changes within treatment schools are a potential concern. We examine this by studying placebo effects on students' entry test scores in lower secondary schools as well as on parental characteristics. Since funding in lower secondary schools cannot influence test scores at entry, any such placebo effects indicate compositional changes correlated with the treatment, which may bias our estimates. Although we control for students' entry test scores when estimating the effect on 9th grade test score, which accounts for such potential bias and improves statistical precision, any within-school changes in students' cognitive abilities correlated with treatment would represent a concern since it suggests that other unobserved factors were co-occurring with our treatment.

Reassuringly, there are no systematic changes in entry test scores nor in parental earnings, education, or immigrant background in treatment schools before, during, or after the funding

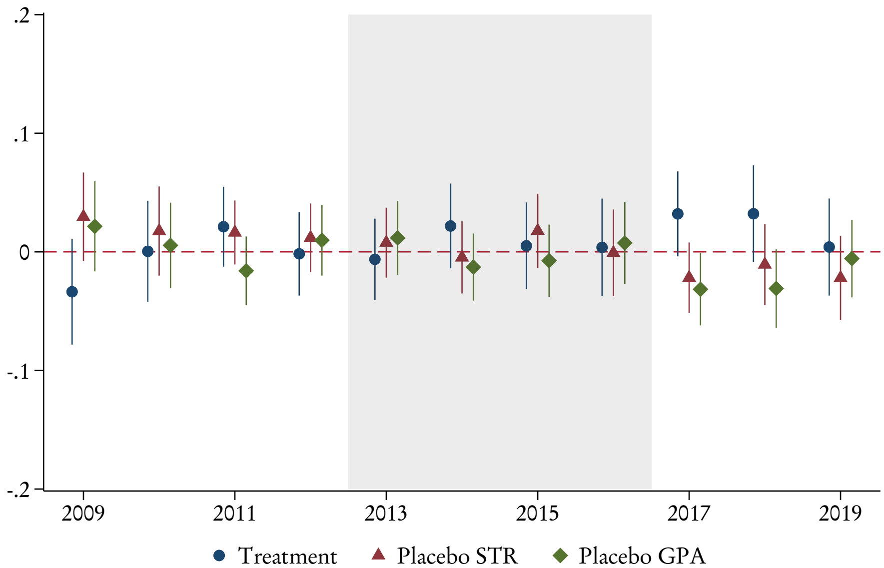
[9] This data source is available only for 2015-2018 cohorts.

Figure 4. RD-estimates of effects on 9th grade test scores



Note: The graphs show RD-estimates for the average score from the 9th grade test. Data are students sitting the grade 8th test in 2013-2016. The graph is based on school-level data and student weights. The lines show the local linear regressions and are estimated using Calonico et al. (2014), with no additional control variables, triangular weights and a fixed bandwidth of 0.5. Bins are quantile-based. Figure notes show coefficients and standard errors from student-level linear and local regressions. Student controls include gender, year dummies, a cubic in the 8th grade test score and parental education. Bias-corrected confidence interval (estimated using higher-order polynomial) in brackets.

Figure 5. Effects on 9th grade test scores



sample mean (SD) = 0.018 (0.908)
pooled estimate = 0.006 (0.011)
p-value joint test of 2013-2016 placebos = 0.8390

Note: The graphs show estimates and confidence intervals from estimating eq. (3). Outcome is 9th grade test scores and the control variables are gender, age, year fixed effects, a cubic in the 8th grade test score, and parental education The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treatment period is shaded. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses student-level data and cluster standard errors at the school level.

period, as seen in Appendix Figure A.8. Nor are there any systematic differences in any of the placebo schools. The 2016 treatment schools cohort seemingly have slightly higher entry test scores, as do the 2014 GPA placebo schools cohort. However, this is expected given the number of estimated coefficients, and there are no indications of any trends or systematic differences.

The precisely estimated zero average effect strongly restricts potential effects for subgroups. Any effect of the extra funding either must be small, limited to a small group of students, or counteracted by a negative effect on other students. Previous research has found that primarily disadvantaged students benefit from extra funding (Jackson, 2020). In Appendix Figure A.9, we investigate heterogeneous effects by student sex and parents' immigrant background, earnings, and education. We do not find effects for any group. Indeed, we can reject effects larger than 4% of a standard deviation for most groups. The single exception is children of immigrants, for whom we can only reject effects larger than 6%. This mostly reflects lower precision for this group (which is smaller than the other studied), since the point estimate of less than 2% hardly indicates any large effect.

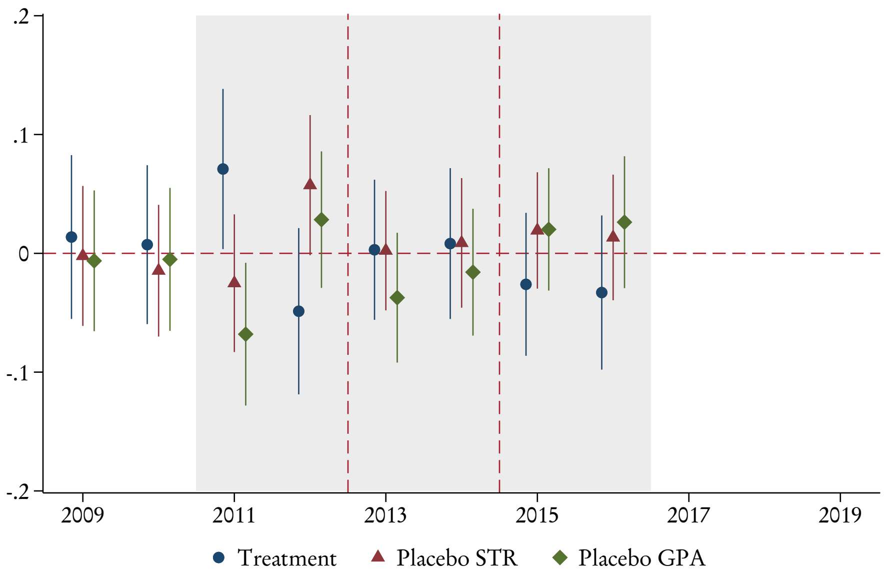
4.3 Effects on longer-term outcomes and the school environment

Treatment school students have been exposed to a lower student-teacher ratio for one year when taking the 9th-grade test. Since exposure over a more extended period may have a more substantial impact, we check whether the funding impacted externally-graded exam scores at the end of compulsory education (in grade 10). For the effects of exam scores, school-cohorts are exposed to one year (2011 and 2016), two years (2012 and 2015), or three years (2013-2014) with extra funding. The cohorts treated for three years are delimited with vertical dashed lines in Figure 6, with exam score results in panel (a). The average effect across all treated cohorts is -1.2% of a standard deviation. Following a dose-response rationale, being exposed to more resources over a more extended period should have a larger impact. However, the average impact for school-cohorts exposed three years to extra funding is also insignificant and close to zero at 0.5% of a standard deviation.

We are also able to investigate effects on a wider set of non test-score measures of academic outcomes as well as effects on non-cognitive outcomes such as the school environment. These are important complementary measures since test scores do not capture the full extent of students' learning. To the extent that non-cognitive skills impact longer-term outcomes, we expect that such effects could show up in teacher-graded tests and school dropout. Specifically, since the funding did not improve the 9th-grade standardized test scores nor 10th-grade exam scores, any effects on 10th-grade teacher-assessed grades and high school completion would likely have been caused by non-cognitive skills. We find no significant effect on teacher-assessed grades, with the average effect across all treated cohorts is -2.5% of a standard deviation for teacher-assessed grades and the impact for those exposed for three years being -3.3% of a standard deviation. Finally, there are no changes in completing the first year of high school, although the precision for this outcome is lower. The 95% confidence interval ranges from a reduction of 2.2 percentage points to an increase of 0.6 percentage points, which is wide given that 87.3% of students complete their first year of high school on time. Appendix Figure A.11 shows that the results are similar without controls.

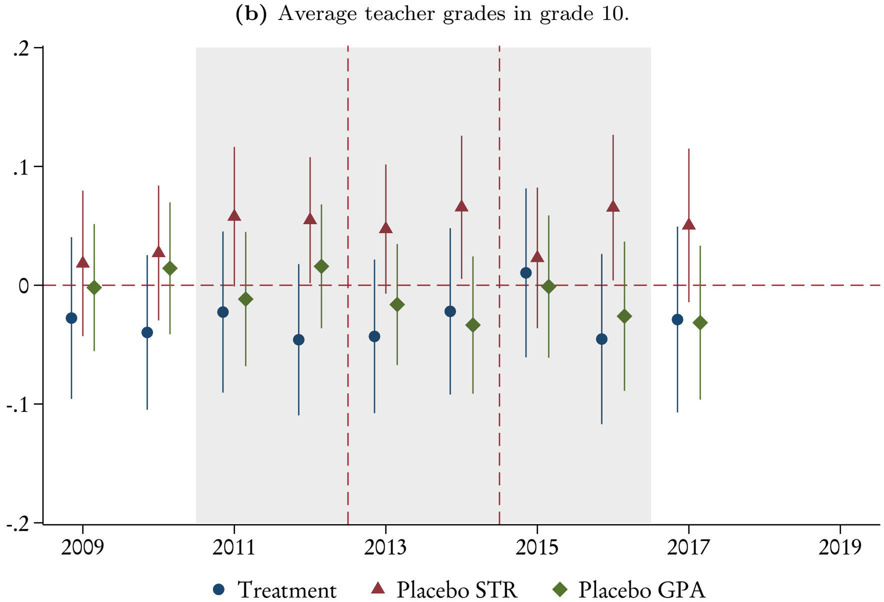
Figure 6. Effects on teacher grades and exam score grade 10. With controls.

(a) Exam score grade 10



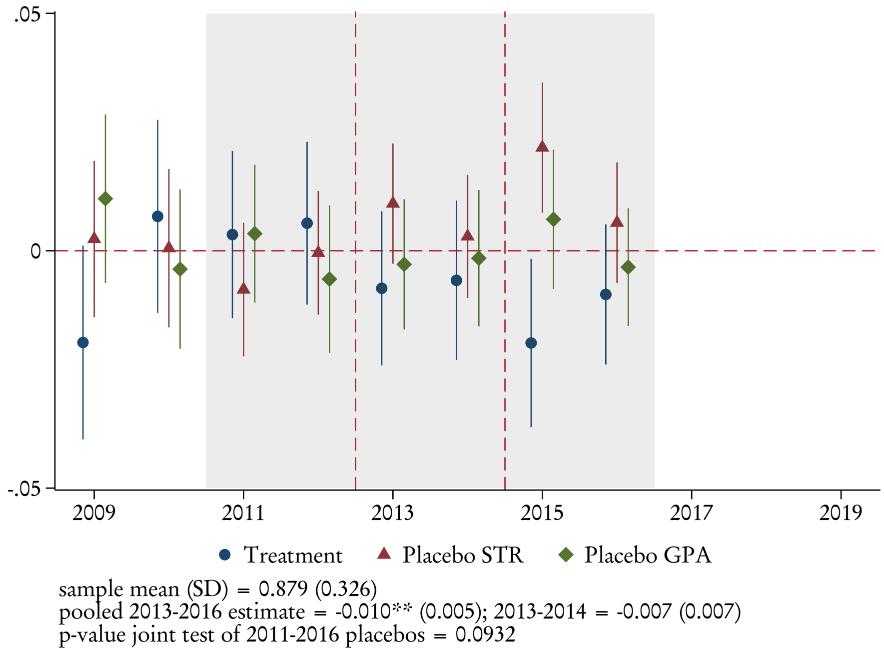
sample mean (SD) = 0.055 (0.974)
pooled 2013-2016 estimate = -0.012 (0.022); 2013-2014 = 0.005 (0.025)
p-value joint test of 2011-2016 placebos = 0.1384

(b) Average teacher grades in grade 10.



sample mean (SD) = 0.028 (0.982)
pooled 2013-2016 estimate = -0.025 (0.028); 2013-2014 = -0.033 (0.030)
p-value joint test of 2011-2016 placebos = 0.1758

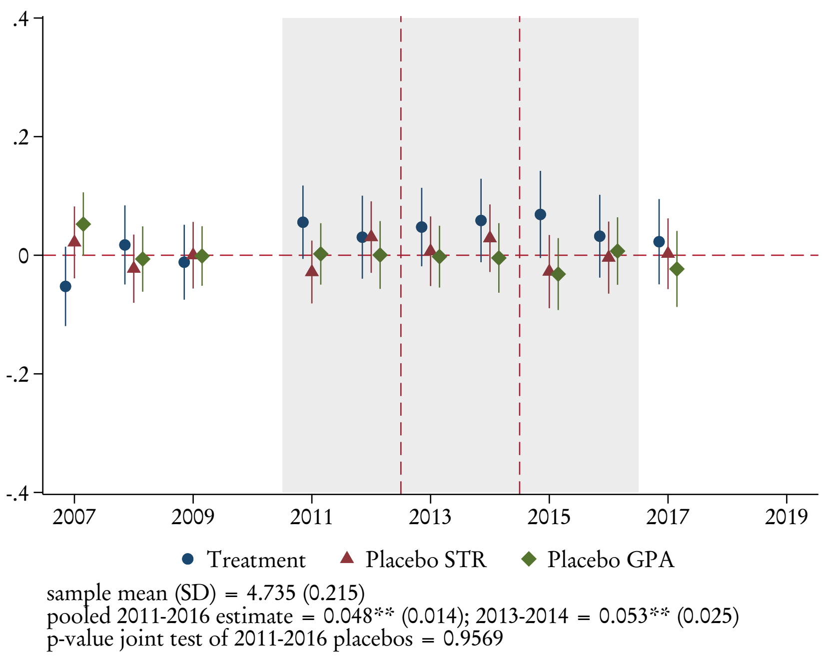
Figure 7. Effects on completion of first year high school



Note: The graph shows estimates and confidence intervals from estimating (3). Outcome is completion of the first year of high school. The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treated cohorts are shaded. The dashed vertical lines indicate delimit the cohorts treated for three years. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses student-level data and cluster standard errors at the school level.

So far, we have seen no indication that the extra funding changed academic outcomes. However, we cannot rule out that the funding impacted students in ways not captured by academic outcomes. To examine the impact on the school environment, we have constructed an index based on student responses to survey items including well-being in school, teacher support, and bullying in a nationwide survey of 10th graders. In Figure 8, we present results where we aggregate responses from 12 different sub-indices into a school environment measure, where higher values imply better school environment. Contrary to the results described above, these analyses suggest that the funding improved the school environment by about 5 percent of a student-level standard deviation. These results are supported by non-significant (placebo) treatment effects in the pre-treatment period. Moreover, no effects are found among placebo schools, suggesting that the identifying assumptions hold. When conducting the same analysis on the separate indices (see Appendix Figures A.12 and A.13), we find that the estimates are consistently in the direction of the school environment being improved . These results suggest that the school environment improved but the precision of these estimates is lower than for individual test scores. In any case, to the extent that the funding positively affected the school environment, this effect was too small to materialize in better academic outcomes.

Figure 8. Effects on the school environment index.



Note: The graphs show estimates and confidence intervals from estimating (3). Outcome is an index summarizing most of the sub-indices from the student survey presented in Figures A.12 and A.13. The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treated cohorts are shaded. The dashed vertical lines delimit the cohorts treated for three years. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses school-level with student weights and cluster standard errors at the municipality level.

5 Conclusion

We have studied a large-scale policy that provided extra funding of 600 teachers per year to 166 lower secondary schools in Norway over a four-year period (1.5 billion NOK in 2012, or 258 million USD). The funding assignment was based on two sharp conditions (margins), allowing us to credibly identify the effects of additional resources using a regression discontinuity framework. We find that the extra funding reduced the student-teacher ratio by about 12% compared to the pre-policy baseline. There is no evidence of crowding out of other school inputs or that the treatment schools met any restrictions when recruiting additional teachers. The large shift in student-teacher ratio did not affect academic outcomes, as measured by 9th-grade test scores, end-of compulsory school grades, and high-school progression. The null effect on test scores is particularly precise as we can reject an effect as small as about 2% of a standard deviation. We do, however, find effects on the school environment but these effects are likely to small to induce effects on academic outcomes. The assignment mechanisms enabled us to examine our identifying assumptions using several placebo dimensions. Together with other robustness tests, these analyses support our conclusions and do not raise any concern that the identifying assumptions are violated. The richness of our register data allows us to rule out several other compensatory side-effects of the policy, such as changes in teacher composition, teacher sickness absence, or funding to special needs education.

While it is difficult to know why the extra school funding did not improve academic outcomes in our context, two factors deserve specific attention. First, if there are diminishing marginal returns to school spending, we may expect additional funding to have small effects in a country like Norway, where spending is already high (Schleicher, 2018). This explanation is challenged by the fact that we identify marginal effects for students in schools with a higher student-teacher ratio and weaker school performance than the average. A related institutional aspect is that Norwegian school administration practice compensating resource allocation in favor towards schools with disadvantaged children. The between-school variance in test scores is low by international standards; only 10% of the variance is between schools in Norway compared to 20% in the US (Schleicher, 2018). Finally, there is no consesus on diminishing marginal returns to school spending and Jackson and Mackevicius (2021) find little evidence of this in the US.

A second explanation relates to differential effects across types of school spending. Although the current consensus is that school spending has positive effects (Jackson and Mackevicius, 2021), the type of spending also seems to matter. Some recent studies suggest that while additional operational expenditures (e.g., teacher pay) impact student outcomes, hiring more teachers seems to have no effects (Brunner et al., 2020; Baron, 2021; Abott et al., 2020). Our study fits with these findings. Moreover, although the class size literature findings are mixed, several studies find no effects of reducing class size, including in Norway (Falch et al., 2017; Leuven et al., 2008; Leuven and Løkken, 2020). Unlike most of the class size literature, which uses natural experiments at margins far from what is relevant for policies (typically halving the class size at a certain cutoff), our study identifies the parameter of interest for funding policies by the national government. To reduce class sizes overall, more teachers are needed and the reform we study was large in that respect. In our treatment schools, the number of students per teacher was reduced from about 20 to about 18. In order to know whether the results generalize to other contexts, and whether hiring more teachers is cost effective, we need more specific policy based evidence from other countries.

References

Abott, C., V. Kogan, S. Lavertu, and Z. Peskowitz (2020). School district operational spending and student outcomes: Evidence from tax elections in seven states. *Journal of Public Economics 183*, 104–142.

Angrist, J. D., E. Battistin, and D. Vuri (2017). In a small moment: Class size and moral hazard in the italian mezzogiorno. *American Economic Journal: Applied Economics 9*(4), 216–49.

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Angrist, J. D., V. Lavy, J. Leder-Luis, and A. Shany (2019). Maimonides' rule redux. *American Economic Review: Insights 1*(3), 309–24.

Argaw, B. A. and P. A. Puhani (2018). Does class size matter for school tracking outcomes after elementary school? quasi-experimental evidence using administrative panel data from germany. *Economics of Education Review 65*, 48–57.

Baron, E. J. (2021). School spending and student outcomes: Evidence from revenue limit elections in wisconsin. *American Economic Journal: Economic Policy*.

Bonesrønning, H. (2003). Class size effects on student achievement in norway: Patterns and explanations. *Southern Economic Journal*, 952–965.

Browning, M. and E. Heinesen (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics 109*(2), 415–438.

Brunner, E., J. Hyman, and A. Ju (2020). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics 102*(3), 473–489.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *Stata Journal 14*(4), 909–946.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2020). Analysis of regression-discontinuity designs with multiple cutoffs or multiple scores. *The Stata Journal 20*(4), 866–891.

Chay, K. Y., P. J. McEwan, and M. Urquiola (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *The American Economic Review 95*(4), 1237–1258.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Coleman, J. S. (1968). Equality of educational opportunity. *Integrated Education 6*(5), 19–28.

Cornelissen, T. and C. Dustmann (2019). Early school exposure, test scores, and noncognitive outcomes. *American Economic Journal: Economic Policy 11*(2), 35–63.

Falch, T., A. M. J. Sandsør, and B. Strøm (2017). Do smaller classes always improve students' long-run outcomes? *Oxford Bulletin of Economics and Statistics 79*(5), 654–688.

Fredriksson, P., B. Öckert, and H. Oosterbeek (2013). Long-term effects of class size. *The Quarterly Journal of Economics 128*(1), 249–285.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments*, Volume 2, pp. 95–322. Elsevier.

Guryan, J., J. Ludwig, M. P. Bhatt, P. J. Cook, J. M. Davis, K. Dodge, G. Farkas, R. G. Fryer Jr, S. Mayer, H. Pollack, et al. (2021). Not too late: Improving academic outcomes among adolescents. *NBER WP* (Number 28531).

Hægeland, T., O. Raaum, and K. G. Salvanes (2012). Pennies from heaven? using exogenous tax variation to identify effects of school resources on pupil achievement. *Economics of Education Review 31*(5), 601–614.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal 113*(485), F64–F98.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jackson, C. K. (2020). *Does school spending matter? The new literature on an old question.* American Psychological Association.

Jackson, C. K., R. C. Johnson, and C. Persico (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics 131*(1), 157–218.

Jackson, C. K. and C. Mackevicius (2021). The distribution of school spending impacts. *NBER WP* (Number 28517).

Jackson, C. K., S. C. Porter, J. Q. Easton, A. Blanchard, and S. Kiguel (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights 2*(4), 491–508.

Kirkebøen, L. (2021). School value-added and long-term student outcomes. *Mimeo SSB*.

Kirkebøen, L. J., A. Kotsadam, O. Raaum, S. Andresen, and J. Rogstad (2017). Effekter av satsing på økt lærertetthet.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics 114*(2), 497–532.

Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics 10*(2), 1–26.

Leuven, E. and S. A. Løkken (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources 55*(1), 309–348.

Leuven, E., H. Oosterbeek, and M. Rønning (2008). Quasi-experimental estimates of the effect of class size on achievement in norway. *Scandinavian Journal of Economics 110*(4), 663–693.

Schleicher, A. (2018). Insights and interpretations. *Pisa 2018 10*.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural bolivia. *Review of Economics and Statistics 88*(1), 171–177.

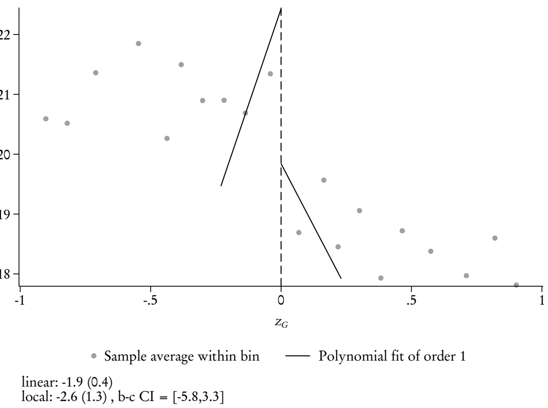
Woessmann, L. and M. West (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in timss. *European Economic Review 50*(3), 695–736.

Table A.1. Compensating school inputs. Association between entry test score and student teacher ratio in the pre-treatment school years (2011-2012).

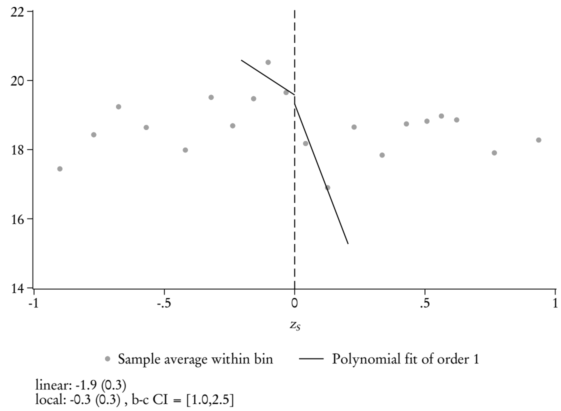
	All schools		Large municipalities
	(1)	(2)	(3)
A. Student-teacher ratio			
Entry test score (8th grade)	3.490**	2.430**	3.235**
	(0.309)	(0.357)	(0.527)
Mean (std dev)	15.9 (2.8)		17.9 (2.2)
B. Student-teacher ratio in regular teaching			
Entry test score (8th grade)	2.568**	0.884**	1.461**
	(0.364)	(0.450)	(0.620)
Mean (std dev)	20.0 (3.2)		22.0 (2.5)
Fixed effects	Year	Year*Municipality	Year*Municipality
N schools	1 574	1 574	239
N students	336 556	336 556	75 665

Note: Each cell is an estimate from a separate model regressing resource inputs in grades 8-10 on average entry test score in 8th grade (beginning of lower secondary) for different samples and specifications. Sample of pre-treament school years 2011-2012. For a year t, test scores are average 8th grade scores of years $t, t-1, t-2$. Top row is total student hours/total teacher hours (Panel A) and bottom row is regular instruction student hours/regular teaching hours (Panel B). Column (1) controls for Year fixed effects, columns (2) and (3) control for Year*Municipality fixed effects. In column (3) we restrict the sample to large municipalities (defined as having more than 9 schools). Standard errors clustered at the school level in parentheses. ** p<.05.

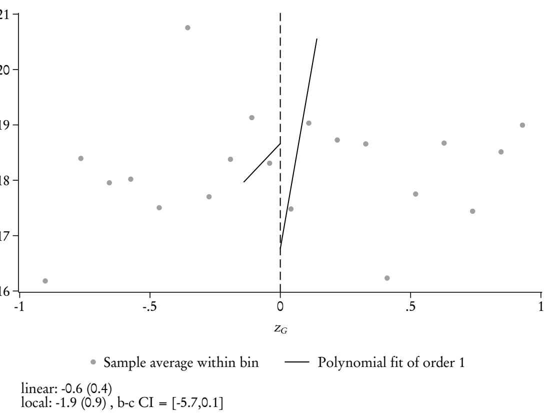
A Results referred to in the text



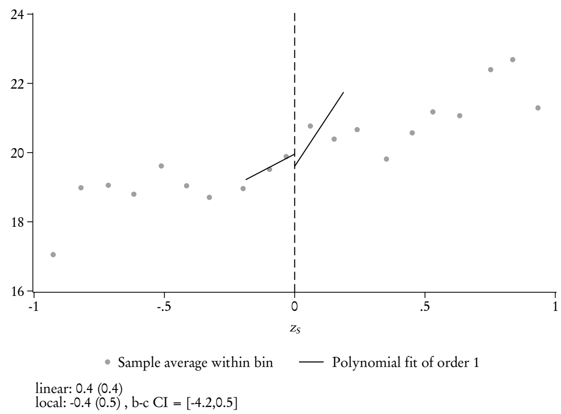
(a) GPA-margin, treatment ($z_S > 0$)



(b) Student-teacher ratio-margin, treatment ($z_G > 0$)



(c) GPA-margin, placebo ($z_S < 0$)

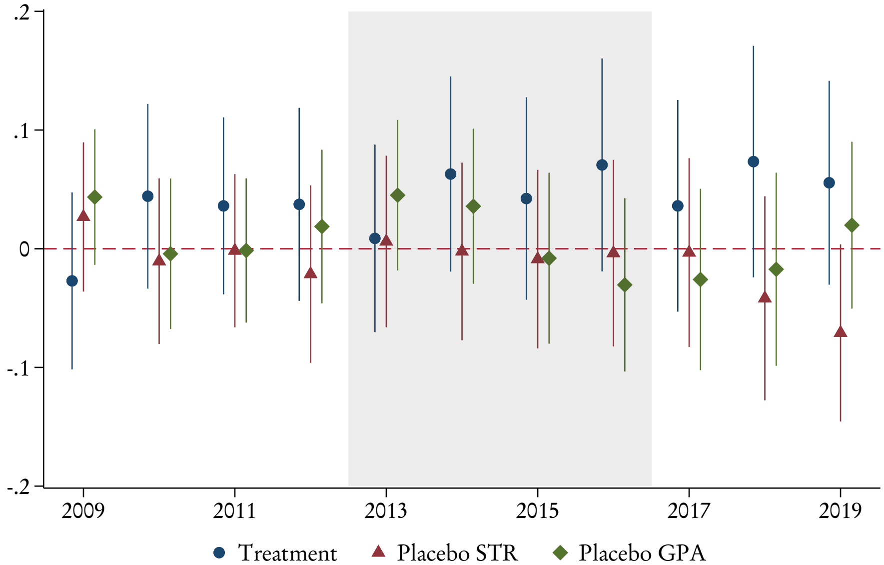


(d) Student-teacher ratio-margin, placebo ($z_G < 0$)

Figure A.1. RD-estimates of the effect on student-teacher ratio, RMSE-optimal bandwidths

Note: The graphs show RD-estimates corresponding to (1). Data are years 2013-2016, outcome is (regular-teaching) student-teacher ratio. Figure notes show coefficients and standard errors from linear and local regressions. All analyses uses student weights. The lines show the local linear regressions and are estimated using Calonico et al. (2014), with triangular weights and RMSE-optimal bandwidth . Bias-corrected confidence interval (estimated using higher-order polynomial) in brackets. Bins are quantile-based.

Figure A.2. Effects on 9th grade test scores, no student level controls.



sample mean (SD) = -0.011 (0.923)
pooled estimate = 0.046 (0.035)
p-value joint test of 2013-2016 placebos = 0.4751

Note: The graphs show estimates and confidence intervals from estimating eq. (3). Outcome is 9th grade test scores. The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treatment period is shaded. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses student-level data and cluster standard errors at the school level.

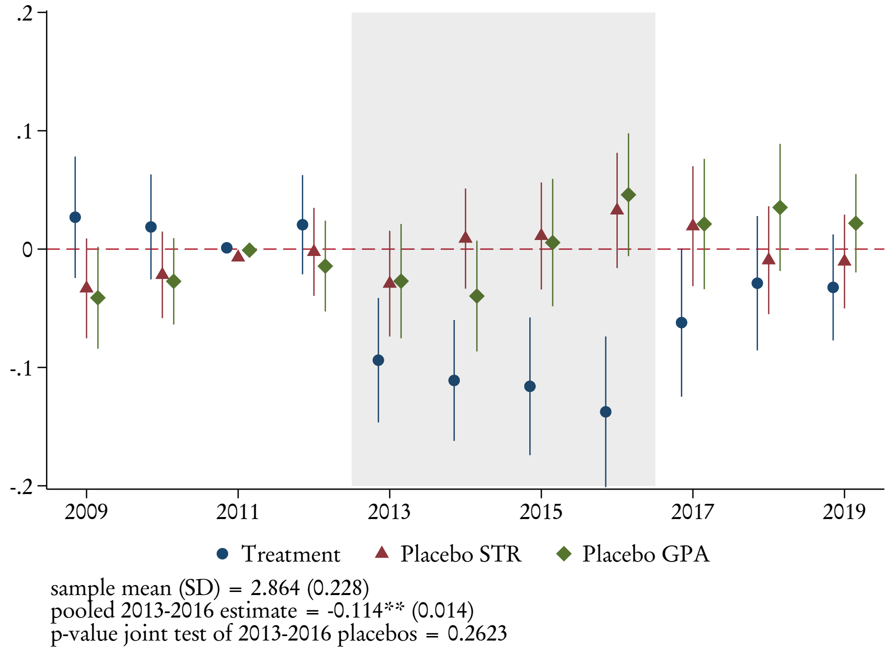
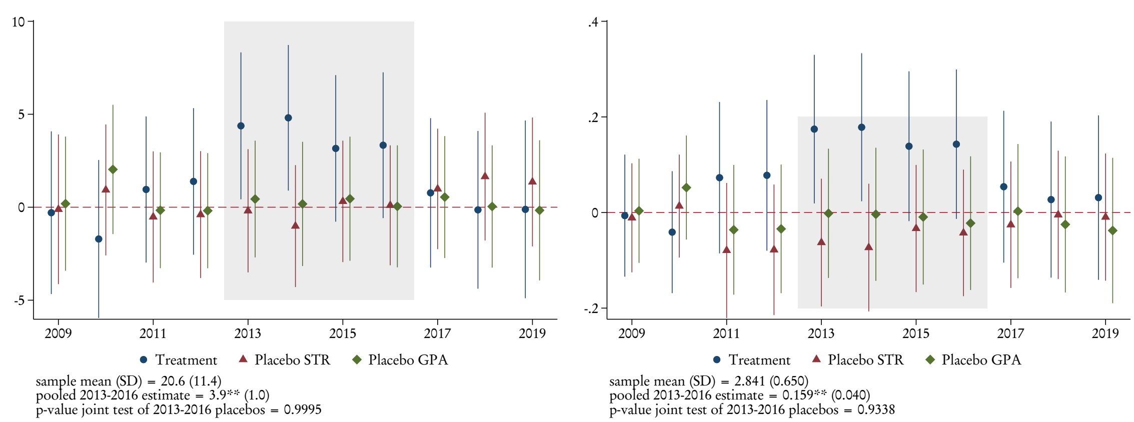


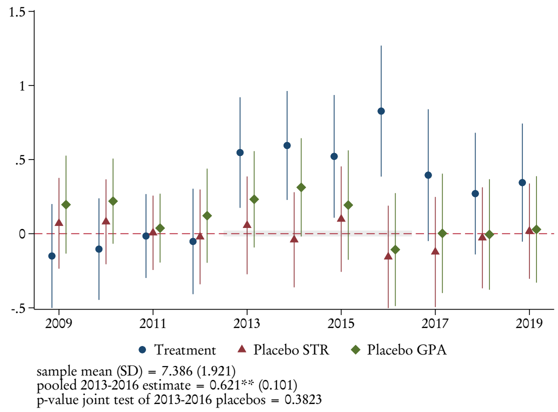
Figure A.3. Effects on log student-teacher ratio
Note: The graph show estimates and confidence intervals from estimating (3). Outcome is log of (regular-teaching) student-teacher ratio. The regression uses student weights and cluster standard errors at the municipality level. See note to Figure 3 for further details.



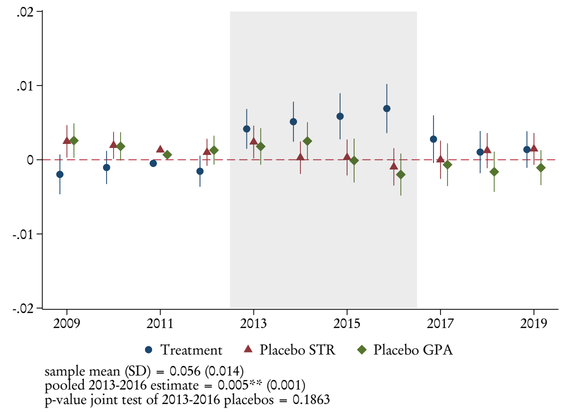
(a) Number of teachers

(b) Log number of teachers

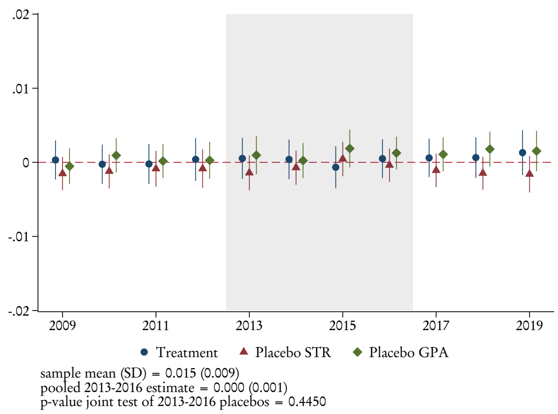
Figure A.4. Effects on number of teachers
Note: The graph show estimates and confidence intervals from estimating (3). Outcome is number of teacher at the school and log number of teachers. The regression uses student weights and cluster standard errors at the municipality level. See note to Figure 3 for further details.



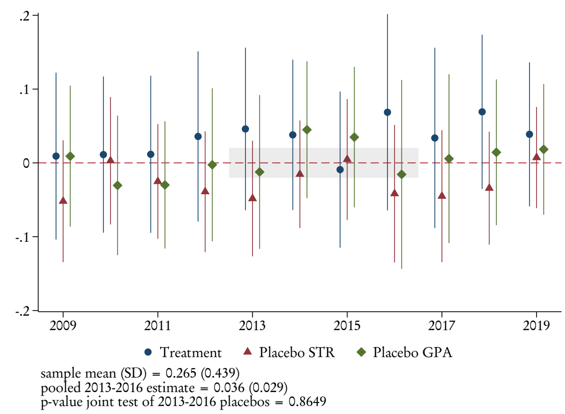
(a) Total teacher hours



(b) Regular-instruction teacher hours



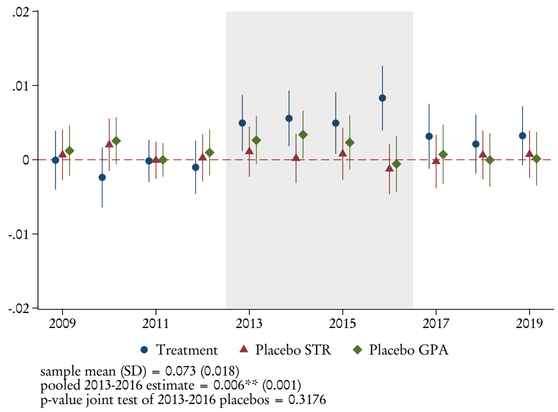
(c) Teacher hours for special needs teaching



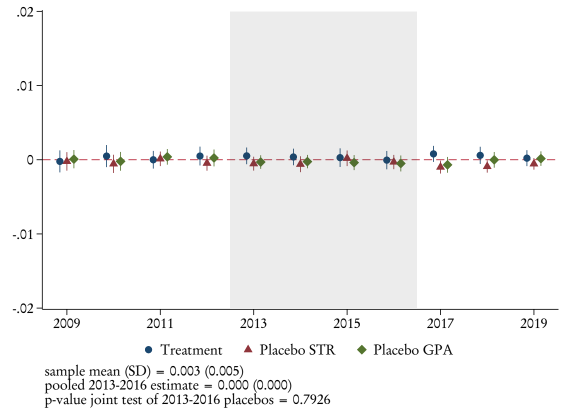
(d) Teacher hours for special needs language instruction

Figure A.5. Effects on total teacher hours and different uses of teacher hours

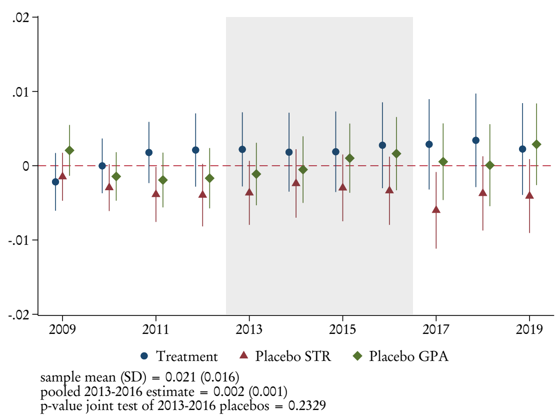
Note: The graph show estimates and confidence intervals from estimating (3). Outcome is total teacher hours and teacher hours decomposed by use, relative to total student hours. The regression uses student weights and cluster standard errors at the municipality level. See note to Figure 3 for further details.



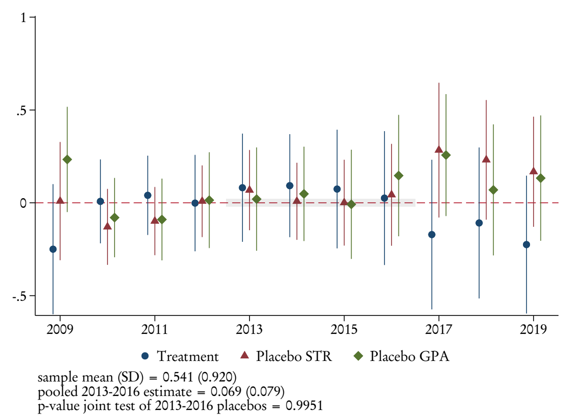
(a) Teacher hours with qualified teachers



(b) Teacher hours without qualified teachers



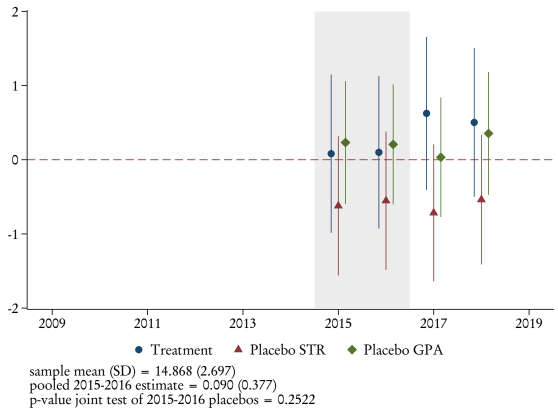
(c) Assistant hours



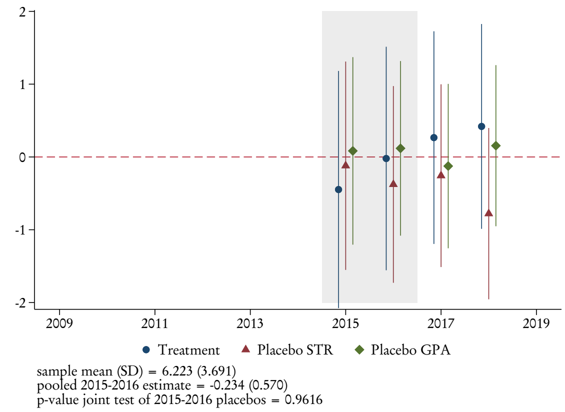
(d) Assistant hours not special

Figure A.6. Effects on teacher and assistant hours by qualifications and use

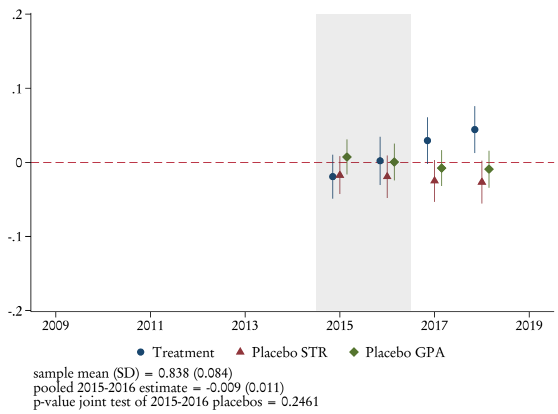
Note: The graph show estimates and confidence intervals from estimating (3). Outcome is total teacher hours and assistant hours by qualifications and use, relative to total student hours. The regression uses student weights and cluster standard errors at the municipality level. See note to Figure 3 for further details.



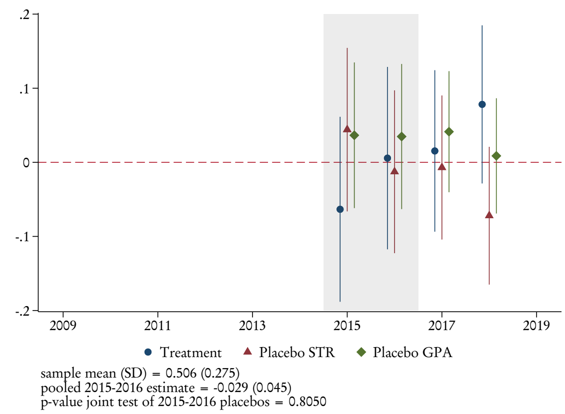
(a) Years since completed education



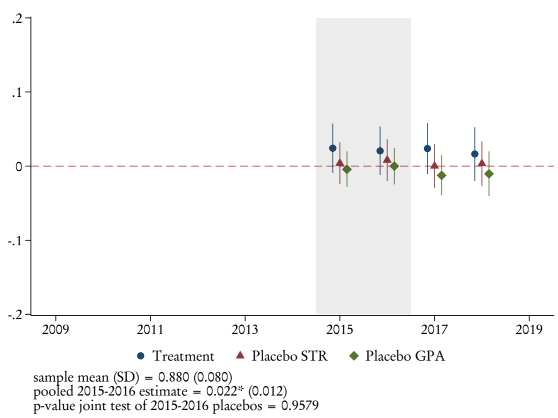
(b) Years working at same school



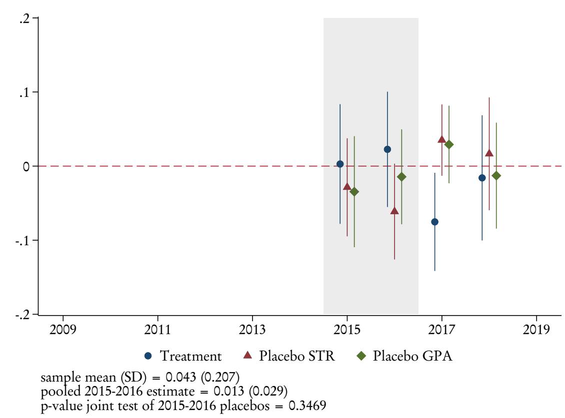
(c) Share completed education > 3 years ago



(d) Share working at school > 3 years



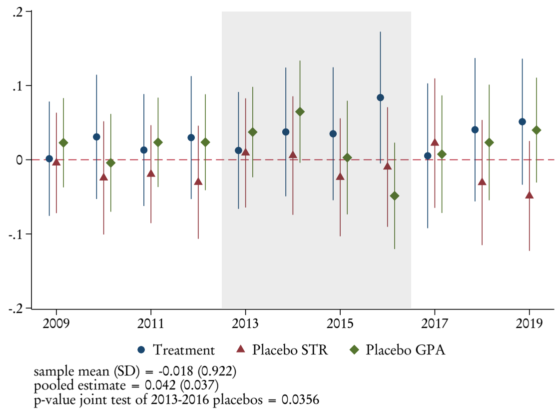
(e) Share teacher education



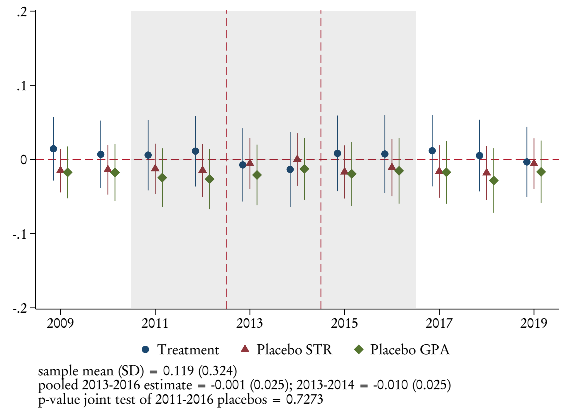
(f) Average sickness absence

Figure A.7. Effects on teachers

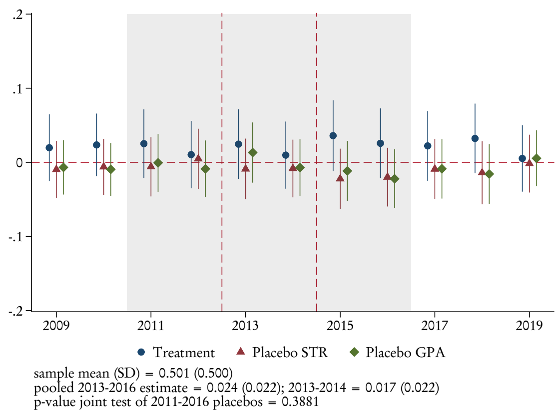
Note: The graph show estimates and confidence intervals from estimating (3). Outcomes are average teacher characteristics from matched employer-employee data. The regression uses student weights and cluster standard errors at the municipality level. See note to Figure 3 for further details.



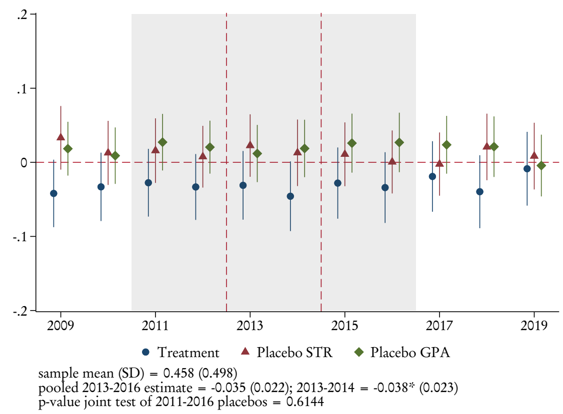
(a) 8th grade test scores



(b) Parents are immigrants



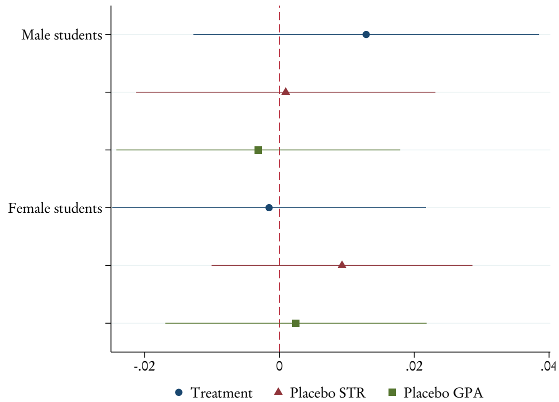
(c) Parental income



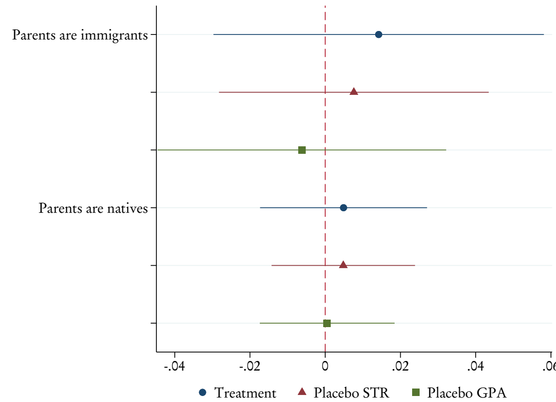
(d) Parental education

Figure A.8. Placebo effects on pre-determined student characteristics

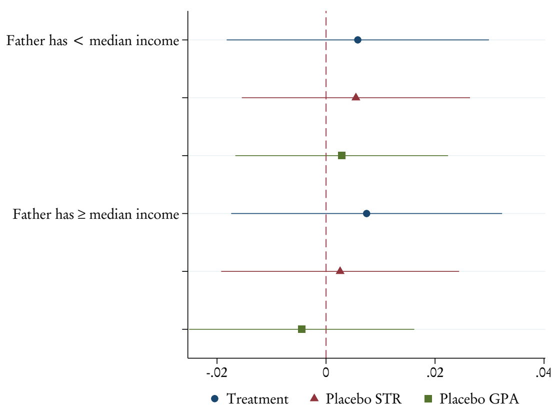
Note: The graphs show estimates and confidence intervals from estimating (3). Outcomes are pre-determined student characteristics. The different series correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The treated cohorts are shaded. The dashed vertical lines indicate delimit the cohorts treated for three years. The figure note show the sample mean and standard deviation of the outcome, estimated effect and standard errors for a pooled analysis of the treatment years and the p-value of a joint test of all placebo effects (γ) for all treatment years. The regression uses student-level data and cluster standard errors at the school level.



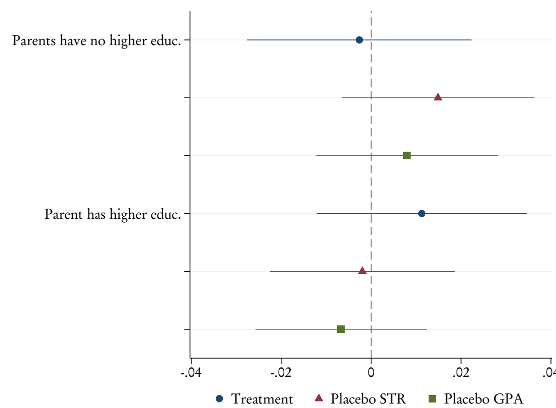
(a) By sex



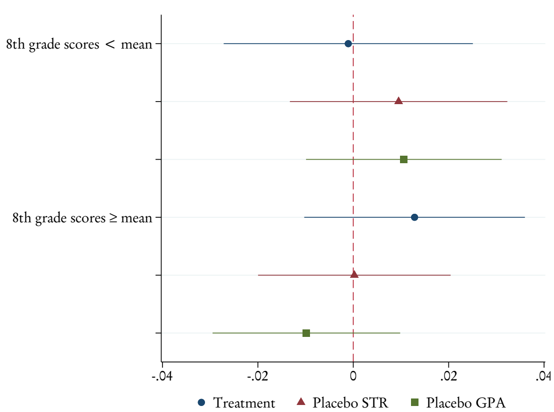
(b) By immigration background



(c) By parental income



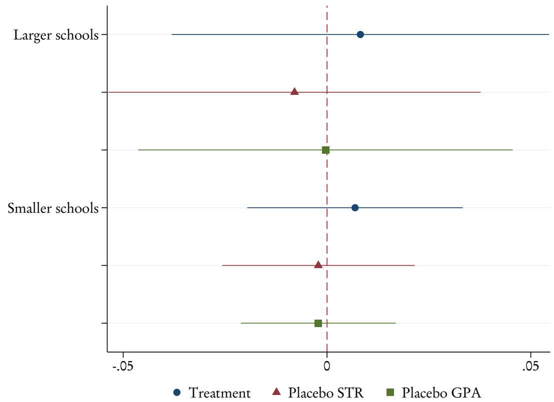
(d) By parental education



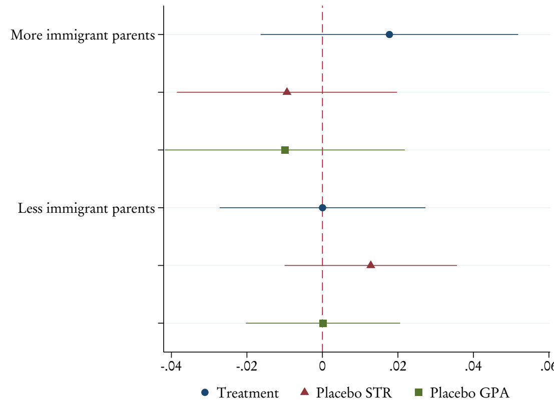
(e) By 8th grade test score

Figure A.9. Heterogeneous effects on standardized test scores

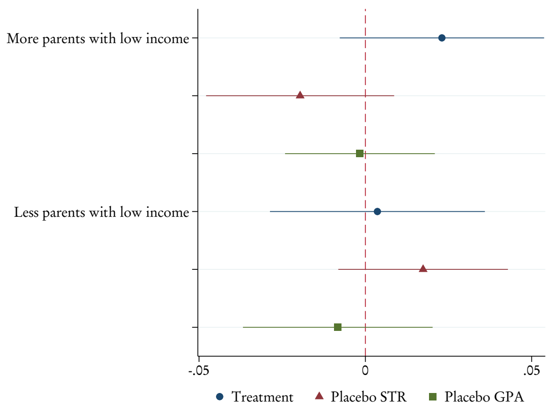
Note: The graphs show estimates and confidence intervals from estimating (3) fully interacted with a binary variable. Outcome is 9th grade test scores. The different estimates correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The regression uses student-level data and cluster standard errors at the school level.



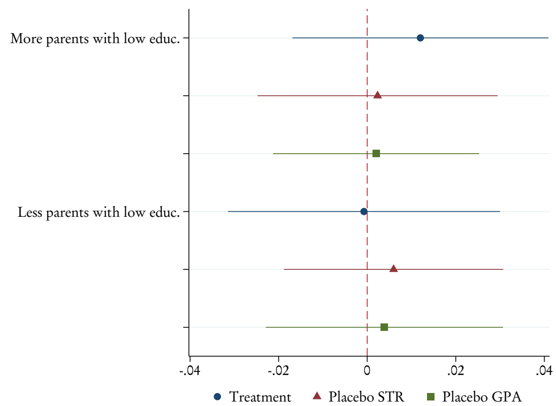
(a) School size



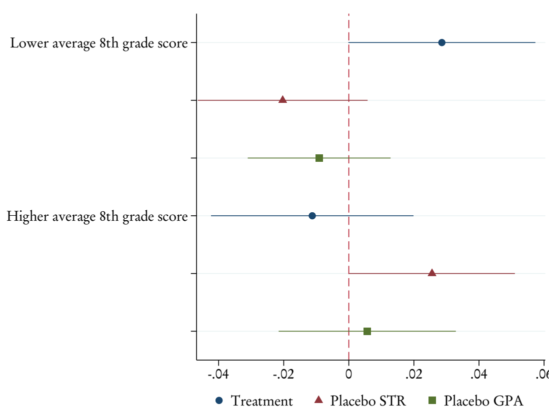
(b) By share immigration background



(c) By share low parental income

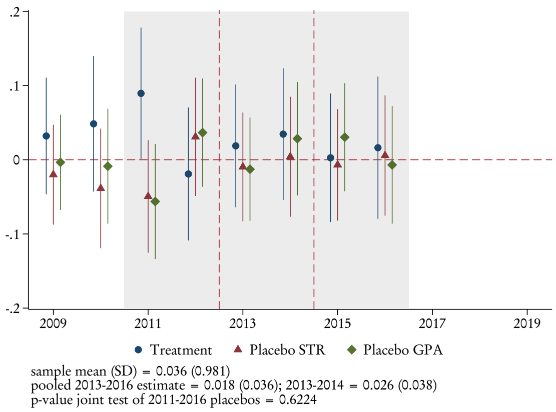


(d) By share low parental education

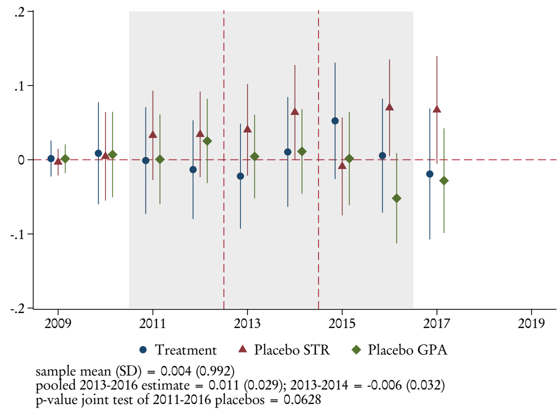


(e) By share low 8th grade test score

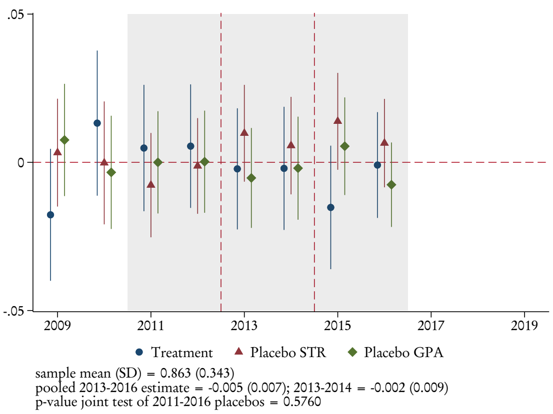
Figure A.10. Heterogeneous effects on standardized test scores, by school characteristics
Note: The graphs show estimates and confidence intervals from estimating (3) fully interacted with a binary variable. Outcome is 9th grade test scores. The different estimates correspond to treatment effects, δ in eq. (3), and placebo effects from the non-treatment margins, γ in eq. (3). The regression uses student-level data and cluster standard errors at the school level.



(a) Exam score



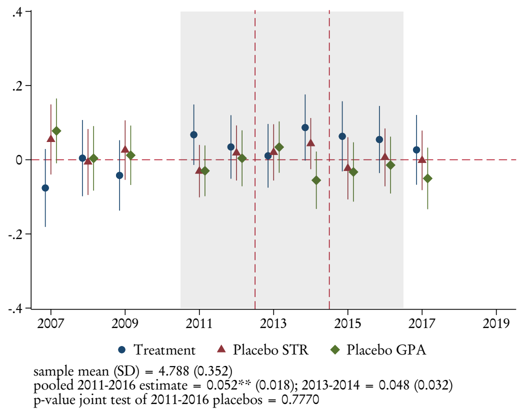
(b) Average end-of-compulsory teacher grades



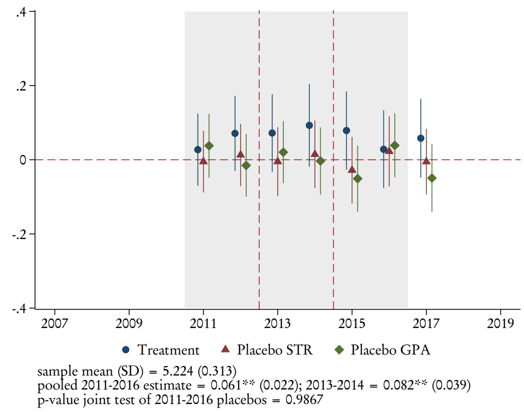
(c) Completion of first year high school

Figure A.11. Effects on exam scores, teacher grades and completion of first year high school. No controls for grade 8 test scores.

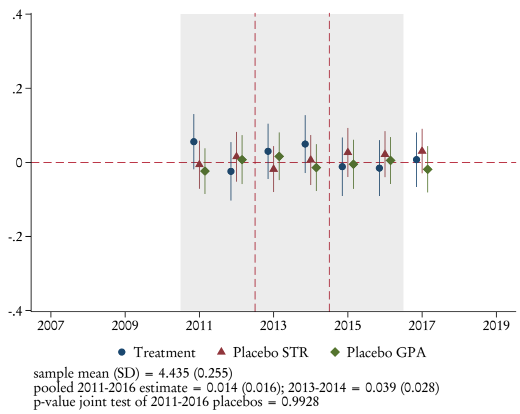
Note: The graphs show estimates and confidence intervals from estimating (3). Outcomes are end-of-compulsory school exam scores and teacher grades and completion of the first year of high school. The regression uses student-level data and cluster standard errors at the school level. See note to Figure 6 for further details.



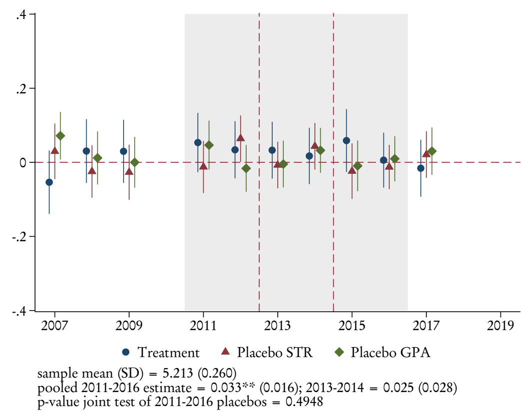
(a) Enjoy school



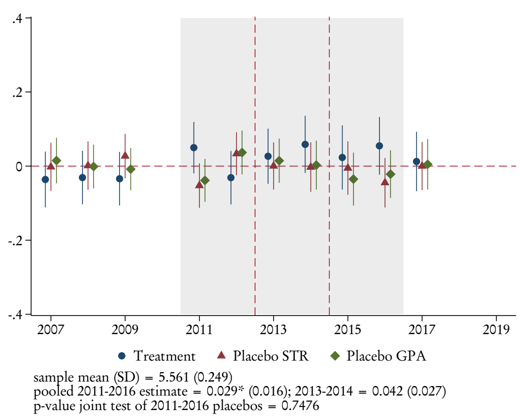
(b) Support from teachers



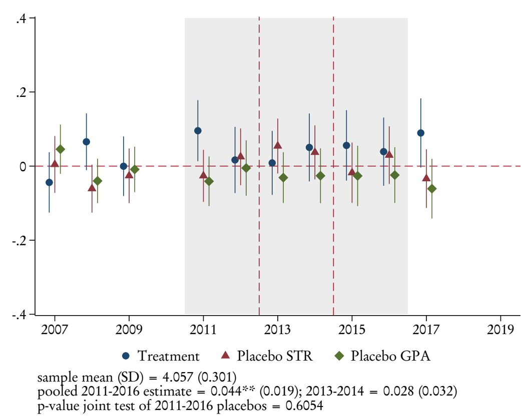
(c) Support from parents



(d) Appropriate degree of academic challenge



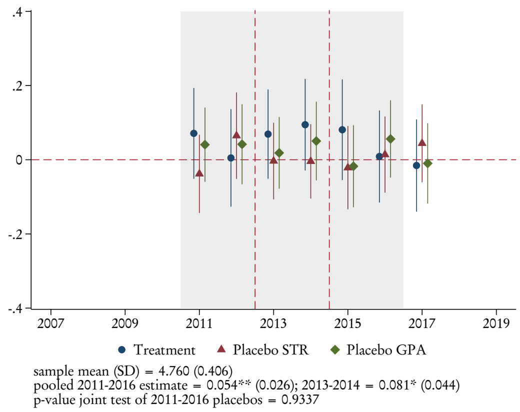
(e) Sense of mastery



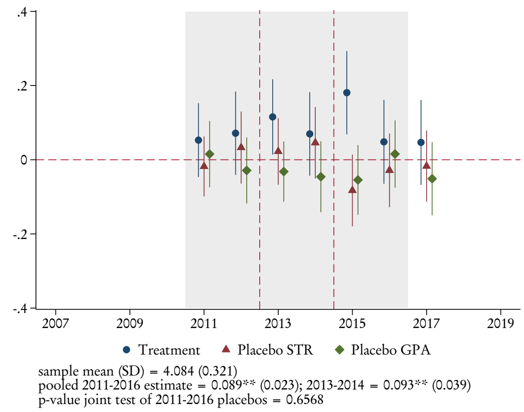
(f) Motivation

Figure A.12. School environment subindices 1

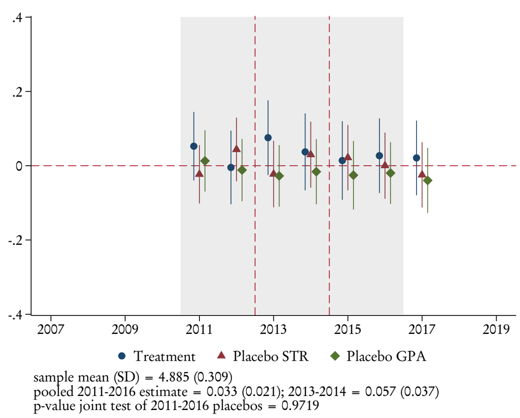
Note: The graphs show estimates and confidence intervals from estimating (3). Outcomes indices from student survey constructed by the Norwegian Directorate of Education and Training. The regression uses school-level with student weights and cluster standard errors at the municipality level. See note to Figure 8 for further details.



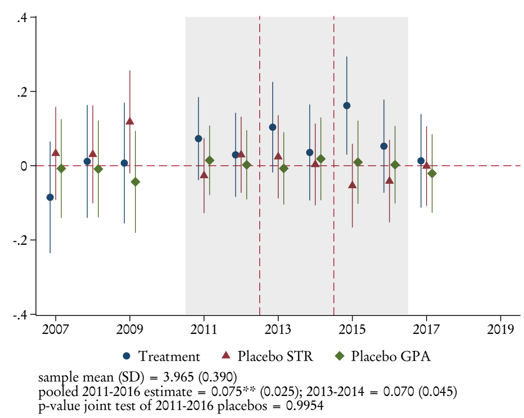
(a) Culture for learning in school



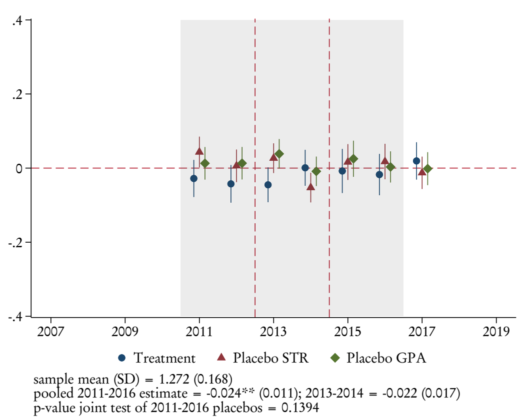
(b) Assessment that supports learning



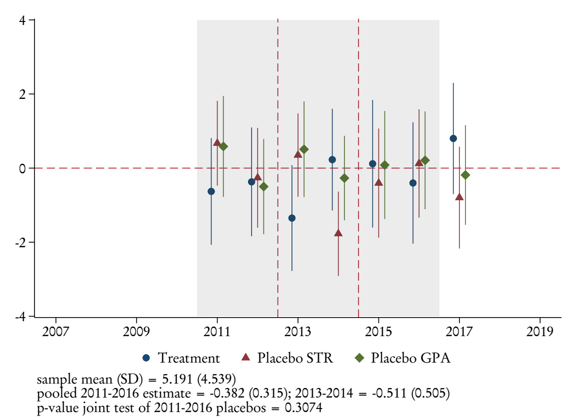
(c) Common rules



(d) Student democracy and involvement



(e) Bullying



(f) Share bullied (percent)

Figure A.13. School environment subindices 2

Note: The graphs show estimates and confidence intervals from estimating (3). Outcomes indices from student survey constructed by the Norwegian Directorate of Education and Training. The regression uses school-level with student weights and cluster standard errors at the municipality level. See note to Figure 8 for further details.