WORKING PAPERS IN ECONOMICS

No. 9/17

LEROY ANDERSLAND

PEER EFFECTS FROM A SCHOOL CHOICE REFORM

Department of Economics
UNIVERSITY OF BERGEN

Peer Effects from a School Choice Reform

Leroy Andersland[†]

This version: 30 August, 2017

Abstract

In 2005 the city of Bergen, Norway, went from a geographical catchment area high school intake system to a grade point average-based (GPA) intake system. The reform changed the composition of high school peer student characteristics substantially for comparable groups of students before and after the reform. This article compares changes in outcomes for students in Bergen before and after the reform to changes in the outcomes of students in control cities. Positive effects are found on test scores and grades at high school.

Keywords: Analysis of Education, Peer effects, Tracking, High school, Natural experiment
JEL codes: I21, J18

[†] Department of Economics, University of Bergen, 5020 Bergen, Norway, email; leroy.andersland@econ.uib.no

1 Introduction

A significant and ongoing debate in educational research and policy concerns how to place students into groups to enhance the learning environment. One important strand of the literature has studied the optimal mixture of a student group in field experiments (Booij et al. 2017; Carrell et al. 2013; Duflo et al. 2011). Do high-ability students benefit from having high-ability peers, or is it preferable to have a mixture of high- and-low ability peers? The answer to this question may vary by student age, whether cognitive or non-cognitive outcomes are studied, type of peer ability, subject, and institutional setting.

A main argument for having a system with similar students grouped together in schools is to allow the teacher to tailor content and pedagogical techniques to a homogeneous group. Another argument is that it can potentially improve student interaction. For example, high-ability students may create a culture for learning or push each other toward better achievement when grouped together. The main arguments for mixing students with different backgrounds rely on the notion that students may gain from being part of groups that include students that are diverse in various ways. High-ability students may gain from ranking in the top of their class, while low-ability students may gain from interacting with high-ability students. There are, however, many ways of grouping students, and it is possible to do so using different dimensions. In the end, the question of which system and what kind of mechanisms is most preferred are empirical.

Evidence regarding peer effects on students in their natural environments relies on finding natural experiments; the most frequent types used in the literature include housing vouchers, busing students to different schools, natural disasters, and school acceptance cutoffs. This paper adds evidence to the literature by using a school choice reform that varied peer characteristics at the school level. Specifically, we examine a reform in Bergen, Norway's second-largest city, which changed its high school intake from a catchment area

approach to a performance-based intake system. High schools located in the city center (downtown) were more attractive to students than high schools outside the city center (suburbs). When the intake system was neighborhood-based, students living downtown attended downtown high schools. Once the intake system became performance-based, most high-ability students living downtown still attended central schools, while many lower-ability downtown students were shut out of downtown schools by high-ability suburban students. In sum, the reform decreased variation within high schools, and increased variation across high schools in Bergen.

This paper compares the outcomes of students in Bergen before and after the reform to comparable students in other cities to uncover their response to changes in peer characteristics. High-ability downtown students attended high schools where the average peer student middle school GPA increased by 0.65 over the middle school GPA standard deviation (SD) from before the reform to after it, compared to similar groups in other cities. This is equivalent to going from a school at the median to a school among the top 10% in both Bergen and control cities before reform. The results show positive and significant effects on centralized, externally evaluated exams in some subjects for high-ability students as a consequence of the reform. For lower-ability students, the reform implied attending high school with less variation in peer characteristics. Consistent with recent findings from field experiments (Boiji et al. 2017, Carrell et al. 2013, Duflo 2011), our results suggest that high school performance for these students increased as a consequence of the reform. The intake reform led to a natural experiment that generated a type of tracking similar to that achieved in experiments. The reform makes it possible to identify effects on high ability students of changing peers from mixed to high ability (high-high). For low ability students it is possible to find effects of changing peers from mixed to low (low-low). The effects of this type of

tracking are relevant in cases where one decides between dividing a group based on prior ability or not.

Furthermore, the analysis suggests that students obtaining the highest grades before and two years after the reform were not necessarily comparable for two separate reasons. First, the average middle school GPA increased for cohorts completing middle school after the reform, largely because the pre-reform middle school GPA was of relatively low importance, since it had very few actual consequences for the students. After intake reform, middle school GPA became the main measure that determined which high school students would attend, so many students showed increased efforts in middle school to achieve the higher grades that would allow them to attend the school of their choice. Second, the predetermined background characteristics of the highest-scoring students at middle schools were different before and two years after the reform, which can only be interpreted as meaning that highest-scoring students are different under low- and high-stakes systems. This insight adds another layer of complexity in finding a comparable group of students in Bergen before and after reform. This analysis therefore employ a difference-in-difference-in-difference (DDD) strategy that uses the fact that high-ability middle school students in suburban areas in Bergen were exposed to the same reform, but experienced a much smaller average change in peer characteristics in high school than high-ability downtown students. The findings from this empirical strategy are weaker, but still support the hypothesis of the existence of some positive effects of the reform for this group.

The main contribution of this paper is to introduce a new type of natural experiment to identify the effect on educational outcomes of significantly changing the peer environment. While changing the characteristics of the peers, we attempt to keep school type and travel distances fixed for a particular group. In addition, this new identification strategy permits an investigation of the consequences of tracking for different groups, and opens novel

perspectives on mechanisms that are explored in detail using high-quality register data. Section 2 reviews the literature and places this paper's contributions in the context of existing scholarship, while Section 3 describes the institutional realities of high school intake in Norway. Section 4 explains how the empirical design was implemented to exploit the exogenous variation in peer characteristics created by the reform. The baseline sample was constructed from administrative records. Section 5 details the data that are important for interpreting the results. Section 6 contains the results, including a discussion of mechanisms and robustness checks, while Section 7 summarizes and concludes the paper.

2 Literature review

Manski (1993) formulated three concepts that are fundamental for the understanding of peer effects. The first is the endogenous effects—students tend to change behavior according to the behavior of the group. The second is exogenous effects, which expresses that students current behavior depends on the groups background characteristics. Finally, the correlated effects addresses the fact that student tend to select into groups based on unobserved characteristics. The reflection issue arises because of the existence of a possible multiplier in that peer behavior can affect one's own behavior, which in turn can affect the peers. Most modern papers address the identification problem stemming from endogenous peer group formation and reflection, but only a few are able to separate exogenous from endogenous effects.

Hoxby (2000) authored one of the first studies that explicitly addressed the selection issue across schools by using exogenous variation in peer characteristics within schools and across years. The results indicated positive peer effects of high-ability peers, which were stronger in a same-race context for primary schools in Texas. Hanushek et al. (2003), Betts and Zau (2004), and Lavy et al. (2012) reported similar findings. Using Norwegian data, Black et al. (2013) studied long-run outcomes such as IQ scores, teenage childbearing,

educational choices, and adult labor market status and earnings. The study found positive effects among girls of having more females in the cohort. No effect was found due to variation in average education of the mother, but average income of the father appeared to play a role in students' long-term outcomes. Other results from Norway using the same method are found in Bonesronning (2008) and Boenesronning and Haraldsvik (2014). These studies showed that school achievement was negatively affected by the presence of classmates from dissolved families and students with less-educated parents respectively.

Carrel et al. (2013) used variations in squadrons' standardized test scores as the peer variable to identify peer effects at the Unites States Air Force Academy. This analysis suggested a positive effect of peer Scholastic Aptitude Test scores on freshman GPAs among low-ability students. Building on these results, a follow-up study was conducted in which low-ability students were randomly assigned to squadrons with high-ability peers. The resulting significant negative effects for low-ability students and lack of effects on high-ability students were taken as at least partial evidence of the importance of endogenous, within-squadron peer group formation. Angrist (2014) cited this study as evidence that the standard approach of regressing outcomes on peer means, with variation mainly coming from naturally occurring variation, is not reliable.

The alternative to using naturally occurring variation is to conduct randomized experiments that manipulate the peer characteristics of individual students. As in Carrell et al. (2013), Duflo et al. (2011) manipulated peer groups in an experimental setting by streaming students into ability groups. Low-ability students were put in groups with other low-ability students, while high-ability students were put into groups with other high-ability students. The results indicated that all students benefited from tracking, including the low-ability students assigned to low-ability groups. The researchers concluded that these results show that students benefit when teachers are able to adjust their teaching approaches to a homogenous

class. Booij et al. (2017) randomized students into tracked groups using an expanded set of track combinations, so they were able to look at the effects of different combinations of group compositions and find results consistent with both Carrell et al. (2013) and Duflo et al. (2011).

2.1 Natural experiments

Many articles on peer effects in schools employs some kind of natural experiment to identify those effects. A frequent choice is policy interventions that are intended to desegregate neighborhoods or schools. Examples of this from the United States are Moving to Opportunity housing vouchers (MTO) and Metropolitan Council for Educational Opportunity (Metco). Kling et al. (2007) based their study on the fact that a lottery assignment mechanism was used to assign MTO vouchers to families; these vouchers gave families the opportunity to move to lower-poverty areas. Comparing families that were offered vouchers to those that were not, the researchers did not find any effect on adult economic self-sufficiency. They did however find beneficial mental health effects for female youth that were offset by negative health effects for male youth. Kling et al. (2005) used the same policy intervention to study criminal behavior and found similar results; female criminality went down when moving to lower-poverty areas, while the effects for males were more mixed. Ludwig et al. (2013) studied long-term outcomes and found results consistent with previous research. Chetty et al. (2016) employed newly available data on children younger than 13 at the time of random assignment. Restricting their sample to this cohort, they found significant positive effects on earnings for all groups in their mid-twenties.

Angrist and Lang (2004) analyzed the effect on test scores for students in suburban schools that received a fraction of new “Metco students” from low-income areas. Metco is a

desegregation program that sends low-income students out of poor Boston districts into schools in the surrounding suburban areas. The study did not find substantial effects on students already attending those schools. Sacerdote (2011) notes that the literature generally shows modestly positive effect on academic achievement gains, but that the effects on non-academic outcomes appear to be much larger. Sacerdote (2001), Duncan et al. (2005), DeSimone (2007), Wilson (2007), Kling et al. (2005), Kling et al. (2007), and Carrel et al. (2008) looked at outcomes such as drinking, smoking, cheating, sexual activity, criminal involvement, health, and racial attitudes. Recently, Rao (2015) studied variations in the proportion of poor children in Indian middle class schools and found that overall attitudes towards the poor became more altruistic.

Another type of natural experiment uses the regression discontinuity framework, studying students that apply to selective high schools; some are accepted and some rejected based on admission scores. Employing this strategy, Clarke (2010) found only small effects on test scores of attending selective UK schools. Jackson (2013) used this design with single-sex schools in Trinidad and Tobago, while Abdulkadiroglu et al. (2014) studied the public school systems in Boston and New York. Both show that students just above the admission cutoffs attend high schools with students that score about on average 0.5 standard deviation higher on a predetermined test than students right below the cutoff. Jackson (2013) only found effects for a group of students that had expressed strong preferences for attending selective single-sex schools and some negative effect on selecting science courses, while Abdulkadiroglu et al. (2014) found no effect of attending elite schools on students near the cutoff for admission to these schools.

One recent study used an empirical design to avoid the issues of endogenous peer groups, correlated effects, and reflection. Dahl et al. (2013) examined social interactions in program participation. Their results showed, among other things, the importance of naturally

occurring preexisting peer groups, as they were able to identify effects on siblings and coworkers on program adoption by varying the "price" of the social program. Translating this to a school setting means that even though a study is able to control for endogenous sorting across schools or classes, the endogenous sorting within schools or classes may be an important point of focus for studies analyzing school-situated peer effects. Other analyses using the partial population approach are found in studies of the PROGRESA program in Mexico. The PROGRESA program provided cash incentives for parents to send their children to school. Peer effects can then be identified on ineligible children that are in the naturally occurring peer groups of eligible children. Angelucci et al (2009), Bobonis and Finnan (2009), and Lalive and Cattaneo (2009) all found substantial positive peer effects on school attendance. To the best of our knowledge, however, no studies have used this method to identify peer effects on academic achievement.

Empirical designs using natural experiments are often unable to separate neighborhood or school effects from peer effects. There may be other differences between high- and low-poverty neighborhoods than resident incomes, and there are other differences between elite schools and other schools than their students. Designs of the type found in Angrist and Lang (2004) explicitly address this by focusing on the effects on students who were already attending schools that experienced a change in student composition. Besides employing a new type of natural experiment, a key contribution of the present study is that it keeps variables such as school type and travel distance fixed, while varying average peer characteristics substantially. The combination of these two features is not often found in the scholarly literature. In addition, the reform that changed the high school intake from a geographical catchment area based system to a GPA based intake system is similar to switching from ability mixing to tracking. Therefore, we contribute with a natural experiment that allows us

to explore the effects of ability tracking at high school for both low and high achieving students.

3 Institutional setting

Children in Norway start school in August of the year in which they reach six years of age. Children normally attend primary school until age 14 and middle school from age 14 to age 16. Most primary and middle schools are public, with intake based on geographical catchment areas. When students finish middle school, most students choose between applying for academic or vocational tracks at high school. Around half of students choose to start on the academic track, with about 75% of that group graduating within three years.

3.1 High school intake

In 2005, 95% of high school students in Norway attended public high schools. High school intake systems are regulated at the county level; private high schools have separate mechanisms to accept students. The approaches adopted can be divided into middle school GPA-based intake systems, geographical catchment area-based intake systems, and combinations of the two models (Brugård 2013). Bergen, where the reform examined in the present study occurred, is located in Hordaland County. Before the school year starting in 2005, Bergen had a system by which most students completing middle school were assigned a high school by the county school administrative office, which was guided by rules that obliged them to divide all middle school students into GPA groups, and then divide these students among high schools so that each high school had a roughly equal proportion of students from each GPA group.[1] One reason for this approach was to try to avoid the development of "good" and "bad" schools. In practice, students were generally assigned to

[1] Source: Nils Skarvhellen, Head of Intake office at Hordaland Fylkeskommune. (20.10.2015), Knudsen, Sortevik and Woldset, Government proposal analysis (2003)

schools was close to their homes to reduce travel time; there was some flexibility on the administrative office's part to deal with students with strong desires to attend a certain high school. The Hordaland school administrative office noted that this system required significant effort on their part. The demands of the intake system, combined with increasing pressure from different interest groups, were the reasons that Hordaland County changed its intake system in the school year beginning in August 2005 to a middle school GPA-based intake system. The county government passed the rule change in October 2004. In the next high school intake students could list schools based on preference, and were accepted to their first choice if their middle school GPA was above that school's admission level. Each school had a limited number of seats, so the admissions level varied depending on the number and middle school GPA of the applicants. Only the central school authority at the county level is involved in the acceptance procedure and not the schools.

The control municipalities used for empirical comparison purposes are Trondheim, Stavanger, Drammen, and Kristiansand, four of the five other largest cities in Norway. Oslo, the capital and largest city, is not included among the control cities because a separate school choice reform took place at the same time as in Bergen.[2] Drammen had a catchment area-based intake system, while Trondheim, Stavanger, and Kristiansand all based intake on middle school GPA.

3.2 Curriculum and grading

Learning structure and course compositions at schools are regulated at the national level in Norway. This means that all students who attend a public school have access to approximately the same range of courses and attend schools with similar learning principles

[2] We chose to focus on the reform that happened in Bergen since the reform there was a total transition, while the reform in Oslo was only a partial change in intake systems.

and goals. For the school year 2006–2007 a reform in learning principles took place, called "Kunnskapsløftet". The most important changes at the high school level were changes in course composition; students were not differentially exposed to the reform within cohorts, only across cohorts.

4 Empirical design

A school intake reform may significantly change the composition of students at high schools. With a catchment area intake system, students generally attend the geographically closest high school. With a performance-based intake system, however, high schools consist of students who apply and are accepted to each high school based on middle school performance. The degree of change in the composition of students after an intake reform will depend on the attractiveness of the high school. A change from a catchment area to a performance-based system will lead to a negative selection of students at less attractive schools, since high-ability students will have the option to leave, which most low-ability students will not. The same change will also lead to a positive selection into attractive schools, since high-ability students from outside the catchment area will be chosen over low-ability students from within the catchment area.

Comparable students before and after an intake reform may end up attending a high school with very different peers. In this paper we use these changes to analyze peer effects. The first step is to find comparable students before and after the reform for which the reform's main effect was changing the characteristics of their high school peers. Besides peer characteristics, a school intake reform can change both daily travel distance and the type of high school for comparable students. The group for which reform is most likely to mainly change peer characteristics are high-ability students living in the catchment areas of attractive schools; they would have attended those schools before reform, and because they still qualify

and there is no obvious reason for them to apply to less attractive schools, they are likely to continue attending those same schools. We first chose to focus on high-ability students living in attractive schools' catchment areas, since we expect that this group should experience a significant change in peer characteristics without any changes in distance traveled and type of school.

High-ability students belonging in downtown middle school districts in Bergen before the reform attended high schools located in downtown Bergen; most students average high school peer ability as measured by middle school GPA was close to the average of Bergen as a whole. After the transition to a GPA-based system, they still attended those same downtown high schools, but now many of their low-ability middle school peers were replaced by high-ability students from the suburbs. To identify the effect of these changed peer characteristics, we compared the change in outcomes of high-ability downtown students in Bergen to the change in outcomes of high-ability downtown students in other large cities in Norway in a difference-in-difference (DD) setup. Model 1 is defined as:

$$Y_{it} = \beta_0 + \beta_1 BERGEN_{it} + \beta_2 REFORM_{it} + \beta_3 (BERGEN_{it} * REFORM_{it}) + \beta_4 X_{it} + \varepsilon_{it} \quad (1)$$

Where i denotes individual student, t denotes cohort, Y_{it} are outcomes that can be affected by changed peer characteristics, $BERGEN_{it}$ is a dummy variable (1 for high-ability student living downtown, 0 if the student has high ability and lives in another city center). High-ability students are defined as those having middle school GPAs in the top 25% of their citywide cohort, while downtown students are those who attended middle school in the downtown area of the city.[3] Middle school attendance is almost exclusively determined by a middle school level geographical catchment area. $REFORM_{it}$ is a dummy indicating 1 for the cohorts

[3] We vary the GPA threshold in the robustness section.

applying to high school after Bergen's reform (born from 1989–1991), and 0 for cohorts applying for high school before it (born 1986–1988). X_{it} is a vector of individual-level control variables and middle school dummies. Individual level controls are parents' earnings, parents' years of education, and gender.

[FIGURE 1]

Figure 1 shows the development of average incoming peer GPA at high schools for each cohort for high-ability students in downtown Bergen and control cities. Average peer GPA remains stable at around 4.4–4.6 for both groups before reform. Peer GPA is slightly higher in the control groups, something that could be explained by the fact that three of the four control cities had a GPA-based intake system in the period studied. The reform was implemented for the cohort born in 1989, and we saw a sharp increase in peer GPA in the treatment group for this cohort, which stabilized at a higher level for the subsequent two cohorts. There was no change for the control cohorts. Peer students' middle school GPA increased by 0.65 of one SD of middle school GPA from before to after reform, compared to comparable groups in other cities. This is equivalent to going from a school at the median to a school among the top 10% in both Bergen and the control cities before reform.

5 Data and variable definition

Data were taken from Norwegian administrative records. Middle school grade information is available through a centralized middle school database, while information on middle school and high school attendance is available through education records detailing the schools and tracks that individuals attend and complete. High school grade information is available from two sources, the school administrative grade records (a database with grade information

collected from the various school administrative systems) and the Norwegian certificate database, which contains high school grade information for all certificates granted. Both databases are used in this analysis because each has strengths and weaknesses. The certificate database contains only grade information on those who complete their certificates at high school, so dropouts' grades are not present. The certificates database however has grade information for more cohorts than the administrative database.

The baseline sample is students who started high school immediately after middle school.[4] Data on school absence is available from the certificates database for those who completed school. This measure comes from teachers' recording the number of hours and days a student was absent from class during the school year.

The data allow us to link students and parents, so we can use parents' years of education and yearly earnings as control variables; these are both measured when the students are 10 years old. Earnings are measured in 1996 NOK.

[TABLE 1]

Table 1 offers descriptive statistics of the sample; the baseline sample consists of high-ability downtown students in Bergen and similar students in Kristiansand, Stavanger, Trondheim, and Drammen. The table is divided into three panels: panel a) show descriptive statistics of covariates, panel b) shows the peer GPA variable, and panel c) shows descriptive statistics of high school outcomes. Column (1) shows the pre-reform means of the treated group, Column (2) shows the SD of that group. Columns (3) and (4) show the difference in means between treatment and control before and after the reform. Column (5) shows the number of observations of each variable. The top 25% downtown students in Bergen, Kristiansand,

[4] We have verified that the reform did not affect applications and intake to academic tracks.

Stavanger, Trondheim, and Drammen give a total baseline sample of 1869 students. Missing observations on covariates are dealt with by replacing them with the value 0 and including a dummy for the missing observation of the covariate.

From the descriptive statistics of covariates we note that students belonging to the treatment group have a lower proportion of females and parents with higher earnings and more education. These differences are somewhat smaller after reform. Peergpa is constructed from the average incoming middle school GPA of the peers of the high school students. We note that before the reform, high-ability downtown students in Bergen on average attended schools with lower-ability peers than in other cities. This is due to the intake system in Bergen pre-reform not being performance-based, while three of the four cities in the control cities did have performance-based intake systems. Comparing the difference between columns (3) and (4) in panel b) shows that the treatment group increased their average peer GPA at high school by 0.41 compared to the control group.

Firstyear GPA is the average grade for the first year of high school, while "High School GPA" is the average of all grades in high school. These two measures largely contain grades assigned locally by the teacher. Absence days and Absence hours are the number of days and hours of recorded absence during all years at high school. Absence days are the number of full days that a student was not recorded as present in any class at school. Absence hours are the number of hours of recorded absences from class, not including full-day absences. "Select basic math year 1" indicates whether the student selected the less advanced math course in the first year of high school.

Norwegian exam in year 3 is a compulsory national exam at the end of high school that is externally administered and graded. Norwegian II exam year 3 is the second formal written language for the student. Students decide themselves which written language is their

main language. For more than 90 % of the students, the Norwegian exam is "Bokmål", while the Norwegian II exam is "Nynorsk".

Exams in years 1, 2, and 3 make use of the fact that students are randomly drawn to take exams in different subjects. Scores in different subjects are pooled for each student by year since the number of students drawn for each subject is relatively small. Only about 30% of students are drawn to take an exam in year 1 in any subject, which explains the smaller sample size. The advantage of the exams in year 1 is that the exam-takers are randomly drawn among mandatory subjects, which means that that the coefficient are not inadvertently capturing mechanisms that involve a change in course composition. A larger proportion of the students take a standardized written exam in years 2 and 3, which are drawn among electives.

6 Results

The first results using Model 1 are shown in columns (1)–(3) in Table 2. The table focuses on the high-ability downtown students who experienced a large increase in peer ability. Column (1) presents the results without any controls, while Column (2) add middle school dummies. Column (3) shows the preferred specifications were controls for background characteristics and middle school dummies are included. The focus of the discussion will be on the empirical specification including middle school dummies and background characteristics. Each row gives the estimate of β_3 from Model 1 with the dependent variable indicated in the row header. For now, β_3 is interpreted as the average treatment effect on the treated (ATT) of the intake reform on high-ability downtown students. Section 7 goes into detail about what may explain the findings.

[TABLE 2]

A positive coefficient is found on both Firstyear GPA and High school GPA, but the effect is only significant on average grades in the first year of high school after including middle school dummies. After including both middle school dummies and background characteristics, the coefficient on Firstyear GPA is 0.10. There is a negative though not significant coefficient for total hours absent from high school and a positive insignificant coefficient for total days absent. Not finding any effect on absence is in consistent with travel time being unaffected by the reform for these students. The next row shows a non-significant increase in the likelihood of selecting a basic as opposed to an advanced math course during the first year of high school. Thus, the results does not give any conclusive evidence of whether high-ability peers encourage more advance course taking, or if it makes it more difficult to get a seat at a limited number of spots at these courses.

The effect on Norwegian exam in year 3 shown in Column (3) are statistically significant at the 10% significance level. The size of the effect is 0.20 and stable across different specifications. The results show that Nynorsk also increased by 0.30. This effect is significant at the 5% level. Even though there are fewer observations of Exam year 1, a statistically significant positive effect at the 1% significance level is found on this measure. The size of the coefficient is 0.48, and is the largest effect on the achievement measures shown in Table 2. In sum, together with the effect on GPA, the effect on national exams suggests the reform's positive effect for the high-ability downtown students. The effects on Exam year 2 and 3 are not significant in any of the specification. A possible explanation is that these exams are drawn among elective courses, which may be affected by the reform, making this exam measure more sensitive to mechanisms that cause students to change course composition at high school.

[FIGURE 2]

Figure 2 plots average outcomes by cohort, allowing for a graphical inspection of how outcomes change over time. Solid circles and triangles indicate averages, while 95% confidence intervals of means are indicated with crosses. Confidence intervals are tighter for the control group since that group is larger. Norwegian exam scores and GPAs averages move relatively coherent before reform, with a trend shift for the treatment group at the time of the reform. This supports the assumption that if the reform had not happened, the two groups would have had the same development in outcomes.

6.1 Reform effects on all groups in Bergen

Table 3 show subsample estimates of β_3 in Model 1 on all groups of students in Bergen. All estimations are performed with middle school dummies and background control variables. Columns (1)–(4) show estimates for the low-, medium-low-, medium-high-, and high-ability downtown students, while columns (5)–(8) show these results for suburban students across the same achievement groups.[5] The first row shows that higher-ability students received higher-ability peers after reform, and that this effect was strongest for downtown students. For low- and medium low-ability students peergpa increased much weaker or not significantly at all. Given the large increase in peergpa for high-ability students, a larger fall in peergpa for other groups of students could be expected. The reason for this is that average middle school GPA in Bergen increased after the reform. Further discussion of the consequences of this fact

[5] Students are split into equal sized groups within their cohort and city based on where they ranked on the middle school GPA distribution.

appears in section 7.1. Another consequence that can be inferred from the first row is that the variation in student characteristics within schools decreases.

[TABLE 3]

The second and third rows show clear positive and significant effects on average high school grades for all groups of students, except for high-ability students. These results show that the intake reform benefited these other groups of students. There is a negative coefficient on hours absent significant at the 5% level for downtown students with medium-low ability, while there is a positive significant coefficient at the 10 % level on days absent for low-ability suburban students. Thus, the effects on absence are inconsistent.

The effect on the centralized exams in the third year is most pronounced for high-ability downtown students. The positive effects on GPA that appeared for the other groups of students are not found to the same degree on exam scores. One explanation for this is that, as measures of academic achievement, the exams are subject to more noise. Alternatively, the school grading captures improvements in abilities that are not measured in exams. An example of this is classroom behavior.

Regarding the effect on the pooled exam score measures, a larger positive effect on first-year exams for high-ability downtown students are found than for most of the other groups. No significant effects are found on second-year exams, while on third-year exams a significant negative effect for high-ability suburban students and a significant positive effect for low-ability suburban students. As noted above, the first-year exam involves fewer students, while second- and third year exams are vulnerable to potential mechanisms that cause students to change course selection because of the reform. This may explain why the

findings for these measures are less coherent. Average outcomes by group over cohort can be visually inspected in Figures A.2 and A.3 in the Appendix.

6.2 Placebo and Robustness

β_3 identifies the ATT effect of the reform on downtown students ranked in the top 25% of their cohort in Bergen, with the assumption that without the reform, they would have had the same trend in outcomes as downtown students in control cities ranked in the top 25% of their cohort and city. To determine if that would have been the case, trends before the reform are examined. This is possible since we observe outcomes for three cohorts of students before reform.

[TABLE 4]

Figure 2 allows for visual inspection, but for a formal test Model 1 is estimated with the adjustment of keeping only the three cohorts before reform and defining two placebo reforms starting in school years 2003–2004 and 2004–2005. Results are shown in columns (1) and (2) of Table 4. Only two of the 22 tests gave a significant coefficient at the 10% significance level. This does not provide strong evidence against the common trend assumption, though one weakness of this test is that the lower sample size offers less precision.

[TABLE 5]

Table 5 show how effects change when the estimation sample or model specifications change. Column (1) gives the baseline estimates reported in the main results. Column (2)

shows estimates in a sample containing downtown students scoring in the top 33% of their city cohort. Coefficients on Norwegian and Norwegian II exams remained positive, but they were no longer significant. One possible explanation is that it was students with the highest ability that gained the most from a change in peers. Column (3) includes high-ability students from more cities than those included in the baseline sample. Both the effects on grades and test scores became less noticeable, though this could be because this sample definition are comparing trends in groups that were less equal than the groups compared earlier. Columns (4)–(6) explore how the estimates are sensitive to adjusting the cohorts included in the sample. Standard errors on effects increased when reducing the sample size to include only cohorts closer to reform, while the effect on exams stay significant.

To correct for possible intragroup correlation in error terms, standard errors were clustered at the high school*year level. Table A.1 in the Appendix shows that the results are somewhat sensitive to the level of clustering. Standard errors on test scores decreased when clustering on middle school or city. One possible reason is few clusters; there were 25 middle schools, while there were five cities.

[FIGURE 3]

The last robustness check performed is based on the permutation method proposed in Buchmueller, DiNardo, and Valleta (2011). We have assumed that the policy change happened in each of the 20 largest cities in Norway (excluding Bergen and Oslo), and estimated DD coefficients for the top 25% students in each of these cities and for the other three ability groups in each city. Figure 3 shows the distribution of these coefficients and the 95th percentile in the distribution. By comparing the coefficient for Bergen to the empirical distribution of DD coefficients for the other groups, we rejected or kept the null hypothesis of

no effect in Bergen for each outcome. The results from this robustness check show that none of the coefficients are above the 95th percentile in the distribution of coefficients. A way to increase precision is to pool the two Norwegian exam scores and inspect inference is for this measure. Pooling Norwegian scores results in nearly a doubling of the sample size, and results are significant at the 1% significance level with robust standard errors, and above the 95ht percentile in the distribution created with the permutation test. Bergen is the second largest city in Norway. Choosing fewer and larger cities decrease the variation in the distribution of coefficients. In total, results from the permutation test do not suggest intragroup correlation in error terms lead to too small standard errors.

7 Mechanism

7.1 Testing for selection and "the incentivizing effect"

The 25% best students from downtown Bergen districts and the 25% best students from downtown districts in the control cities are student-group categories that students could switch in or out of because of reform. The ATT of a school choice reform on high-ability downtown students would be biased if the top 25% downtown students were different under a catchment area system and a GPA-based system, as for example if students changed their catchment area to one with their preferred school as a response to the introduction of a catchment area system. This could prevent the treated group from being comparable before and after reform. Machin and Salvanes (2010) showed that house prices in Oslo remained sensitive to school intake reform that took place in 1997, even a decade later. This may be less of an issue in Bergen since the system before the reform was not strictly based on catchment area.

A similar situation would arise if students ranked in the top 25% of their cohort were not the same before and after reform. Haraldsvik (2014) studied the effect on middle school

grades of students in Bergen as a consequence of reform. The study revealed that those grades increased in the district as a whole with the transition from the catchment system to the GPA system. Haraldsvik proposed that a performance-based system incentivized some or all students to work for better grades in order to increase their chances of attending their preferred high school.

The reform was announced in the fall the year before it was implemented. For the first cohort applying to high school after the reform, the adjustment time was less than a school year. The first cohort after reform should also have been less incentivized to increase their grades, since there were fewer observable differences between high schools. The second and third cohorts had more time to adjust to intake reform, and the differences between schools would have been more evident.

[TABLE 6]

One way to test whether there was selection of students into the treated group due to reform is examining changes in predetermined background variables, and if students were incentivized by the reform, it would be revealed by grades determined before high school. To test for selection and the incentivizing effect, Model 2, a modified version of Model, 1 is implemented:

$$Y_{it} = \alpha_0 + \alpha_1 BERGEN_{it} + \alpha_2 Cohort1989_{it} + \alpha_3 Cohort1990_{it} + \alpha_4 Cohort1991_{it} + \alpha_5(Cohort1989_{it} * BERGEN_{it}) + \alpha_6(Cohort1990_{it} * BERGEN_{it}) + \alpha_7(Cohort1991_{it} * BERGEN_{it}) + \alpha_8 X_{it} + \epsilon_{it} \quad (2)$$

The main change from Model 1 is that the Bergen indicator now is interacted with indicators for each post-reform cohort. The dummies $Cohort1989_{it}$, $Cohort1990_{it}$, and $Cohort1991_{it}$ are the indicators for the three post-reform cohorts.

Table 6 show estimates of α_5, α_6, and α_7 in Model 2. Panel a) shows results when the dependent variables are background characteristics and middle school GPA. Panel b) show results when dependent variable are high school peergpa and measures of academic performance in high school. There are generally smaller differential changes in predetermined background characteristics for the first cohort than for the rest. The second cohort shows larger differences, while for the third cohort there were negative significant coefficients on fathers' and mothers' years of education. Coefficients on parents' income were negative but insignificant, while a positive insignificant coefficient appeared for the female dummy. These results confirm a hypothesis of dynamic response to school choice reform. Students scoring in the top 25% of their cohort in the city center of Bergen were different before the reform and two years after the reform.

Table 6 shows that middle school GPA increased. This finding could be explained by the "incentivizing effect" that the school choice reform had on student middle school grades. The top 25% of downtown Bergen students had higher middle school grades after the reform, a finding that is in line with Haraldsvik (2014). The second main explanation for the ATT of the reform is therefore that high-ability students in Bergen became better because of the reform before entering high school.

Panel b) focuses on high school academic outcomes with the new specification. It indicates that selection into the top 25% of downtown students in Bergen affected the high school outcomes of this group. The results in Table 2 show that including background variables did not change the results significantly. However, if the significant coefficients in panel a) indicated changes in unobservable factors, that would suggest that the DD coefficient

is a lower bound on the effect of the reform. The incentivizing effect indicated by the positive effect on middle school GPA is not an issue if it was a transitory shock to abilities that do not affect outcomes in high school. If it was not a transitory shock in abilities, then part of the observed effect of the reform could be explained by this phenomenon. The next section specifies a model that is designed to take into account both selection into the group of the top 25% of students in Bergen and the incentivizing effect due to school choice reform.

7.2 Accounting for selection and direct effects

A school choice reform could affect the high school outcomes of the top 25% of downtown Bergen students through channels other than a change in peer characteristics in high school. The ATT would then not only reflect a peer effect but also these alternative mechanisms. The first of the two main alternative explanations indicated in the last section is that the top 25% of students in Bergen were not the same before and after reform. The second is that the top 25% of downtown Bergen students had higher ability after the reform because they studied harder at middle school in order to be accepted into selective high schools.

[TABLE 7]

One way to separate the peer effect from these explanations is employing the fact that all students in Bergen underwent the school choice reform, but not all of them experienced the same change in high school peer characteristics. As shown in Column (8) in Table 3, the top 25% of suburban students did not experience the same change in peer characteristics, even though they were equally subject to the reform. The procedure would then compare changes in test scores between downtown students in Bergen and in control cities to changes in test scores between suburban students in Bergen and in control cities. Changes in group

composition and the incentivizing effect would no longer pose a concern if these effects were equal in the downtown and suburban areas. To implement this strategy, the following specification (Model 3) is estimated:

$$Y_{it} = \gamma_0 + \gamma_1 BERGEN_{it} + \gamma_2 REFORM_{it} + \gamma_3 DOWNTOWN_{it} + \gamma_4(BERGEN_{it} * REFORM_{it}) + \gamma_5(BERGEN_{it} * DOWNTOWN_{it}) + \gamma_6(REFORM_{it} * DOWNTOWN_{it}) + \gamma_7(BERGEN_{it} * REFORM_{it} * DOWNTOWN_{it}) + \gamma_8 \mathrm{X}_{it} + \varepsilon_{it} \quad (3)$$

Columns (1) in Table 7 show estimates of γ_7 in Model 3. First we note that since both downtown and suburban students experienced an increase in average peergpa, the relative effect on peergpa goes down. The coefficient of GPA measures are about the same, but significance disappear for Firstyear GPA. The coefficient for Norwegian exam year 3 is 0.16 and insignificant, while the coefficient with controls on Norwegian II is 0.34 and significant at the 5 % significance level. Significance disappears for Exam year 1, while the coefficient change sign for Exam year 3. The disappearance of significance for some of the outcomes could be explained as a direct effect of the incentivizing effect of the reform. Alternatively, the disappearance could be explained by the lower relative increase in peergpa using this strategy. The coefficient for Norwegian II suggests that high-ability downtown students still gain from the reform relative to high ability suburban students.

[FIGURE 4]

To test for whether the DDD strategy accounts for the selection and the incentivizing effect, Model 3 is estimated with background characteristics and middle school GPA as dependent variables. Table A.3 in the Appendix shows that the significant effect from middle school

GPA disappears, while there is still a negative effect on maternal years of education. This means that we cannot reject a null hypothesis that there was no incentivizing effect, while there may still be some unaccounted-for selection in the model using the DDD identification strategy.

7.3 School effects

The reform in Bergen allowed students to choose which school to attend. For at least 10 years before reform, a catchment area design in which students' ability as measured by middle school GPA was used to distribute students across schools. This could indicate that at the time of reform the schools were relatively similar, since their classes had for a long time consisted of similar student.

[TABLE 8]

There is, however, a systematic pattern to the change in which schools different types of students attended before and after reform. Table 8 reveals some of the changes in composition in downtown high schools. Columns (1)–(5) show the proportion of high-ability downtown students at each downtown school in Bergen before and after reform. Column (6) reports the proportion at any downtown school, while Column (7) report the proportion of high-ability downtown students at private schools. The p-value of a two-proportion z-test for differences in proportions is reported in the last row. The table shows that high-ability downtown students moved between downtown high schools because of reform; specifically, they moved from Tanks and Bjørgvin to Katten and BHG. The table also shows that the reform did not influence the decision of high-ability downtown students of whether to attend downtown public schools or private schools.

All downtown high schools are not equally attractive to students. Since the reform induced more high-ability downtown students to attend a particular high school, the ATT could reflect a change in school quality, rather than simply peer effects. The school effects observed cannot explain all of the effect identified. The proportion of high ability downtown students at BHG and Katten increased by 18 percentage points as shown in Table 8. Even if the effect of attending BHG or Katten is large, a potential school effect can only explain a small proportion the identified effect of the reform.

8 Summary and Conclusion

There are many studies in the peer effect literature that rely on naturally occurring variation in peer characteristics to estimate peer effects. Commonly used natural experiments are school vouchers, desegregation schemes, or school assignment lotteries. This study used a school choice reform process in Bergen, Norway to investigate the effects of changes in peer characteristics at high schools for high-ability students.

A change from a catchment area-based intake system to a performance-based intake system, or the reverse, will have different consequences for different types of students. This study focused on a group—the high-ability downtown students in Bergen—for which the reform primarily resulted in a considerable shock in peer characteristics at high school. The ATT of the reform was identified by comparing the change in high school outcomes of this group of students before and after reform to comparable students in other cities. The analysis showed that this group of students attended high schools where peer students' average middle school GPA increased by 0.65 of one SD after reform, as against comparable groups in other cities.

The results showed that exam scores of downtown high-ability students in Bergen increased significantly due to the reform. Since the reform meant that this group of students

attended high school with higher-ability peers, it is tempting to draw the conclusion from this finding that high-ability students gain from attending high school with other high-ability students. Secondary findings urge caution about the effect of the school choice reform on high-ability students, since the results suggested that middle school students adjust rapidly to the new high school intake system. Using a DDD identification strategy that aimed to account for middle school students' adjustment to school choice reform, we found some positive effects on exams taken at the end of high school.

Implementing a performance-based intake system is one way of creating a tracked system where similar students attend school together based on an achievement measure. Detailed policy recommendations regarding intake systems require more in-depth analysis on the total effects of the intake reform. The present study's results suggest that reform had a largely positive effect on students at all ability levels, although it may be more challenging to understand the underlying mechanisms that caused this effect in other groups.

References

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. 2014. The elite illusion: achievement effects at Boston and New York exam schools. *Econometrica*, 82 (1), pp.137–196.

Angelucci, M., De Giorgi, G., Rangel, M. A., and Rasul, I. 2010. Family networks and school enrolment: Evidence from a randomized social experiment. *Journal of Public Economics*, 94 (3), pp.197–221.

Angrist, J. D. 2014. The perils of peer effects. *Labour Economics*, 30, pp. 98–108.

Angrist, J.D. & Lang, K. 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. *American Economic Review*, 94 (5), pp.1613–1634.

Betts, J. R., and Zau, A. 2004. Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript.

Black, S. E., Devereux, P. J., and Salvanes, K. G. 2013. Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31 (1), pp.119–153.

Bobonis, G. J., and Finan, F. 2009. Neighborhood peer effects in secondary school enrollment decisions. The Review of Economics and Statistics 91 (4), pp.695–716.

Bonesronning, H. 2008. Peer group effects in education production: Is it about congestion? The Journal of Socio-Economics 37 (1), pp.328–342.

Bonesronning, H., and Haraldsvik, M. 2014. Peer effects on student achievement: Does the education level of your classmates parents matter? Working paper.

Booij, A. S., Leuven, E., & Oosterbeek, H. 2017. Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84 (2), pp.547–578.

Brugard, K. H. 2013. Does school choice improve student performance? Working paper.

Buchmueller, T. C., DiNardo, J., and Valletta, R. G. 2011. The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from Hawaii. *American Economic Journal: Economic Policy*, 3 (4), pp.25–51.

Carrell, S. E., Malmstrom, F. V., and West, J. E. 2008. *Peer effects in academic cheating. Journal of Human Resources*, 43 (1), pp.173–207.

Carrell, S. E., Sacerdote, B. I., and West, J. E. 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81 (3), pp.855–882.

Chetty, R., Hendren, N., & Katz, L.F. 2016. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *The American Economic Review*, 106 (4), pp.855–902.

Clark, D. 2010. Selective schools and academic achievement. *The BE Journal of Economic Analysis & Policy*, 10 (1).

Dahl, G. B., Løken, K. V., and Mogstad, M. 2014. Peer effects in program participation. *The American Economic Review*, 104 (7), pp.2049–2074.

DeSimone, J. 2009. Fraternity membership and drinking behavior. *Economic Inquiry*, 47 (2), pp.337–350.

Duflo, E., Dupas, P., and Kremer, M. 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *The American Economic Review*, 101 (5), pp.1739–1774.

Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., and Eccles, J. 2005. Peer effects in drug use and sex among college students. *Journal of Abnormal Child Psychology,* 33 (3), pp.375–385.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18 (5), pp.527–544.

Haraldsvik, M. 2014. Does performance-based admission incentivize students? Working paper.

Hoxby, C. 2000. Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867.

Jackson, C.K. 2012. Single-sex schools, student achievement, and course selection: evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics*, 96 (1), pp.173–187.

Jonsson, J. O., and Mood, C. 2008. Choice by contrast in Swedish schools: How peers' achievement affects educational choice. *Social forces*, 87 (2), pp.741–765.

Kling, J.R., Liebman, J.B., & Katz, L.F. 2007. Experimental analysis of neighborhood effects. *Econometrica*, 75 (1), pp.83–119.

Kling, J.R., Ludwig, J., & Katz, L.F. 2005. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120 (1), pp.87–130.

Lalive, R., and Cattaneo, M. A. 2009. Social interactions and schooling decisions. *The Review of Economics and Statistics*, 91 (3), pp.457–477.

Lavy, V., Paserman, M. D., and Schlosser, A. 2012. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122 (559), pp.208–237.

Ludwig, J., Duncan, G.J., Gennetian, L.A., Katz, L.F., Kessler, R.C., Kling, J.R., & Sanbonmatsu, L. 2013. Long-term neighborhood effects on low-income families: evidence

from moving to opportunity. *American Economic Review,* 103 (3), pp.226–231.

Machin, S., and Salvanes, K. G. 2010. Valuing school quality via a school choice reform. IZA Discussion Paper.

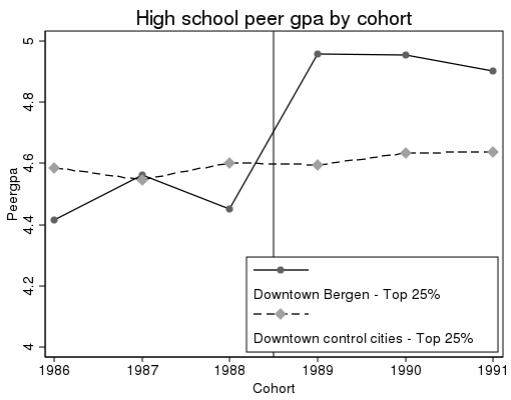
Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60 (3), pp.531–542.

Rao, G. 2013. Familiarity does not breed contempt: Diversity, discrimination and generosity in Delhi schools. Working paper.

Sacerdote, B. 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? Handbook of the Economics of Education, 3, pp.249–277.

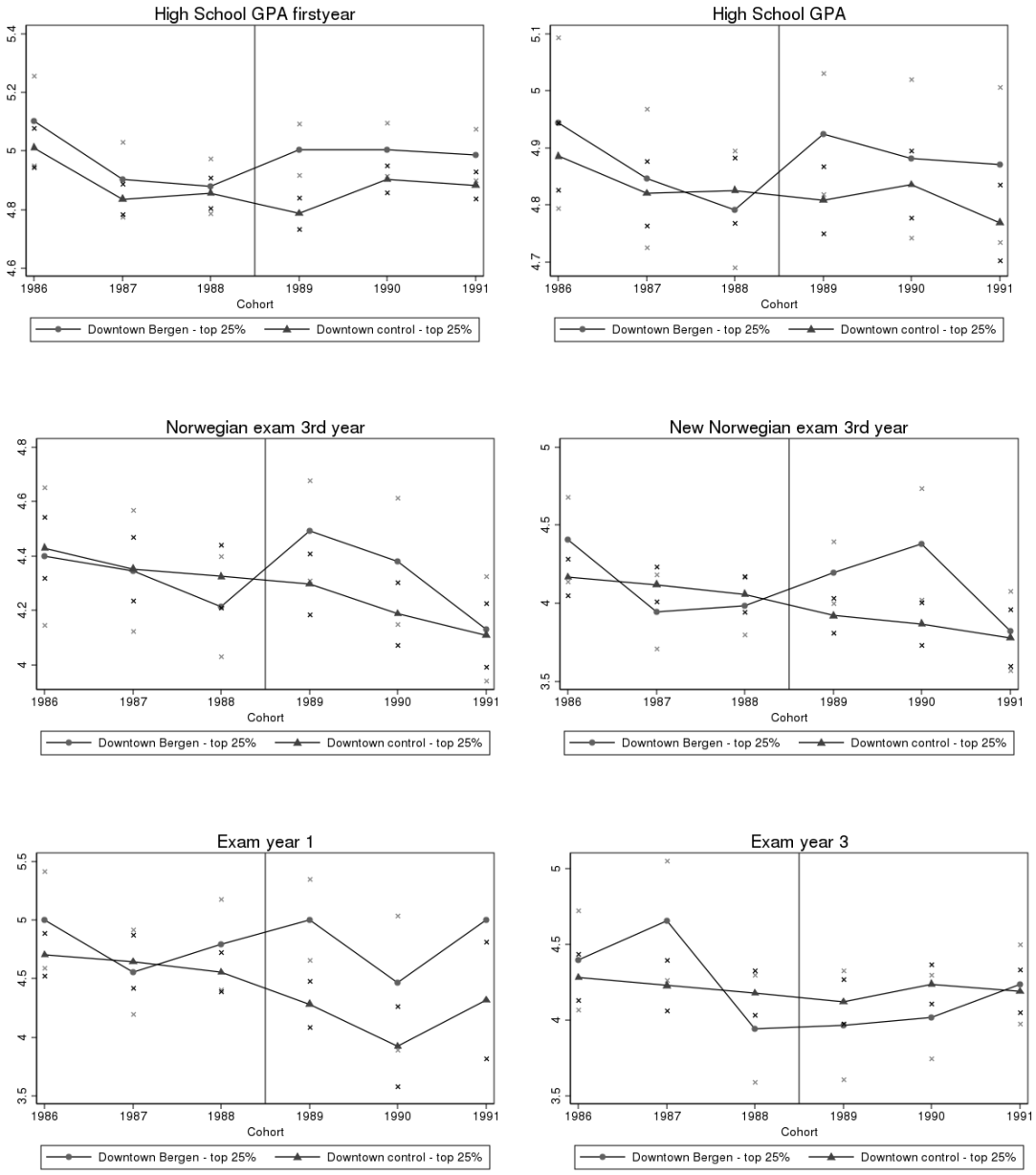
Wilson, J. 2007. Peer effects and cigarette use among college students. *Atlantic Economic Journal*, 35 (2), pp.233–247.

Figure 1



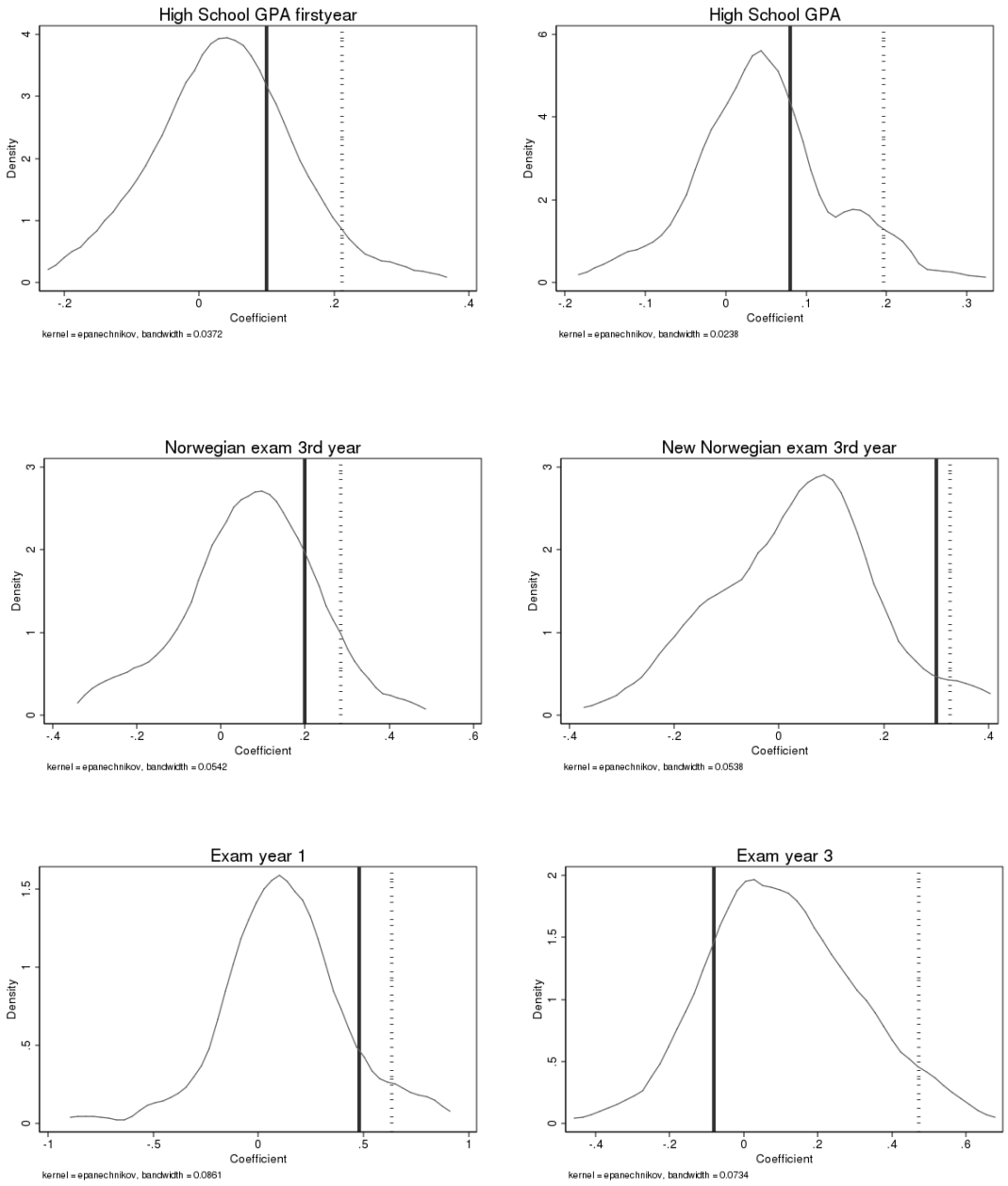
Notes: Figure shows peer gpa by treatment and control groups across cohorts. Cohorts born in 1989 finish middle school in the spring of 2005 and are the first that apply to high school after the school choice reform in Bergen. The treatment group consists of students that attended middle school downtown Bergen and are ranked among the top 25 % at middle school of their cohort in Bergen. The control group consists of students that attended middle school in Kristiansand, Stavanger, Trondheim or Drammen and are ranked among the top 25 % of their cohort.

Figure 2: Outcome trends DD



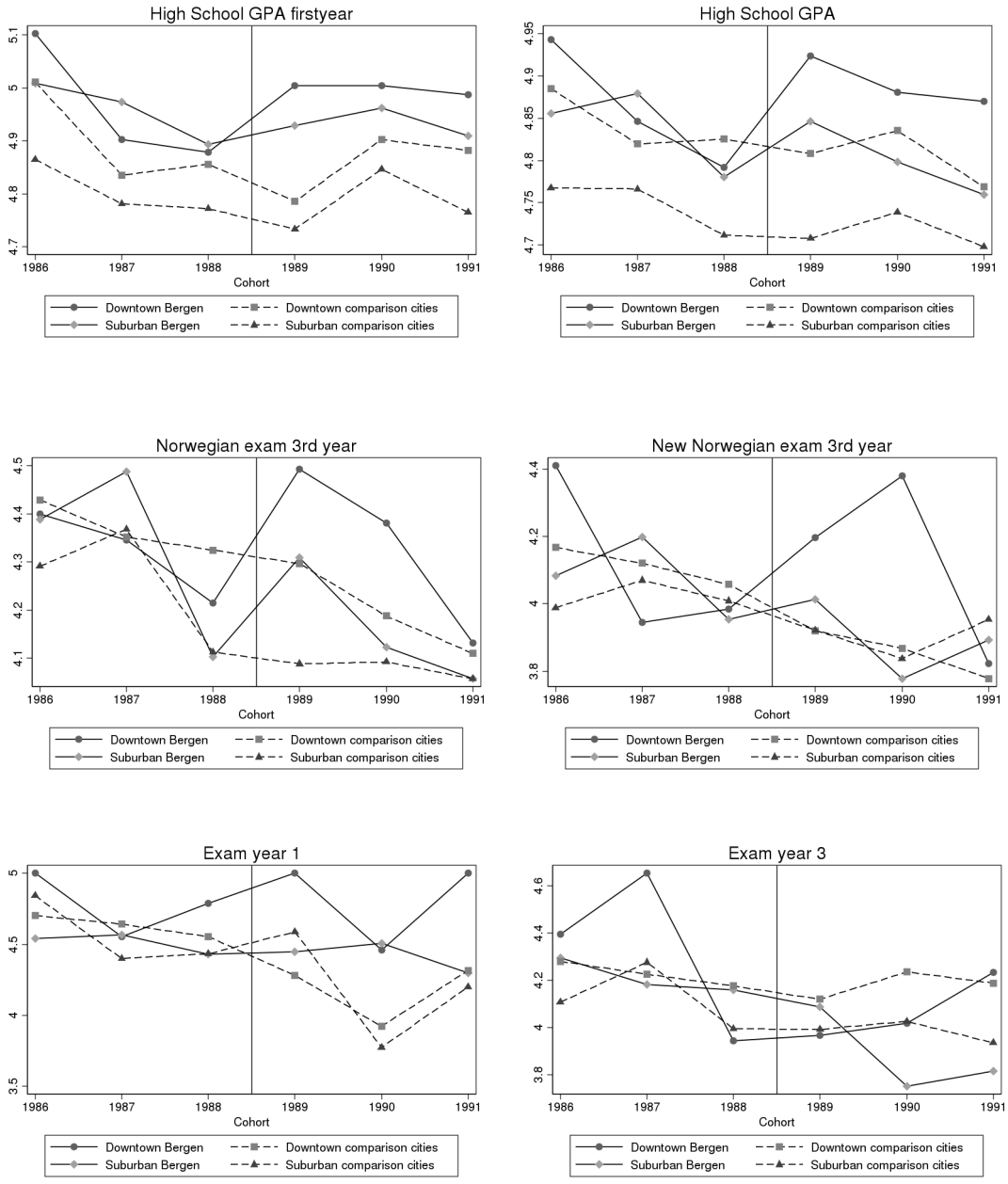
Notes: Figures show outcomes by groups over cohorts. 95 % confidence intervals of means are shown.

Figure 3: Permutation test



Notes: Figures show distribution of coefficents from estimating the effect of placebo reforms. We have assumed that the policy changed happened for each of the 20 largest cities in Norway, and estimated DD coefficient for the top 25 % in each of these cities as well as for the 3 other ability groups of students for each city. This gives a total of 80 coefficents. Dotted line represents the 95 percentile in the distribution of coefficients, while the full line is the estimate for Bergen.

Figure 4: Outcome trends DDD



Notes: Figures show outcomes by groups over cohorts. Only top 25 % of students included. 95 % confidence intervals of means are shown.

Table 1: Descriptive

	Treated		Diff: Treated - control		
	Pre reform		Pre reform	Post reform	
	Mean	SD	Mean	Mean	N
	(1)	(2)	(3)	(4)	(5)
Panel a) - Covariates					
Female	0.59	0.49	-0.08	-0.07	1869
Mother years of education	15.09	2.58	0.91	0.50	1736
Father years of education	15.83	2.68	0.93	0.42	1699
Mother earnings	100353	63944	5801	4492	1818
Father earnings	200699	136088	9561	-4503	1778
Middle School GPA	5.21	0.21	0.08	0.15	1869
Panel b) - Peer characteristic					
Peergpa	4.48	0.14	-0.10	0.31	1869
Panel c) - Outcomes					
Firstear GPA	4.94	0.47	0.05	0.14	1844
High School GPA	4.85	0.46	0.01	0.09	1656
Absence hours	32.23	30.11	-6.53	-7.73	1606
Absence days	14.19	11.60	-3.80	-1.50	1607
Select basic math year 1	0.05	0.21	-0.05	-0.01	1201
Norwegian exam year 3	4.31	0.82	-0.06	0.14	1734
Nynorsk exam year 3	4.09	0.86	-0.03	0.32	1342
Exam score year 1	4.76	0.81	0.13	0.62	420
Exam score year 2	4.49	0.99	0.02	-0.04	1040
Exam score year 3	4.25	1.19	0.02	-0.09	1432
# observations treatment			177	225	
# observations control			694	773	

Notes: Panel a) show descriptive statistics of covariates. Panel b) show endogenous variable peergpa. Peergpa are average middle school GPA of students at highs school. Panel c) show descriptive statistics of high school outcomes. Column (1) show pre reform means of the treated, Column (2) show the standard deviation of the treated. Columns (3) and (4) show difference in means between treatment and control before and after the reform. Treatment consists of students that attended middle school in the downtwon area of Bergen are ranked in the top 25 % of their cohort in Bergen. Control group consists of students that attended middle school in Kristiansand, Stavanger, Trondheim or Drammen and are ranked among the top 25 % of their cohort.

Table 2: Results

Dependent variable: High school outcomes

	(1)	(2)	(3)	(4)
Peergpa	0.42***	0.42***	0.43***	1869
	(0.09)	(0.07)	(0.07)	
GPA firstyear	0.09	0.09*	0.10*	1844
	(0.06)	(0.05)	(0.05)	
HS GPA	0.08	0.07	0.08	1656
	(0.06)	(0.06)	(0.05)	
Hours absent	-1.20	-1.03	-1.07	1606
	(4.38)	(4.33)	(3.97)	
Days absent	2.29	2.78	3.04	1607
	(2.66)	(2.32)	(2.24)	
Select basic math	0.04	0.05	0.04	1201
	(0.08)	(0.07)	(0.07)	
Norwegian exam year 3	0.20	0.20	0.20*	1734
	(0.12)	(0.12)	(0.11)	
Nynorsk exam year 3	0.34**	0.31**	0.30**	1342
	(0.14)	(0.14)	(0.12)	
Exam year 1	0.45*	0.46**	0.48***	420
	(0.24)	(0.21)	(0.18)	
Exam year 2	-0.00	-0.07	-0.02	1040
	(0.15)	(0.14)	(0.14)	
Exam year 3	-0.12	-0.10	-0.08	1432
	(0.20)	(0.19)	(0.18)	
Spesification				
Middle school dummies		x	x	
Background characteristics			x	

Notes: Table show coefficient estimates of β_3 in model 1 in colums (1)-(3) and sample size in Column (4). Sample consist of high ability downtown students in Bergen and control cities. Background characteristics are parents earnings, years of education and gender. Cohorts born 1986-1991. Standard errors in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Table 3: DD estimates for all groups in Bergen

Dependent variable: High school outcomes

	Downtown students				*Suburban students*			
Ability level	*L*	*M-L*	*M-H*	*H*	*L*	*M-L*	*M-H*	*H*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peergpa	0.03	0.13*	0.24***	0.43***	-0.03	0.04	0.11*	0.23***
	(0.09)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)
Firstyear GPA	0.17	0.21***	0.07	0.10*	0.16*	0.22***	0.15***	0.01
	(0.12)	(0.07)	(0.08)	(0.05)	(0.08)	(0.05)	(0.04)	(0.03)
High school GPA	0.22**	0.13**	0.13**	0.08	0.18***	0.20***	0.11***	0.02
	(0.09)	(0.06)	(0.07)	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)
Hours absent	-7.41	-17.70**	-6.11	-1.07	2.36	-4.60	-2.61	1.71
	(11.18)	(7.39)	(6.43)	(3.97)	(7.42)	(4.82)	(3.38)	(3.28)
Days absent	2.32	2.94	2.03	3.04	4.27*	1.64	-0.77	-0.56
	(3.51)	(3.03)	(2.00)	(2.24)	(2.18)	(1.58)	(1.28)	(1.32)
Select basic math	0.02	-0.07	0.10	0.04	-0.07	-0.01	-0.09*	0.02
	(0.09)	(0.08)	(0.08)	(0.07)	(0.05)	(0.05)	(0.05)	(0.03)
Norwegian exam year 3	0.16	0.28**	0.08	0.20*	0.09	0.12*	0.13*	0.03
	(0.15)	(0.11)	(0.13)	(0.11)	(0.09)	(0.07)	(0.08)	(0.08)
Nynorsk exam year 3	0.03	-0.17	0.03	0.30**	0.02	0.13*	0.09	-0.03
	(0.18)	(0.11)	(0.11)	(0.12)	(0.09)	(0.07)	(0.10)	(0.08)
Exam year 1	0.24	0.64*	0.10	0.48***	0.24	0.28	0.23	0.17
	(0.39)	(0.33)	(0.20)	(0.18)	(0.21)	(0.18)	(0.16)	(0.15)
Exam year 2	-0.00	0.20	0.22	-0.02	0.09	0.10	0.09	0.15
	(0.18)	(0.18)	(0.18)	(0.14)	(0.12)	(0.12)	(0.12)	(0.11)
Exam year 3	0.01	0.11	-0.07	-0.08	0.22**	0.05	0.05	-0.20*
	(0.22)	(0.14)	(0.16)	(0.18)	(0.11)	(0.10)	(0.09)	(0.10)
Middle school dummies	x	x	x	x	x	x	x	x
Background characteristics	x	x	x	x	x	x	x	x
Observations	1957	1911	1859	1869	3525	3475	3332	3219

Notes: Table show coefficient estimates of β_3 in equation 1 for all groups. Column (1) compares the change in outcomes of low-ability downtown students from before to after the reform to low ability downtown students in control cities. Column (2) compares the change in outcomes of medium low ability downtown students in Bergen to medium low ability students in control citites. Observations refer to the number of students registered starting academic track. Cohorts born 1986-1991. Standard errors are shown in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Table 4: Placebo

High school outcomes			
	Placebo reforms		
	2003	2004	
	(1)	(2)	(3)
Peergpa	0.08	-0.10	871
	(0.09)	(0.09)	
GPA firstyear	-0.03	-0.07	847
	(0.07)	(0.07)	
HS GPA	-0.05	-0.08	814
	(0.06)	(0.07)	
Hours absent	5.06	1.46	796
	(7.25)	(5.29)	
Days absent	4.25	-1.14	796
	(2.86)	(2.98)	
Select basic math	0.01	0.00	871
	(0.05)	(0.05)	
Norwegian exam year 3	-0.05	-0.08	814
	(0.14)	(0.15)	
Nynorsk exam year 3	-0.36*	-0.12	797
	(0.20)	(0.17)	
Exam year 1	-0.21	0.28	271
	(0.23)	(0.23)	
Exam year 2	0.28	0.16	514
	(0.18)	(0.17)	
Exam year 3	-0.14	-0.48*	633
	(0.28)	(0.24)	
Spesification			
Background chars	x	x	
Middle school dummies	x	x	

Notes: Table show coefficient estimates of β_3 in equation 1. That is, a regression of future outcomes on reform dummy, treatment status dummy and interaction. In columns (1)-(2) the sample consists of pre reform cohorts. Column (1) sets the reform for the school year starting 2003 while Column (2) sets the reform for the school year starting in 2004. Column (3) show the sample size. Cohorts born 1986-1991. Sample includes high ability downtown students in Bergen and control cities. Standard errors are shown in parentheses and are clustered at high school - year level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table 5: Robustness

Dependent variable. High school outcomes						
			Control group	Keeping cohorts		
	Baseline	Top 33%	21 cities	1988-1989	1986-1989	1986-1990
	(1)	(2)	(3)	(4)	(5)	(6)
Peergpa	0.43***	0.40***	0.44***	0.53***	0.46***	0.45***
	(0.07)	(0.07)	(0.06)	(0.11)	(0.10)	(0.08)
GPA firstyear	0.10*	0.09	0.06	0.22***	0.17**	0.11*
	(0.05)	(0.06)	(0.05)	(0.08)	(0.07)	(0.06)
HS GPA	0.08	0.07	0.07	0.17**	0.10	0.07
	(0.05)	(0.06)	(0.05)	(0.08)	(0.07)	(0.06)
Hours absent	-1.07	0.71	0.33	-6.76	-5.56	-3.12
	(3.97)	(3.84)	(4.90)	(4.45)	(4.55)	(3.99)
Days absent	3.04	3.59*	3.06	3.04	2.59	1.90
	(2.24)	(2.00)	(2.09)	(3.26)	(2.69)	(2.37)
Select basic math	-0.02	-0.02	-0.02	0.05*	0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Norwegian exam year 3	0.20*	0.14	0.15	0.33**	0.24*	0.25**
	(0.11)	(0.11)	(0.10)	(0.16)	(0.13)	(0.11)
Norwegian II exam year 3	0.30**	0.18	0.22*	0.45***	0.29*	0.36***
	(0.12)	(0.13)	(0.13)	(0.14)	(0.16)	(0.13)
Exam year 1	0.48***	0.43**	0.15	0.37*	0.53***	0.48**
	(0.18)	(0.17)	(0.16)	(0.19)	(0.19)	(0.19)
Exam year 2	-0.02	-0.03	0.07	-0.03	0.04	-0.10
	(0.14)	(0.15)	(0.12)	(0.19)	(0.17)	(0.15)
Exam year 3	-0.08	-0.24	-0.06	0.28	-0.09	-0.16
	(0.18)	(0.17)	(0.15)	(0.22)	(0.17)	(0.17)
Background char.	x	x	x	x	x	x
Middle school dummies	x	x	x	x	x	x
High school dummies						
Observations	1869	2420	12666	651	1201	1544

Notes: Table show coefficient estimates of β_3 in equation 1. That is, a regression of future outcomes on reform dummy, treatment status dummy and interaction. Column (2) keeps students ranked in the top 33 % of students within year and city. Column (3) includes students scoring in the top 25 % in more cities in Norway in the control. Column (4)-(6) only keeps students belonging to the cohorts indicated in the table header. Column 8 includes a spesification with high school dummies. Oslo not incuded. Cohorts born 1986-1991. Standard errors are shown in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Table 6: Selection and dynamic response

	Years of education		Earnings			
	Mother	Father	Mother	Father	Female	MS GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel a)						
Short (α_5)	-0.24	-0.39	12679.12	-528.81	-0.02	0.08*
	(0.35)	(0.38)	(10350.45)	(18665.42)	(0.06)	(0.05)
Middle(α_6)	-0.37	-0.19	-13840.84	-22632.64	-0.07	0.03
	(0.62)	(0.46)	(9097.84)	(23835.71)	(0.06)	(0.03)
Long(α_7)	-0.67**	-0.87**	-5590.74	-12197.42	0.09	0.11***
	(0.33)	(0.40)	(11620.11)	(18273.99)	(0.08)	(0.03)
N	1736	1699	1818	1778	1869	1869

Panel b)	Peergpa	Norwegian exam year 3	Nynorsk exam year 3	Exam year 3	GPA firstyear	HS GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Short (α_5)	0.46***	0.24*	0.27	-0.10	0.17**	0.10
	(0.10)	(0.13)	(0.17)	(0.18)	(0.07)	(0.07)
Middle(α_6)	0.43***	0.25	0.53***	-0.23	0.05	0.03
	(0.11)	(0.16)	(0.12)	(0.20)	(0.06)	(0.06)
Long(α_7)	0.38***	0.11	0.07	0.09	0.08	0.13
	(0.10)	(0.15)	(0.15)	(0.21)	(0.06)	(0.09)
N	1869	1734	1342	1432	1844	1656

Notes: Table show coefficient estimates of α_5, α_6 and, α_7 in model 2. Panel a show estimates when the dependent variable is predetermined background variables and middle school GPA, while panel b show estimates when dependent are high school outcomes. Middle school dummies included in all regressions, and covariates included in regressions in panel b. Sample includes high ability downtown students in Bergen and control cities. Cohorts born 1986-1991. Standard errors are shown in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Table 7: Results II

Dependent variable: High school outcomes		
	DDD	
	(1)	(2)
Peergpa	0.19***	1869
	(0.06)	
GPA firstyear	0.09	1844
	(0.05)	
HS GPA	0.07	1656
	(0.06)	
Hours absent	-2.81	1606
	(-5.15)	
Days absent	3.38	1607
	(2.24)	
Select basic math	0.02	1201
	(0.07)	
Norwegian exam year 3	0.16	1734
	(0.12)	
Nynorsk exam year 3	0.34**	1342
	(0.14)	
Exam year 1	0.31	420
	(0.25)	
Exam year 2	-0.18	1040
	(0.16)	
Exam year 3	0.11	1432
	(0.19)	
Spesification		
Middle school dummies	x	
Background characteristics	x	

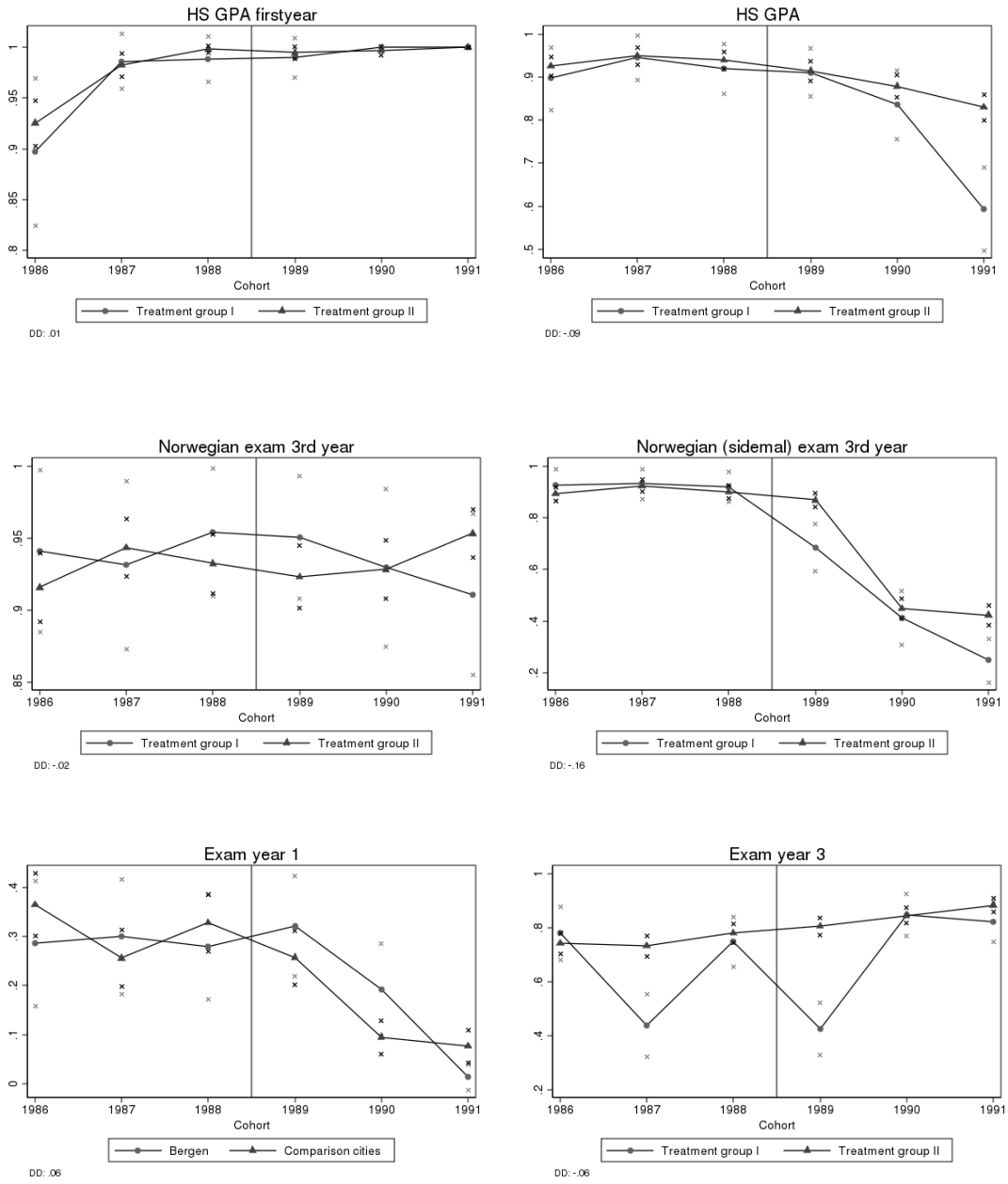
Notes: Table show coefficient estimates of γ_6 in Model 3 in Colum (1). Column (2) show sample size. Background characteristics are parents earnings, years of education and gender. Cohorts born 1986-1991. Standard errors are shown in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Table 8: Proportion of high ability downtown students attending school

	Katten	BGH	Langhaugen	Tanks	Bjørgvin	Downtown	Private
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pre	0.23	0.12	0.25	0.26	0.04	0.90	0.05
Post	0.32	0.21	0.31	0.08	0.00	0.93	0.04
Diff	0.09	0.09	0.06	-0.18	-0.04	0.03	-0.01
P-value	0.04	0.03	0.20	0.00	0.01	0.37	0.97
Observations	402	402	402	402	402	402	402

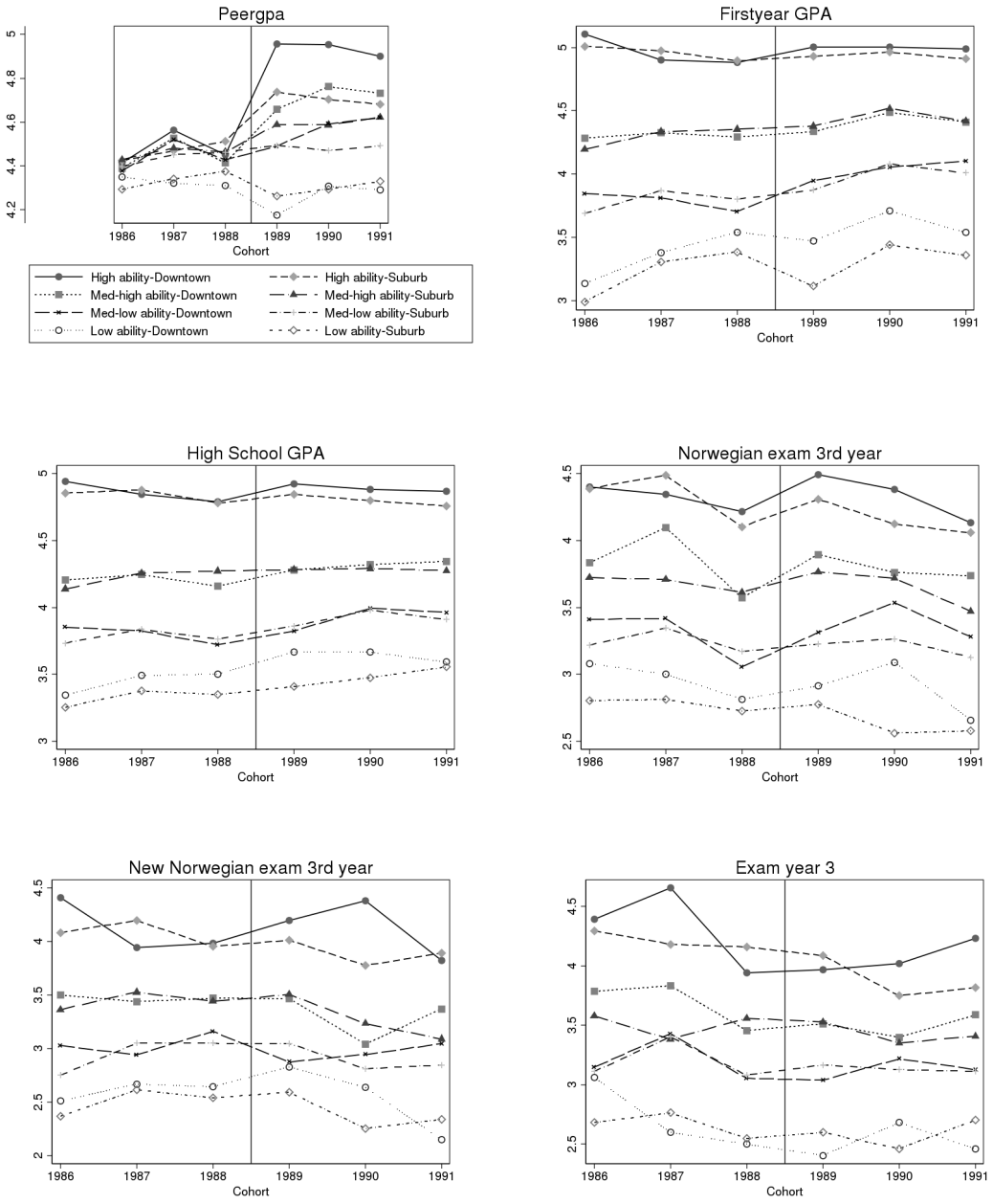
Notes: Table show proportion of high ability students attending each downtown high school in Bergen before and after reform in columns 1-5. Column 6 report the proportion at all downtown schools, while Column 7 report the proportion of high ability downtown students at private schools. P-values from a two-tailed test of proportions are reported.

Figure A.1: Attrition rate trends



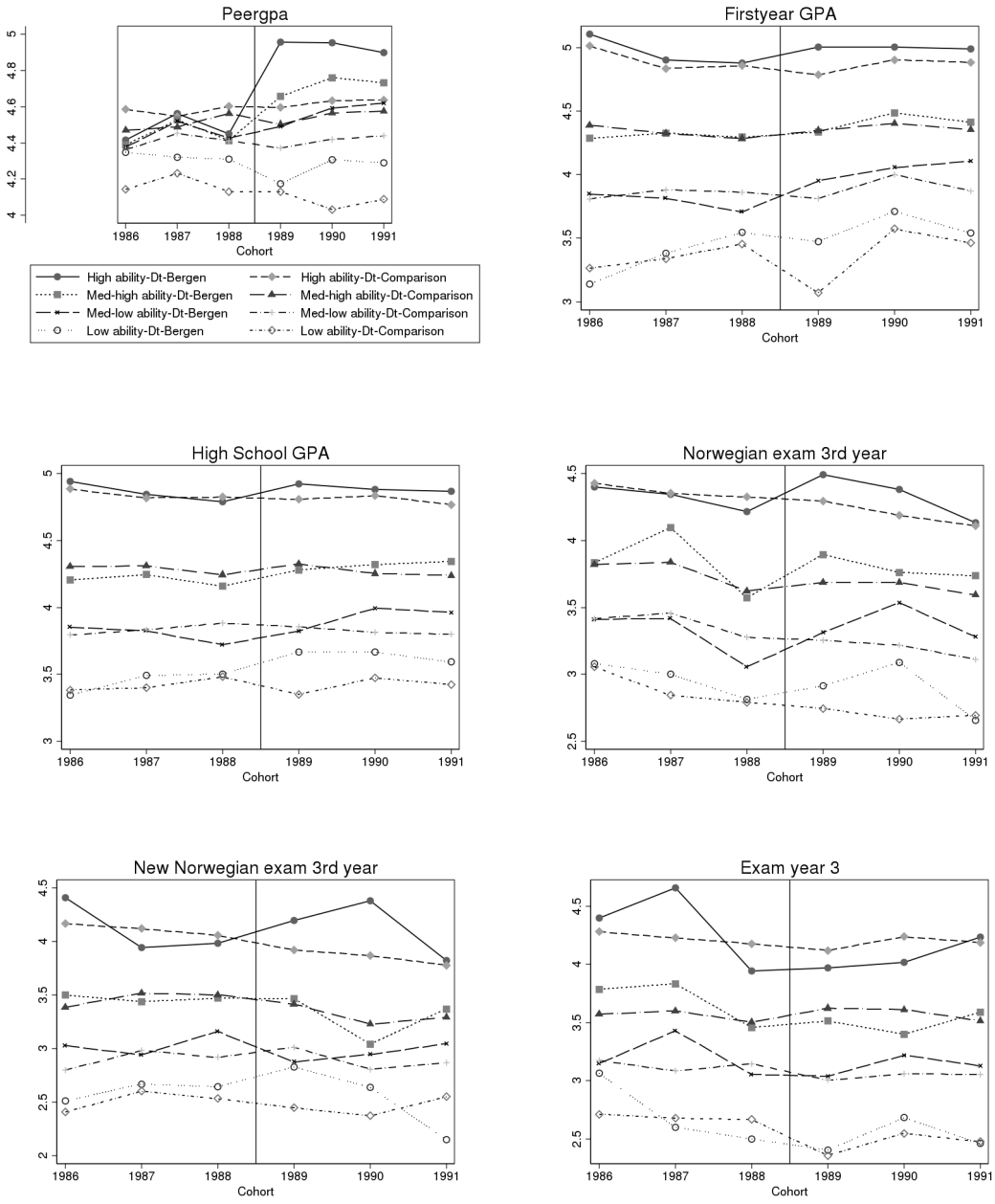
Notes: Figures show rates for which we observe test score outcomes over time. Difference in difference estimates are shown in the corner of each figure. Only significant difference in trends are detected for HS GPA at 10 % significance level. The jump in level of HS GPA firstyear from cohort 1986 is caused by a lack of administrative grades from the first year for this cohort.

Figure A.2: Descriptive figures - Bergen downtown and Bergen suburban students



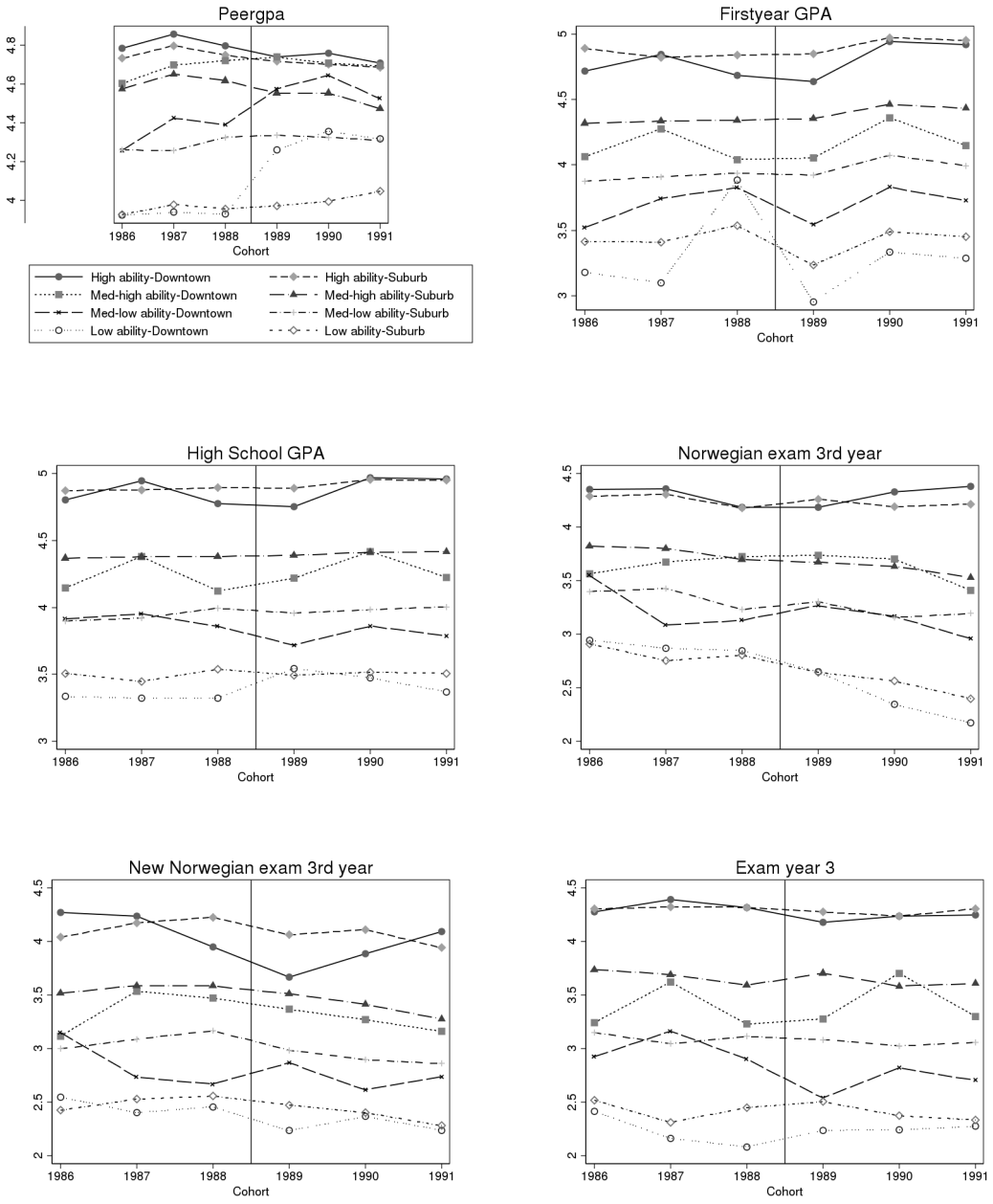
Notes: Figures show mean outcomes by cohort and group. Legends indicated in figure a.

Figure A.3: Descriptive figures - Bergen downtown and control downtown students



Notes: Figures show mean outcomes by cohort and group. Legends indicated in figure a.

Figure A.4: Descriptive figures - Oslo downtown and Oslo suburban students



Notes: Figures show mean outcomes by cohort and group. Legends indicated in figure a.

Table A.1: Estimation - Clustering on different levels

Dependent variable. High school outcomes							
	No	Middle school	High school	City	City 21	MS*year	HS*year
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peergpa	0.42***	0.42***	0.43***	0.42***	0.44***	0.42***	0.43***
	(0.02)	(0.03)	(0.10)	(0.01)	(0.03)	(0.05)	(0.07)
GPA firstyear	0.10**	0.10**	0.10	0.10*	0.06***	0.10**	0.10*
	(0.05)	(0.05)	(0.06)	(0.03)	(0.02)	(0.04)	(0.05)
HS GPA	0.08	0.08**	0.08	0.08*	0.07***	0.08**	0.08
	(0.06)	(0.03)	(0.06)	(0.04)	(0.02)	(0.04)	(0.05)
Hours absent	-1.07	-1.07	-1.07	-1.07	0.33	-1.07	-1.07
	(4.31)	(3.04)	(3.30)	(3.23)	(1.23)	(2.92)	(3.97)
Days absent	3.04	3.04**	3.04	3.04***	3.06***	3.04*	3.04
	(2.12)	(1.36)	(1.93)	(0.48)	(0.49)	(1.67)	(2.24)
Select basic math	0.04	0.04	0.04	0.04	0.08***	0.04*	0.04
	(0.04)	(0.05)	(0.08)	(0.03)	(0.01)	(0.02)	(0.07)
Nor exam year 3	0.20*	0.20**	0.20***	0.20**	0.15***	0.20**	0.20*
	(0.10)	(0.09)	(0.07)	(0.06)	(0.03)	(0.07)	(0.11)
Norwegian II exam year 3	0.30**	0.30***	0.30***	0.30***	0.22***	0.30**	0.30**
	(0.12)	(0.08)	(0.07)	(0.03)	(0.03)	(0.11)	(0.12)
Exam year 1	0.48**	0.48***	0.48***	0.48**	0.15***	0.48***	0.48***
	(0.20)	(0.14)	(0.13)	(0.14)	(0.05)	(0.15)	(0.18)
Exam year 2	-0.02	-0.02	-0.02	-0.02	0.07*	-0.02	-0.02
	(0.16)	(0.09)	(0.17)	(0.15)	(0.04)	(0.15)	(0.14)
Exam year 3	-0.08	-0.08	-0.08	-0.08	-0.06**	-0.08	-0.08
	(0.14)	(0.08)	(0.11)	(0.11)	(0.03)	(0.18)	(0.18)
Spesification							
Middle school dummies	x	x	x	x	x	x	x
Background chars	x	x	x	x	x	x	x

Notes: Table show coefficient estimates of β_3 in model 1. Treatment group are high ability downtown students in Bergen, while control group are high ability students in control cities Trondheim, Savanger, Kristiansandand Drammen. Standard errors in parentheses. Cohorts born 1986-1991. Level of clustering indicated in column headers. City means clustering at city/muncipal level, while city 21 is clustering on city level when expanding to include 21 cities. MS*year is clustering on middle school - year, and HS*year is clustering on high school - year. * p<0.10, ** p<0.05, *** p<0.01

Table A.2: School quality

	Downtown students				Suburban students			
Ability level:	Low	Med-low	Med-high	High	Low	Med-low	Med-high	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Downtown								
Prop. pre	0.10	0.09	0.09	0.12	0.16	0.17	0.13	0.13
Prop. post	0.07	0.10	0.11	0.15	0.05	0.12	0.16	0.23
Diff	-0.03	0.01	0.02	0.02	-0.11	-0.05	0.03	0.10
P-value	0.01	0.57	0.05	0.11	0.00	0.00	0.02	0.00
Private								
Prop. pre	0.01	0.02	0.03	0.03	0.11	0.19	0.29	0.33
Prop. post	0.08	0.05	0.02	0.02	0.24	0.21	0.18	0.20
Diff	0.08	0.03	-0.01	-0.01	0.13	0.01	-0.10	-0.13
P-value	0.00	0.05	0.33	0.44	0.00	0.67	0.00	0.00
Katten								
Prop. pre	0.13	0.11	0.11	0.14	0.13	0.13	0.11	0.15
Prop. post	0.02	0.01	0.07	0.29	0.03	0.04	0.10	0.44
Diff	-0.10	-0.10	-0.03	0.15	-0.09	-0.09	-0.01	0.29
P-value	0.00	0.00	0.17	0.00	0.00	0.00	0.61	0.00
BHG								
Prop. pre	0.08	0.07	0.07	0.08	0.20	0.18	0.18	0.14
Prop. post	0.00	0.04	0.07	0.18	0.00	0.03	0.24	0.44
Diff	-0.08	-0.03	0.00	0.09	-0.20	-0.15	0.07	0.30
P-value	0.00	0.17	0.89	0.00	0.00	0.00	0.06	0.00
Bjorgvin								
Prop. pre	0.17	0.11	0.09	0.04	0.21	0.18	0.11	0.10
Prop. post	0.26	0.12	0.05	0.00	0.19	0.22	0.10	0.06
Diff	0.09	0.01	-0.04	-0.04	-0.02	0.04	-0.01	-0.05
P-value	0.02	0.77	0.13	0.00	0.59	0.26	0.86	0.07
Langhaugen								
Prop. pre	0.08	0.11	0.11	0.15	0.19	0.14	0.11	0.10
Prop. post	0.02	0.11	0.15	0.19	0.02	0.12	0.22	0.18
Diff	-0.06	-0.01	0.04	0.04	-0.17	-0.02	0.11	0.07
P-value	0.00	0.83	0.09	0.23	0.00	0.46	0.00	0.01
Tanks								
Prop. pre	0.06	0.07	0.09	0.17	0.11	0.21	0.15	0.14
Prop. post	0.06	0.21	0.22	0.07	0.02	0.20	0.14	0.08
Diff	-0.00	0.14	0.13	-0.10	-0.08	-0.02	-0.01	-0.06
P-value	0.90	0.00	0.00	0.00	0.00	0.59	0.78	0.03

Notes: Table show proportion students at each downtown school in Bergen before and after the school choice reform in column 1-5. Also show the proportion at all downtown schools, and the proportion students at private schools. P-value of a two-proportion z-test for differences in proportions are reported.

Table A.3: DDD placebo

	Years of education		Earnings			
	Mother	Father	Mother	Father	Female	MS GPA
$\hat{\gamma}_7$	-0.65**	-0.59*	-2326.65	635.34	0.01	0.04
	-0.33	-0.34	-8593.76	-16737.4	-0.07	-0.03
N	4760	4674	4974	4894	5081	5081

Notes: Table show estimates of γ_6 in model 3. Dependent variable are predetermined background variables and middle school GPA. Middle school dummies included in all regressions. Cohorts born 1986-1991. Standard errors are shown in parentheses and are clustered at high school - year level. * p<0.10, ** p<0.05, *** p<0.01

Department of Economics
University of Bergen
PO BOX 7800
5020 Bergen
Visitor address: Fosswinckels gate 14
Phone: +47 5558 9200
www.uib.no/econ/