Ability Grouping's Effects on Grades and the Attainment of Higher Education –

A Natural Experiment [†]

Magnus Bygren

Department of Sociology, Stockholm University

Institute for Analytical Sociology, Linköping University

Abstract

To test the effect of ability grouping on grades and the attainment of higher education, this study examines a naturally occuring experiment—an admission reform that dramatically increased ability sorting between schools in the municipality of Stockholm. Following six cohorts of students (N = 79,020) from the age of 16 to 26, I find a mean effect close to zero and small positive and negative differentiating effects on grades. With regard to the attainment of higher education, I find a mean effect close to zero, the achievement-group gap was unaffected, the immigrant–native gap increased, and the class-background gap decreased. These results are consistent with much previous research that has found small mean effects of ability grouping. They are inconsistent with previous research, however, in that I find ability grouping's effects on gaps are rather small and point in different directions.

Introduction

Educational systems tend to allocate more able students into more demanding tracks. In the scholarly literature, "ability grouping" refers to curricular, or at least instructional, differentiation on the basis of some measure of student ability or achievement (LeTendre, Hofer and Shimizu 2003). This kind of grouping is uncontroversial in tertiary education, but it is more disputed in primary and, to some extent, secondary education. Estimating the causal effects of such sorting involves difficulties in netting out selection effects from instruction and peer effects.

This study focuses on ability grouping across secondary schools based on grades, with the aim of estimating the causal effects of ability grouping on grades and the attainment of higher education, net of any effects of selection. To identify the effects of ability grouping, I analyze the effects of an admission reform implemented in the municipality of Stockholm in 2000. Before the reform, school admission was based on residence; students living close to a school had priority in the admission process. After the reform, admission was based on grades; students with high grades were given priority in admission to schools. As a consequence of the reform, grade sorting across schools increased dramatically (Söderström and Uusitalo 2010). In practice, the reform turned a small number of old schools in the central and affluent parts of Stockholm into elite institutions dominated by students at the upper end of the achievement distribution. The student body of schools in lower socioeconomic areas developed in the opposite direction.

Many methodological difficulties are typically associated with studying the consequences of grouping students on the basis of achievement and related sorting effects in the educational system. The reform examined here affected ability grouping across schools, but the schools and the formal curricula remained the same, as did the population from which students were recruited.[1] My empirical case can thus be treated as a natural experiment, making it possible to identify the causal effects of ability grouping across schools with a relatively high degree of confidence.

Potential effects of ability grouping

All modern school systems engage in some differentiation of the learning environment, and the advantages and disadvantages of ability grouping have been debated since at least the 1920s (Slavin 1987). Proponents of ability grouping argue that narrowing the range of student abilities allows teachers to align the level and pace of instruction more closely with student needs, which provides better opportunities for students to make progress commensurate with their abilities and thus helps maintain student interest and motivation (Oakes and Guiton 1995). Furthermore, because ability grouping separates students by some measure of achievement, "pond effects" are minimized: students' work is compared only to that of similar-achievement peers, preventing possible negative effects on low-achievement students' self-esteem, which might result from comparisons with high-achievement students (cf. Marsh 1991, Seaton et al. 2008). Critics of ability grouping claim that less able students lose the opportunity to benefit from positive peer effects, and they argue that being labeled as less able communicates low expectations, which may become self-fulfilling (Figlio and Page 2002, Slavin 1987,

Terwel 2005). Furthermore, this kind of sorting places racial and ethnic minority students and children from working-class homes into low-achievement tracks (see Alba, Sloan and Sperling 2011, Terwel 2005), contributing to the social reproduction of elite and underclass groups in society. Other scholars argue that the composition of classes, or peer effects for that matter, are not very important compared to the type and quality of the learning environment in a classroom (Hattie 2009).

From a theoretical standpoint, the effects of ability grouping on educational outcomes are not clear. First, one must consider potential peer effects; students are influenced by their peers' attitudes, achievements, and choices (see Owens 2010, Sacerdote 2001). One might also find instruction effects; the quality of instruction commonly varies between groups in ability grouping systems, and if "high-achievement schools" have a higher-quality learning environment than schools with a lower average achievement level, this would systematically generate increased differences between groups. Second, one must also consider potential social comparison effects. Students compare themselves to others in their immediate social surroundings, and depending on how these *relative* comparisons turn out, we should expect different effects on students' self-concepts and educational choices (e.g., Davis 1966, Mood and Jonsson 2008). All else being equal, the more successful the school environment, the lower students' academic self-concepts and the less ambitious their educational choices will be (Marsh 1991). If social contrast effects are important, increased ability grouping should attenuate differences between groups. If students' perceived relative standing with regard to achievement matters for their educational choices (e.g., whether they choose to continue into higher education), these choices should become more similar across groups in settings with higher ability sorting.

Studies on the effects of ability grouping

Most early studies of ability grouping found that grouping raises the achievement and aspirational standards of students located at the upper end of the achievement distribution but is detrimental to students located at the lower end (e.g., Alexander, Cook and McDill 1978, Gamoran 1987, Kerckhoff 1986). However, many of these studies are methodologically problematic due to the difficulties of netting out selection effects from estimates of the effect of ability grouping. If students self-select, or are selected, into certain tracks, and if it is difficult to pin down the actual selection criteria, then the observed effect might be a spurious effect of unobserved ability differences between groups, and not an effect of tracking. Furthermore, if the quality of instruction varies between tracks, and these differences are not accounted for in the estimation of effects, this might also give rise to spurious effects of sorting (Argys, Rees and Brewer 1996, Betts and Shkolnik 2000, Hattie 2002).

One way of netting out these differences is to use regression-based methods, whereby these kinds of observed differences are controlled for statistically to obtain unbiased estimates of the effect of ability grouping. Studies using this kind of design typically report zero, or close to zero, effects of ability grouping on mean achievement, and differential achievement effects across high- and low-achievement groups, but results are not uniform. Hoffer (1992) compared students across U.S. middle schools that varied in their use of tracking and found no mean achievement growth effects of ability grouping; he did find that tracking widened the achievement gap between high- and low-achievement groups (see Jackson 2009, Saleh, Lazonder and De Jong 2005 for similar results). Using a similar design, Argys, Rees and Brewer (1996) found a small positive mean effect of tracking on test scores

for U.S. 10th graders, but they noted that ability grouping only benefited students from the high-achievement group and was detrimental to the achievement of students in the low-achievement group. In contrast, Betts and Shkolnik (2000) found much smaller differential effects across U.S. middle and high school achievement groups compared to previous studies, once more adequate control groups were used for the comparison. And Figlio and Page (2002) compared achievement gains across similar students attending tracked versus untracked U.S. schools and found that ability grouping appeared to help low-achievement students.

These two latter studies were methodologically more sophisticated than much previous research, which makes the absence and even reversal of expected differential effects noteworthy. Systematic reviews of the literature have concluded that ability grouping, or tracking, has minimal effects on average learning outcomes but differential effects on high-tracked versus low-tracked students (Hattie 2009, Kao and Thompson 2003), but these reviews do not take the quality of study design into consideration.

Taken together, research indicates zero or close to zero mean achievement effects of ability grouping. Most studies also show that ability grouping is detrimental to the achievement of low-achievement students or students from a less advantageous social background, and grouping benefits high-achievement students and students from highly resourced families. However, a few well-designed studies show much smaller or even reversed differential effects.

Empirical Expectations and Analytic Strategy

One limitation of previous research is that studies usually focus on single outcomes in the short term (e.g., grades or test scores), making the assessment of long-term effects of ability grouping difficult. If ability grouping effects are either temporary or accumulate over time, then a relatively early observation of outcomes could substantially over- or under-estimate the impact of ability grouping on educational outcomes (Brunello and Checchi 2007, Ireson and Hallam 1999).

As Boudon (1974) noted, between-group differences in educational attainment can be conceptualized as the combined outcome of between-group differences in grades (primary effects) and educational choices, given grades (secondary effects). When we change ability sorting, we simultaneously change the learning environment and, to some extent, the social environment in which choices to continue into higher education are made. Previous research suggests that the direction of peer effects on achievement is positive, but this is not necessarily true for educational choice, because high-performance environments often discourage ambitious educational choices (cf. Marsh 1991, Mood and Jonsson 2008).

Thus, I first evaluate the mean effects of increased ability grouping on grades and the attainment of higher education. I expect these effects will be small or absent. Second, I evaluate whether ability grouping increases grade gaps and attainment gaps between student groups defined by country of birth, class background, and students' (prior) achievement levels. Low-achievement student groups and children from low-resource families may be more vulnerable to a downward shift in peers' average achievement, because these students, on average, have fewer resources to compensate for conditions at school (Jackson 2009, Saleh, Lazonder and De Jong

2005). Consequently, I expect these gaps will grow with increased ability grouping. Additionally, if students are disadvantaged with regard to more than one dimension—for example, an immigrant student with low grades—we should find even greater gaps.

With regard to higher-education attainment gaps, expectations are less clear. On the one hand, increased grade gaps should translate into increased attainment gaps. On the other hand, because high-performing peers can depress a student's belief in the likelihood of succeeding in higher education, discouraging ambitious educational choices, we should see smaller between-group differences in educational choices, given grades. For these analyses, I thus investigate whether, and if so the degree to which, increased ability grouping across schools affects education attainment gaps through changes in grades (primary effects) or changes in propensities for making the transition to higher education (secondary effects).

The hypothesized mechanism in this scenario is the effect of ability grouping on peer composition in schools. I thus test whether changes in peers' average grade level can account for any estimated treatment effects; these effects should attenuate once peer-achievement composition is accounted for.

The Admission Reform: School Choice for High Achievers

The educational system in Sweden is goal-oriented: the government decides on the framework of laws and regulations, but operations are decentralized to the country's 290 autonomous local municipalities. In 2000, the municipality of Stockholm implemented an upper-secondary school choice reform. Until 1999, residential proximity to a given school determined a student's priority in the

admission process. Students applied not to a specific school, but rather to enter a specific upper-secondary school program (there are 18 national programs; 12 vocational and six academic); program admissions were determined by grades. Once students were admitted to a program, they were referred to the closest school that offered it. Applicants could express preferences about which school they wanted to go to, but people living in the school's catchment area had priority.

The cohort that enrolled in the fall of 2000 was the first that applied to both a program and a school. Applicants were ranked according to their grades, and students with the highest grades were admitted first, regardless of their residential proximity to the school (Söderström and Uusitalo 2010, USK 2002). Admissions were handled by the municipality's central admissions unit. In contrast to classic tracking policies, the admission reform was not designed to increase ability grouping, nor was it self-evident it would have this effect. It has become clear, however, that the new rules have had consequences for student composition across schools (see Söderström 2006), and some schools have closed because they failed to attract students under this new policy.[2]

Research Design

A well-known challenge for social research that aims to estimate the causal effects of social environments is that individuals select into these environments on the basis of unobservable factors, which makes it difficult to disentangle the effects of pre-selection factors from post-selection environmental factors. Schools are no exception to this rule, because student selection into schools is far from random. For the present project, the observed effect of being "placed" in a particular school might

9

be a spurious effect of unobserved pre-selection ability differences between groups, and not an effect of ability grouping per se.

I circumvent this problem by raising the level of analysis and comparing students across different institutional settings (municipalities) that vary their degree of ability grouping. Because the reform affected schools in only one municipality, I can compare student outcomes before and after the reform in this municipality and use student outcomes in neighboring municipalities as a comparison. Specifically, I assess the effects of the reform using a difference-in-differences (DD) design (e.g., Angrist and Pischke 2009). The basic logic of DD is to examine the effect of a treatment by comparing the development of y in the treatment group to the development of y in a control group. For the present case, the treatment group subsequent to treatment consists of students admitted to upper-secondary schools in the municipality of Stockholm in the year 2000 and onward; the treatment group before treatment consists of students admitted to upper-secondary schools in the municipality of Stockholm prior to 2000. I use the control group to net out other simultaneous changes, assuming these other changes would be the same in the treatment and control groups in the absence of treatment.

The basic model specification is as follows. Let a treatment group indicator A denote the Stockholm group. In a pooled cross-section, let $T2$ denote the second, post-reform period and $AT2$ its product. A simple equation for evaluating the total impact of the reform reads:

$$y = \beta_0 + \beta_1 T2 + \beta_2 A + \beta_3 AT2 + \varepsilon \qquad (1)$$

where β_3 is the estimated treatment effect. A vector of control variables X may also be included in this equation. The key identifying assumption of the DD approach is

that the average difference in *y* between treatment and control group (conditional on *X*), in the absence of treatment, would remain unaltered in the post-treatment period. That is, had the treatment not occurred, both groups would have experienced the same time trends (Lechner 2011). One threat to the common trend assumption, and to a causal interpretation of the treatment, is that changes in the composition of students between the treatment and control groups may give rise to diverging trends and a spurious treatment effect. The regressions thus include control variables that capture social background, prior grades, and other known predictors of educational outcomes.[3] I also investigate pre-treatment trends in outcomes to rule this out as a potential confounder. Because there are theoretical reasons to investigate whether there is treatment-effect heterogeneity across subgroups, I perform such an analysis, details of which are in the Results section.

Data and variables

I used an extract of a compilation of population registers at Statistics Sweden. Because the admission reform was implemented in 2000, I define individuals with upper-secondary entry years 1997 to 1999 as the pre-treatment group, and individuals with entry years 2000 to 2002 as the group entering upper-secondary school during the treatment period. Although there are differences between the municipality of Stockholm and the surrounding county of Stockholm, trends with regard to socioeconomic composition, immigrant composition, and residential segregation were extremely similar in these areas during the study period (cf. USK 2006). To allow sufficient time to attain some amount of tertiary education, I do not include cohorts younger than the 2002 cohort.

11

Among the six cohorts, a total of 108,725 students exited lower-secondary school during the years examined. Of these students, 28,936 never completed the upper-secondary level, either because they never entered the upper-secondary level, or, more commonly, because they entered but dropped out before graduation. I report whether the reform affected mean and group-specific dropout rates. However, I am primarily interested in students who were fully exposed to the changes in student composition; for the main analysis, I thus include only students who actually completed the upper-secondary level. Of these, I excluded 769 individuals due to their having missing values on the higher-education dependent variable. Following these adjustments, the main analytic sample consists of 79,020 individuals, of whom 38 percent belong to the treatment group.

Because I base the analysis on a population sample of students who fulfilled the selection criteria, the results may be seen as factual for the Stockholm area during the time of the study, but generalizations beyond this context are justifiable only on analytic grounds. Furthermore, it is less appropriate to use significance tests as indicators of the probability of the identified statistical correlations given a null correlation in the population. In the present case, the sample is the population, and the estimated coefficients therefore represent population parameters. For this reason, I do not report standard errors and significance tests for the coefficients. To make interpretation of the coefficients as intuitive as possible, I used linear probability models to estimate effects.[4]

Dependent Variables

I derived the first dependent variable, *grades*, from students' GPA at graduation from the upper-secondary level, which usually occurs at age 19. To weed out possible

effects of grade inflation and changes in the grading system, I transformed grades into graduation-year percentile rank scores, with 0 indicating the lowest possible score and 100 the highest possible score.[5] The second dependent variable, *tertiary educational attainment*, is a dichotomous measure indicating whether the individual attained at least one semester of tertiary education, seven years after the upper-secondary graduation year (usually the year in which a student turns 26).[6]

Independent Variables

Treatment group has the value one if an individual went through the upper-secondary level in a school in the municipality of Stockholm, and zero otherwise. *T*2 is a dummy variable with the value zero for students who entered the upper-secondary level in 1997 to 1999, and the value one for students who entered this level in 2000 to 2002. *Immigrant* has the value one if a student and both parents were all born abroad. I measure *class background* as the student's dominant Erikson-Goldthorpe-Portocarero (EGP) household class during the period 1980 to 1990 (Jackson et al. 2007).[7] I measure *GPA lower-secondary level* by transforming GPAs into graduation-year percentile rank scores, with 0 indicating the lowest possible score and 100 the highest possible score. I measure *peers' grades* as the mean lower-secondary grade rank in a student's upper-secondary school program cohort (excluding ego from the calculation).

[Table 1 about here]

Table 1 reports descriptive statistics by treatment group and treatment period. The treatment and control groups are rather similar, but the treatment group includes a larger proportion of students with higher-service-class parents, higher grades, and tertiary education at age 26. The outcome variables do not

13

change much over time, but grades and higher-education attainment decline

somewhat for both the treatment and control groups during the study period,

suggesting that the mean effect of the treatment is small.

Results

To provide an overview of the effects of the admission reform on the sorting of

students across schools, I first report in Figure 1 indicators of sorting across schools

cohort by cohort. I measure the degree of sorting in the form of school segregation

in the final graduation year for each cohort, using the dissimilarity index.[8] As Figure

1 shows, the reform led to a dramatic increase in sorting by grades between schools.

Comparing the pre-treatment average segregation to the post-treatment average

segregation in the treatment group, we see that segregation on this dimension

increased from 21.2 to 33.5, close to a 60 percent increase (see Table 2). The

corresponding change in the control group was an increase from 16.6 to 18.8, a 13

percent increase.[9]

[Figure 1 about here]

Surprisingly, however, estimated reform effects on class segregation and

immigrant–native segregation across schools are very close to zero. Judging from

the changes in segregation patterns, many students with high grades used the

opportunity to opt out of their closest school. Given that students with high-SES and

Swedish-born parents typically have higher grades, we would expect the reform to

also increase segregation along these dimensions. Because country-of-birth

segregation and class segregation remained more or less unaltered, it appears that

14

immigrants and students with low-SES parents who had high grades changed schools disproportionately more often.

How did the reform play out in terms of the typical schooling contexts experienced by students? For treatment-group students with below-average grades, peers' grades shifted downward, but a similar albeit less pronounced development also occurred in the control group (see Figure 2a). We see the mirror image of this for students with above-average grades in Figure 2b. The figure shows a density shift to the right, and students in the treatment group ended up in schools with much higher average grades. A similar but much less radical change occurred in the control group. These patterns were very similar regardless of students' ethnicity and class background (results not presented but available on request).

[Figure 2 about here]

[Table 2 about here]

I did not find any systematic treatment effects on the probability of dropping out of the upper-secondary level (28,936 students out of 108,725 did not finish this level "on time"). Conditional on controls, the mean effect on the dropout rate is equal to –.002, suggesting that the reform decreased the dropout rate by .2 of a percentage point. Treatment effects on dropout rates for specific groups are also close to zero: the immigrant dropout rate was unaffected, and dropout rates decreased marginally for low-achievement students and increased marginally for students with a working-class or lower-service-class background (see Table 3).

[Table 3 about here]

I next turn to an analysis of treatment-effect estimates on grades and tertiary educational attainment; Column 1 in Table 4 reports the estimated mean effect on

15

grades. In this and the following models I collapsed the time dimension to just two periods: a pre- and a post-treatment period, but I also report a more flexible specification. The estimated effect, captured by the interaction term *Stockholm × T*2, is positive but very close to zero. When I add controls in the form of GPA rank at the lower-secondary level and background characteristics (column 2), the estimated effect increases in size, but the size of the effect is rather small in magnitude, equal to one-eleventh of a standard deviation of the dependent variable, which roughly corresponds to the effect of moving down four percentile points on the GPA ranking at the lower-secondary level. In a separate analysis, I found that it is the addition of the variable GPA rank at the lower-secondary level that changes the estimate for the treatment effect. The relative difference in GPA rank between the municipality of Stockholm and the county of Stockholm increased over the observation period (see Table 1, final row), to the municipality's advantage, and when I control for this change, the estimated negative effect of the admission reform increases in size. GPA rank at the lower-secondary level is, perhaps unsurprisingly, a very potent predictor of GPA rank at the upper-secondary level, and this alone accounts for 73 percent of the explained variance in the model reported in column 2. In summary, the effect of the reform appears to have been negative for grades, but rather small in magnitude, on average.

Columns 3 to 5 in Table 4 report mean effects on higher-education attainment. The unconditional model indicates the average treatment effect to be positive but very small, equal to a .5 percentage-point increase in the probability of attaining tertiary education, which should be related to the grand mean of this outcome in the population sample, which is 57 percent. When I condition on GPA rank at the lower-secondary level and background characteristics, the estimated

16

effect turns negative. The mechanism is the same here as in the analysis with grades as the dependent variable: lower-secondary mean GPA changed over the study period, to the treatment group's advantage. Once I control for this change, I observe a small negative treatment effect. In the final model, I add GPA rank at the upper-secondary level (the dependent variable in the analysis reported in columns 1 and 2) as a predictor. This attenuates the estimated treatment somewhat, indicating that part of the (small) negative effect is driven by the negative treatment effect on grades at the upper-secondary level. That is, the reform appears to have had a small average negative effect on grades, translating into a small average negative effect on tertiary educational attainment.

[Table 4 about here]

To check whether underlying pre-treatment trends were behind these estimated effects, Figure 3 examines the year-by-year residual gap between treatment and control groups. The figure shows a clear pre-treatment trend of decreasing advantage for the treatment group in relation to both grades and tertiary educational attainment. Some of the treatment-group decline in educational attainment can be attributed to a relative decline in grades (compare the solid black line to the dashed black line). The (small) treatment effects observed in the regression model might be spuriously generated by this underlying trend. Thus, we cannot reject the null of no mean effect of the reform.

[Figure 3 about here]

I next turn to possible heterogeneity across groups in treatment effects. As mentioned earlier, I focus on whether the reform disproportionately affected immigrants, low-achievers, and low-SES students. Figure 4 shows how between-

group gaps developed during the observation period, using the control group as a reference for changes in the treatment group. Positive values indicate that the gap in question increased more in the treatment group than in the control group. Panel A in Figure 4 shows that the correlation between grades at the lower-secondary and upper-secondary levels increased modestly with the treatment, but this did not translate into an increased achievement gap in tertiary educational attainment.

Panel B in Figure 4 reports the development of the class differential. This gap, in contrast, seems to have declined somewhat in the treatment group compared to the control group with regard to both grades and tertiary educational attainment. However, these changes are rather small in magnitude when the pre-treatment average gaps are compared to the treatment-period average gaps.

Finally, the immigrant–native gap fluctuates with regard to grades (see Panel C in Figure 4), but there is no clear increasing or decreasing trend during the observation period. However, the immigrant–native gap grew sharply with regard to tertiary education in the treatment group during the treatment period, and changes in grades at the upper-secondary level cannot explain this change (compare the dashed and solid black lines). In the pre-treatment period, the gap was around two percentage points larger in the treatment group than in the control group. During the treatment period, this difference increased to between five and ten percentage points. For the immigrant–native gap, grades at the upper-secondary level do not explain the change in the tertiary education gaps, indicating that this gap is driven primarily by changes in students' likelihood of making the transition to higher education (i.e., a secondary effect, see Boudon 1974).

[Figure 4 about here]

To obtain more precise estimates of treatment effects on these gaps, I estimated a model that included immigrant status, class-background dummies, and grade rank at the lower-secondary level, by treatment period (two periods) and treatment group. I then calculated the treatment-group change over time in the effect of each variable, and from this I subtracted the control-group change over time in the effect of each variable. More precisely, each value represents the difference-in-differences $\Delta\beta_T - \Delta\beta_C$, where $\Delta\beta_T$ is the treatment-group change in effect between the pre-treatment period and the treatment period, and $\Delta\beta_C$ is the control-group change in effect between the pre-treatment period and the treatment period. These figures tell us how much and in what direction the treatment affected the effect/gap, netting out time-constant group-specific effects, time-variant effects common to both groups, and time-variant compositional changes with regard to class background, immigration status, and grades. The upper panel of Table 5 reports differential effects of the treatment on grades, and the lower panel shows differential effects of the treatment on higher-education attainment.

The first column in the upper panel of Table 5 shows that the estimated treatment effect on the correlation between grades at the lower- and upper-secondary levels is positive but close to zero. This and the group-specific effects are rather small given the size of the coefficient for the total sample (.727, see Table 4). Furthermore, the differentiating effect of the treatment on the correlation between lower-secondary grades and tertiary educational attainment is uniformly equal to, or very close to, zero, irrespective of the controls included (see columns 1, 2, and 3, lower panel). In summary, the treatment has no or only negligible differentiating effects on achievement-group gaps.

19

Consistent with the results presented in Figure 4, there is a small negative effect of the treatment on the class-background differential in grades, and the decline in the class differential is much larger among immigrants. Compared to the specified expectation, the size of the gap between higher-service-class students and unskilled working-class students declined by almost 11 percentile points among immigrants (see column 4, upper panel). This decline appears, to some extent, to be transmitted into a decline in the class differential for tertiary education as well. The net decline in the class differential was 3.3 percentage points, and 5.6 percentage points for immigrants. For immigrants, this decline is entirely explained by grades at the upper-secondary level. That is, the decline in the class–grade gap among immigrants appears to have been transmitted into a decline in the educational attainment–class gap among immigrants (see columns 4 and 5 in the lower panel).

Also consistent with results in Figure 4, the immigrant–native gap with regard to grades is more or less unaffected by the treatment, on average (see column 7 in the upper panel). Immigrants appear to have benefited, however, relatively speaking, among low-achievement students and students with a less-advantaged class background. Working-class/lower-service-class immigrant students improved their grades by 4.4 percentile points relative to natives with the same class background. These improvements were not transmitted into higher transition rates to higher education, however. Compared to the specified expectation, immigrants' average probability of attaining tertiary education declined by 5.4 percentage points. This effect is noticeably smaller among students from working-class homes, but it is about the same for low- and high-achievement students (see column 7, lower panel). Controlling for grades at the upper-secondary level does not alter these estimates, suggesting that the effect is primarily educational choice-related.

The average effect is large in comparison. The gap between immigrants and natives in tertiary educational attainment, conditional on grades, is equal to 6.2 percentage points in the population sample, to the benefit of immigrants. The negative treatment effect, in other words, eradicated most of this particular gap.

To investigate potential mechanisms in terms of school composition changes that were a consequence of the reform, the next step in the analysis includes peers' achievement level in the equations, because there is an obvious link in the literature between this dimension and potential peer effects of ability grouping; moreover, ability sorting between schools was *the* kind of sorting affected by the reform (see Figure 1). To take possible nonlinear effects into account, I included a nonparametric specification of decile indicators of school-program-cohort mean entry grades in the model. Columns 3, 6, and 9 in Table 5 report these results. In contrast to the expectation that effects should be attenuated once this dimension is taken into account, the estimated effects are not much affected by inclusion of this variable. It appears that the estimated treatment effect is not mediated by school-achievement composition. As a sensitivity test, I redefined peers into gender-immigrant-class cells within schools, but this did not alter the estimated effects.

[Table 5 about here]

Discussion

I conducted a rigorous test of whether and in what direction ability grouping at the upper-secondary level affects students' grades and their future attainment of higher education. Previous studies typically consider only single outcomes, usually short-

21

term outcomes, making assessments of the overall effectiveness of ability grouping difficult (cf. Argys, Rees and Brewer 1996, Ireson and Hallam 1999). A more serious flaw in many previous studies, however, is a lack of attention to issues of unobserved selection into achievement-based groups, and the likely bias this gives rise to.

I used a reform of the admissions process into schools in one municipality as an exogenous source of variation in student ability grouping across schools. I evaluated consequences of this change using students in neighboring municipalities—where the admission system did not change—as a comparison. Because curricula and schools were identical before and after implementation of the new admission regime, I am in a better position to isolate the causal effect of ability grouping, netting out the effects of observed and unobserved selection into ability groups on outcomes, as well as the potential confounding effects of curricular heterogeneity. The reform was followed by a 60 percent increase in ability sorting across schools, but sorting by ethnicity and class background was not affected. In the present case, the change in ability grouping across schools occurred in isolation from sorting dimensions that usually go hand-in-hand with this type of sorting, giving us the opportunity to isolate the effects of ability grouping more convincingly.

Mean effects of ability grouping are usually close to zero but can, in theory, increase or decrease educational outcome differences between groups, depending on the relative importance of peer effects (Owens 2010, Sacerdote 2001) and social comparison effects (Marsh 1991, Mood and Jonsson 2008). Based on a review of the empirical literature on ability grouping, I expected grade gaps would widen as a consequence of increased ability grouping. With regard to higher-education

attainment gaps, I expected widened grade gaps to translate into widened educational attainment gaps, thus increasing the primary effect of any background variable. At the same time, the secondary choice effect might close attainment gaps, because social comparison effects might discourage ambitious educational choices among more privileged groups (cf. Boudon 1974). Furthermore, I expected students who were disadvantaged in more than one dimension (e.g., immigrant students with low grades) to be more vulnerable to an increase in ability grouping.

As expected, the mean effects of increased ability grouping on grades and tertiary educational attainment were very small in magnitude, and these small effects might have occurred in the absence of treatment. Counter to my expectations, increased ability grouping appears to have only marginally affected grade gaps between groups defined by country of birth and achievement. However, contrary to my predictions, the class gap decreased somewhat, and among immigrants it decreased substantially.

Effects on higher-education attainment gaps went in different directions. The immigrant–native gap in tertiary educational attainment increased substantially with increased sorting. However, the class-background gap in educational attainment decreased somewhat, more so among immigrants, and there was no effect on achievement-group gaps. Effects on the immigrant and class gaps are secondary choice effects: they could not be accounted for by upper-secondary grades. Finally, neither mean nor group-specific dropout rates from the upper-secondary level were affected by increased ability grouping.

In summary, the short-term effects of ability grouping on grades were rather small, but there was a tendency toward a decrease in the class-background gap, and

I found long-term effects that hurt immigrant students' educational attainment but benefited working-class students. The results are consistent with previous research on ability grouping—zero or close to zero mean effects of ability grouping are a common finding (Hattie 2002, Hattie 2009, Kao and Thompson 2003, Slavin 1990). With regard to ability grouping's effect on gaps, the results point in different directions, and we can draw no clear conclusions with regard to the effects of ability grouping on longer term stratification outcomes. Contrary to my expectation, I found no evidence that students who were disadvantaged on more than one dimension experienced any additional loss. In fact, immigrant students with a less-advantaged class background tended to fare better than other groups in the new admission regime.

The decreased class gap in higher-education attainment among immigrants was the only gap change that could be explained as a primary effect of changing gaps in grades at the upper-secondary level (cf. Boudon 1974). However, the mean increase in the immigrant gap cannot be attributed to grades and appears to have been caused primarily by changes in transition probabilities into higher education driven by factors other than grades (i.e., secondary effects). This is the largest estimated effect of the reform, both in absolute terms and in relation to the underlying baseline gaps. Although a secondary choice effect dominated over a primary grade effect, I cannot, of course, rule out the possibility that the informal curricula and learning environment in schools changed in such a way as to benefit some students while leaving others behind. Such a development may have failed to show up in student grades because grades, at least partially, are set relative to other students within schools. When immigrant students from low-achievement schools

24

entered the tertiary level, their prior academic knowledge might, on average, have been too low to succeed at this level.

Such instruction effects may have been generated as a consequence of changing the composition of students, because teachers adjust their expectations and standards to their students' average achievement. Many teachers left schools that experienced an increased inflow of low-achieving students (Karbownik 2014), which likely negatively affected the quality of teaching in these schools. Speaking against such a conjecture is the fact that school grade composition could not in any way account for the estimated treatment effects. This pattern undoubtedly poses a broader challenge for research. If there is any truth to theories on the effects of ability grouping, at least some of the estimated reform effects should be accounted for by peers' achievement levels.

The main strength of this study is its high internal validity, but there are some obvious external validity limitations. The data are limited in time and space to Stockholm at the turn of the millenium, an affluent city in a well-developed welfare state. Compared to the United States, Sweden has less segregation and less variation in school quality, and higher education is tuition free. In light of these institutional differences, the results, in an international comparison, likely represent estimates in the lower range of the effects of increasing ability grouping across schools. One of the main goals of Swedish educational policy has been to decrease the correlation between social background factors and educational attainment. Inequality of educational opportunity, and social inequality in general, is relatively limited in Sweden (Gornick and Jäntti 2013). Any negative effects of increasing ability grouping across schools is probably buffered by these policies, suggesting that any

effects of school choice and ability grouping are probably higher in countries and institutional contexts with more variation in standards across schools, and where families bear more of the costs associated with attainment of higher education. Relatedly, immigrant students, who obviously have not benefited from Swedish welfare institutions from birth, appeared to be negatively affected by the increase in ability grouping.

Bearing this in mind, the main take-home point of this research is that ability grouping appears to have negligible mean effects and unclear differential effects on educational outcomes. Most previous research finds that low-achievement students and students from low-resource families are more vulnerable to a downward shift in peers' average achievement, because these students lack resources at home to compensate for conditions at school (e.g., Hattie 2009, Jackson 2009, Kao and Thompson 2003). However, results from more well-designed studies do not conform to this pattern and report zero or even reversed differential effects (Betts and Shkolnik 2000, Figlio and Page 2002). The present study adds to this pattern, suggesting that inadequate control groups in prior research have given rise to a misperception that ability grouping is harmful for disadvantaged students.

Research Ethics

My research protocol was reviewed and approved by the Regional Ethical Review Board in Stockholm. The research involves an analysis of register data that do not include personal identifiers. Because of the nature of the data, informed consent from subjects has not been obtained, but all results are presented in such a way that no single individual can be identified.

Acknowledgments

Table 1: Means (standard deviations) of variables used in the analyses, by treatment group and treatment period.

Variable name	Definition	Statistics Sweden Source	Treatment group, pre-treatment period[a]	Treatment group, post-treatment period[b]	Control group, pre-treatment period[c]	Control group, post-treatment period[d]
Grades	Within-graduation-year GPA percentile rank (at age 19).	The pupil register	54.224 (29.392)	53.934 (29.770)	48.082 (28.218)	47.515 (28.612)
Tertiary educational attainment	At least one semester of tertiary education seven years after upper-secondary graduation (at age 26).	The LISA register	0.629	0.618	0.552	0.536
Immigrant	Individual and both parents born outside of Sweden.	The background data register	0.110	0.109	0.101	0.093
Class background	Highest recorded (EGP) class of any parent (in 1980, 1985, or 1990).	Population censuses				
	Unskilled worker		0.066	0.077	0.085	0.095
	Skilled worker		0.075	0.082	0.106	0.124
	Lower service		0.113	0.111	0.125	0.129
	Middle service		0.224	0.220	0.246	0.232
	Higher service		0.343	0.313	0.271	0.243
	Self employed		0.063	0.063	0.074	0.071
	Unknown/missing		0.116	0.134	0.093	0.106
GPA rank lower-secondary level	Within-cohort GPA percentile rank in 9th grade (at age 16).	The pupil register	61.347 (27.502)	64.949 (26.918)	57.635 (26.636)	57.268 (26.519)
Peers' grades	Upper-secondary school-program-cohort mean of the GPA rank at lower-secondary level (excluding ego)	The pupil register	62.246 (18.140)	64.658 (20.724)	58.534 (16.277)	57.190 (17.107)

[a] Entered upper-secondary schools in the municipality of Stockholm, in the period 1997 to 1999, n = 13,863.

[b] Entered upper-secondary schools in the municipality of Stockholm, in the period 2000 to 2002, n = 16,155.

[c] Entered upper-secondary schools in the county of Stockholm, excluding the municipality of Stockholm, in the period 1997 to 1999, n = 23,022.

[d] Entered upper-secondary schools in the county of Stockholm, excluding the municipality of Stockholm, in the period 2000 to 2002, n = 25,980.

Table 2: Treatment (admission reform) effects on school segregation

Segregation dimension	Treatment group		Control group		Difference-in-Differences[a]
	Pre-treatment period	Treatment period	Pre-treatment period	Treatment period	
Entrance grades	21.2	33.5	16.6	18.8	+10.1
Class background	14.8	16.6	13.2	16.0	-1.0
Country of birth	19.2	19.1	10.3	10.2	0.0

Note: Multigroup segregation indices (Reardon and Firebaugh 2002) of systematic segregation (Carrington and Troske 1997) for pre-treatment period (school entrance years 1997 to 1999) and treatment period (school entrance years 2000 to 2002). Class includes 7 categories, grades 20 categories (based on the percentile distribution of grades in 9th grade), and country of birth 28 categories.

[a] (School segregation in the treatment group after reform – School segregation in the treatment group before reform) – (School segregation in the control group after reform – School segregation in the control group before reform).

Table 3: Unconditional and conditional admission reform effects on dropout rates from the upper-secondary level

	1. Mean effect	2. Conditional mean effect	3. Conditional mean effect, sample restricted to students with below-median grades (in 9th grade)	4. Conditional mean effect, sample restricted to students with working-class or lower-service-class background	5. Conditional mean effect, sample restricted to immigrants only
Stockholm	0.012	0.022	0.030	0.013	0.030
$T2$	-0.075	-0.071	-0.098	-0.097	-0.083
Stockholm$T2$	-0.016	-0.002	-0.010	0.011	0.000
Controls	NO	YES	YES	YES	YES
R-squared	0.009	0.225	0.164	0.219	0.219
N	108,725	108,725	65,324	36,723	13,493

Note: OLS regression estimates (linear probability model estimates) of dropping out of the upper-secondary level on independent variables. Controls included in the regression model: GPA rank lower-secondary level (continuous), class background (seven dummies), and an immigrant indicator.

Table 4: Unconditional and conditional admission reform effects on grades at the upper-secondary level at age 19, and attainment of tertiary education at age 26

|	Dependent variable: Grades (age 19)		Dependent variable: At least one semester of tertiary education (age 26)		
	1	2	3	4	5
Stockholm	6.142	3.106	0.077	0.033	0.018
$T2$	-0.567	-0.099	-0.016	-0.007	-0.006
Stockholm$T2$	0.277	-2.613	0.005	-0.030	-0.018
GPA rank lower-secondary level		0.727		0.008	0.005
Class background (ref.: unskilled worker)					
Skilled worker		1.005		-0.005	-0.010
Lower service		2.678		0.026	0.013
Middle service		3.884		0.103	0.084
Higher service		7.701		0.183	0.146
Self-employed		3.083		0.053	0.038
Unknown		2.148		0.087	0.077
Immigrant		-3.223		0.049	0.065
GPA rank secondary level					0.005
Constant	48.082	2.595	0.552	-0.032	-0.044
R-squared	0.011	0.509	0.006	0.268	0.308
N	79,020	79,020	79,020	79,020	79,020

Note: Column 1 and 2: OLS regression estimates of GPA grade rank at upper-secondary graduation on independent variables. Column 3, 4, and 5: OLS regression estimates (linear probability model estimates) of tertiary educational attainment at age 26 on independent variables.

Table 5: Treatment effects on gaps in grades at the upper-secondary level and gaps in the attainment of tertiary education

	(Treatment group change in lower-secondary GPA coefficient) –(Control group change in lower-secondary GPA coefficient)			(Treatment group change in higher-service-class coefficient) – (Control group change in higher-service-class coefficient)			(Treatment group change in immigrant coefficient) – (Control group change in immigrant coefficient)		
	1	2	3	4	5	6	7	8	9
Dependent variable: Grades at the upper-secondary level (rank 0-100)									
Conditional on peers' grades	NO		YES	NO		YES	NO		YES
All (*n* = 79,020)	0.025		0.003	-1.394		-1.663	-0.055		-0.118
Immigrant students (*n* = 8,029)	0.022		-0.042	-10.956		-11.060			
Low-achievement students (*n* = 39,178)	0.033		0.051	-0.764		-1.227	1.320		1.342
Working-class/lower-service-class students (*n* = 24,182)	-0.009		-0.005				4.410		5.133
Dependent variable: Tertiary education									
Conditional on grades at upper-secondary level	NO	YES	YES	NO	YES	YES	NO	YES	YES
Conditional on peers' grades	NO	NO	YES	NO	NO	YES	NO	NO	YES
All (*n* = 79,020)	0.000	0.000	-0.001	-0.033	-0.030	-0.029	-0.054	-0.053	-0.052
Immigrant students (*n* = 8,029)	0.001	0.000	0.001	-0.056	-0.001	0.036			
Low-achievement students (*n* = 39,178)	0.000	0.000	-0.001	-0.032	-0.030	-0.030	-0.058	-0.066	-0.069
Working-class/lower-service-class students (*n* = 24,182)	0.000	0.000	-0.001				-0.001	-0.021	-0.030

Note: Each value represents the difference-in-differences $\Delta\beta_T - \Delta\beta_C$, where $\Delta\beta_T$ is the treatment-group change in effect between the pre-treatment period and the treatment period, and $\Delta\beta_C$ is the control-group change in effect between the pre-treatment period and the treatment period. The following covariates are included as controls, within period and group: GPA rank lower-secondary level (continuous), class background (seven dummies), and an immigrant indicator (see Table 1 for definitions). The class-background differential is the difference between students with higher-service-class parents and students with unskilled working-class parents.

Figure 1: Degree of sorting across schools, with regard to entrance grades, class background, and country-of-birth background, before and after admission reform in 2000 in the municipality of Stockholm, compared to neighboring municipalities
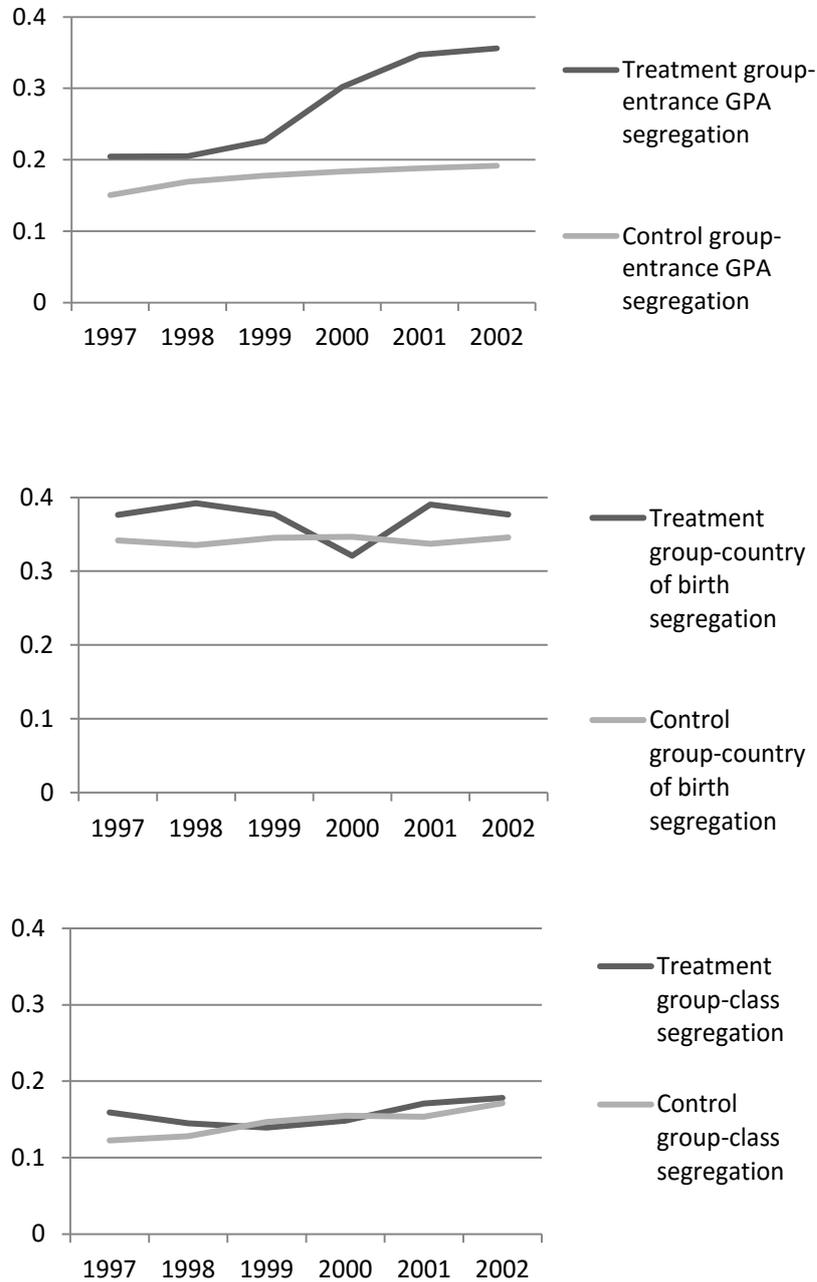
Note: Panels show the multigroup index (Reardon and Firebaugh 2002) of systematic segregation (Carrington and Troske 1997) for each year in the pre-treatment period (school entrance years 1997 to 1999) and the treatment period (school entrance years 2000 to 2002). Class includes 7 categories, grades 20 categories (based on the percentile distribution of grades in 9th grade), and country of birth 28 categories.

Figure 2a: Density graphs of upper-secondary school cohort mean-entry GPA for students with below-median GPA at the lower-secondary level, by treatment group and treatment period
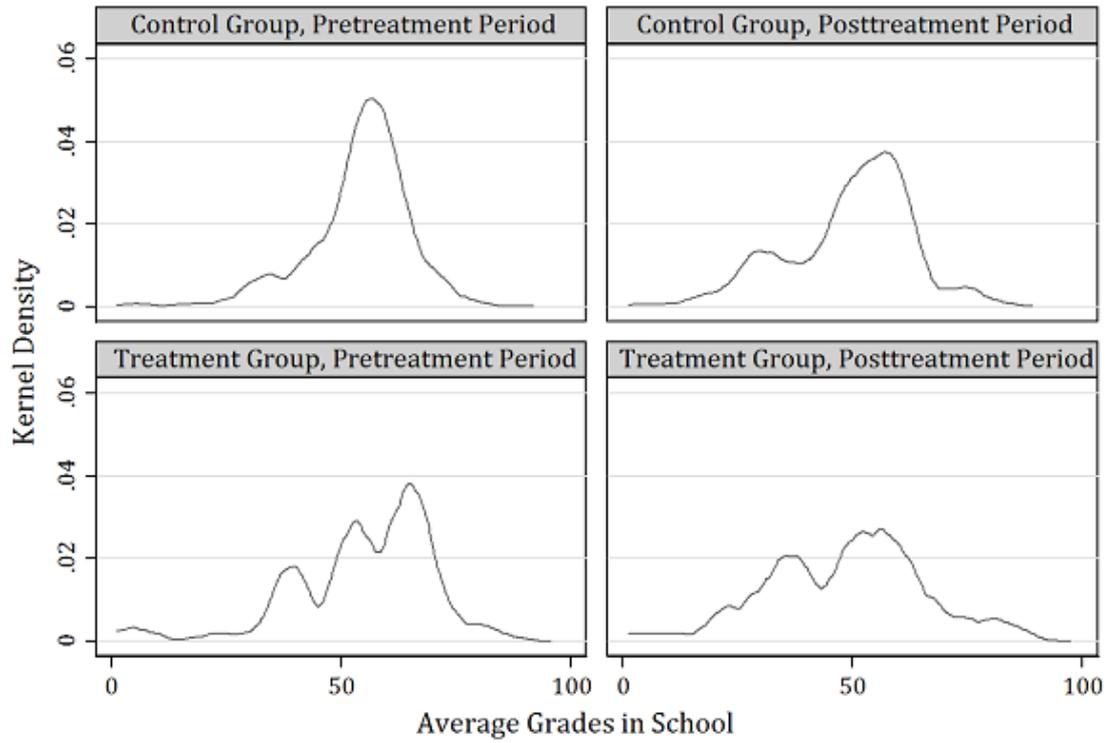
Figure 2b: Density graphs of upper-secondary level school cohort mean entry GPA for students with above-median GPA at the lower-secondary level, by treatment group and treatment period
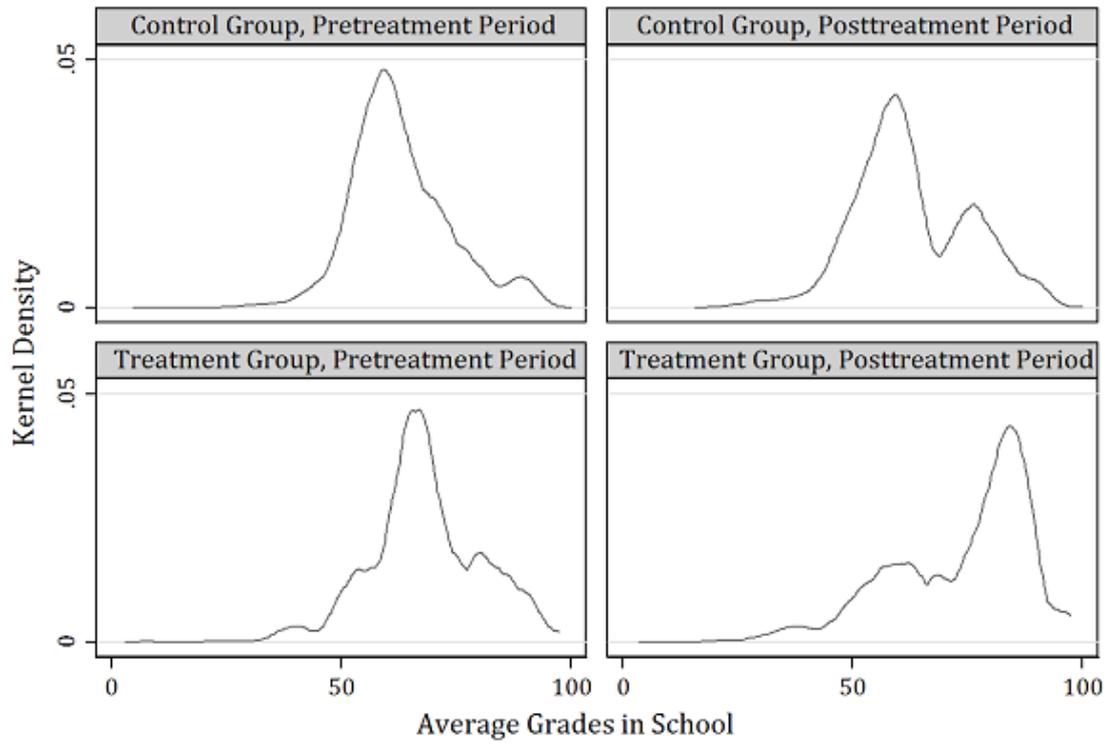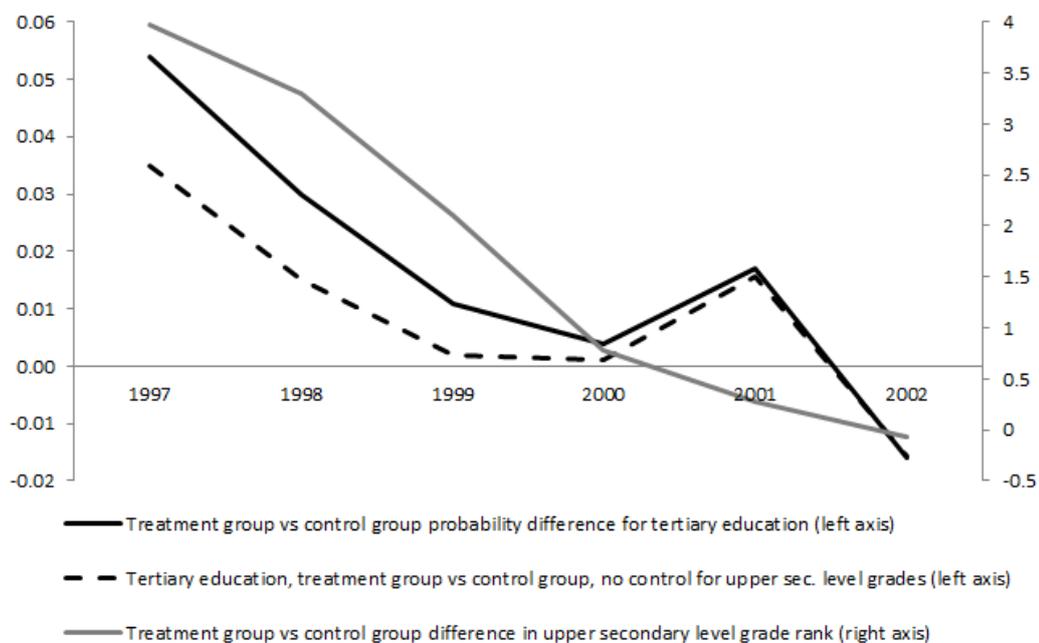
Figure 3: Differences in mean educational attainment at age 26 and mean grades at age 19 between treatment and control groups, before (enrollment years 1997 to 1999) and after (enrollment years 2000 to 2002) treatment

Note: Each data point represents a treatment-group coefficient from a year-specific OLS regression model with the following covariates included: GPA rank lower-secondary level (continuous), class background (seven dummies), and an immigrant indicator (see Table 1 for definitions). The dashed line is also conditioned on upper-secondary level GPA rank (continuous).

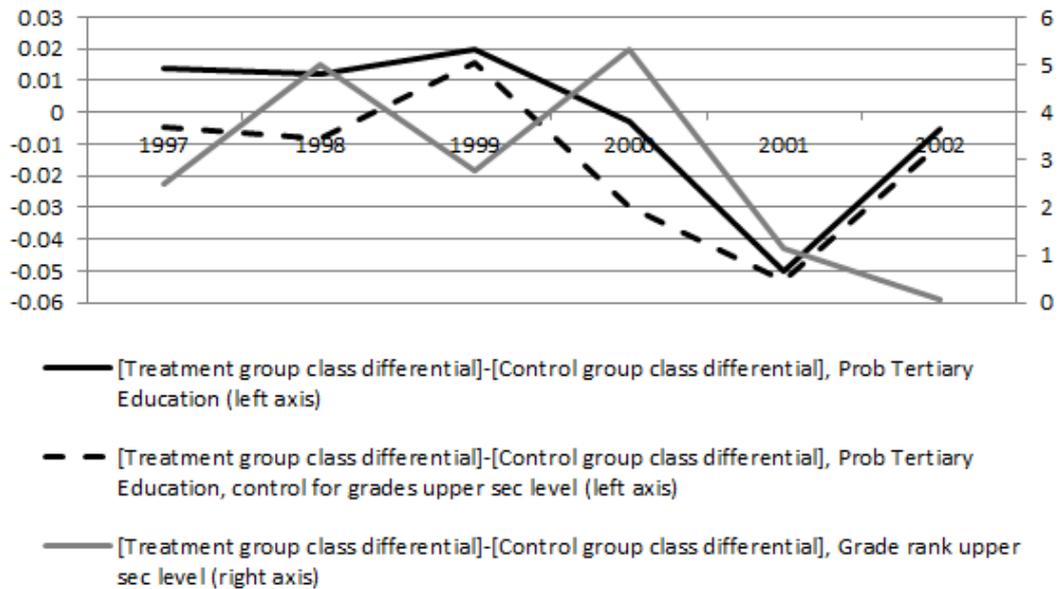**Figure 4:** Year-by-year gap (by immigrant status, class background, and achievement) differences between treatment and control groups in mean educational attainment and mean grades, before (enrollment years 1997 to 1999) and during (enrollment years 2000 to 2002) the treatment period

A

———[Treatment group 9th grade GPA effect differential]-[Control group 9th grade GPA effect differential], Prob Tertiary Education (right axis)

– – [Treatment group 9th grade GPA effect differential]-[Control group 9th grade GPA effect differential], Prob Tertiary Education, control for grades upper sec level (right axis)

———[Treatment group 9th grade GPA effect differential]-[Control group 9th grade GPA effect differential], Grade rank upper sec level (left axis)

B

———[Treatment group class differential]-[Control group class differential], Prob Tertiary Education (left axis)

– – [Treatment group class differential]-[Control group class differential], Prob Tertiary Education, control for grades upper sec level (left axis)

———[Treatment group class differential]-[Control group class differential], Grade rank upper sec level (right axis)

C

Legend:

——— [Treatment group immigrant differential]-[Control group immigrant differential], Prob Tertiary Education (left axis)

– – – [Treatment group immigrant differential]-[Control group immigrant differential], Prob Tertiary Education, control for grades upper sec level (left axis)

——— [Treatment group immigrant differential]-[Control group immigrant differential], Grade rank upper sec level (right axis)
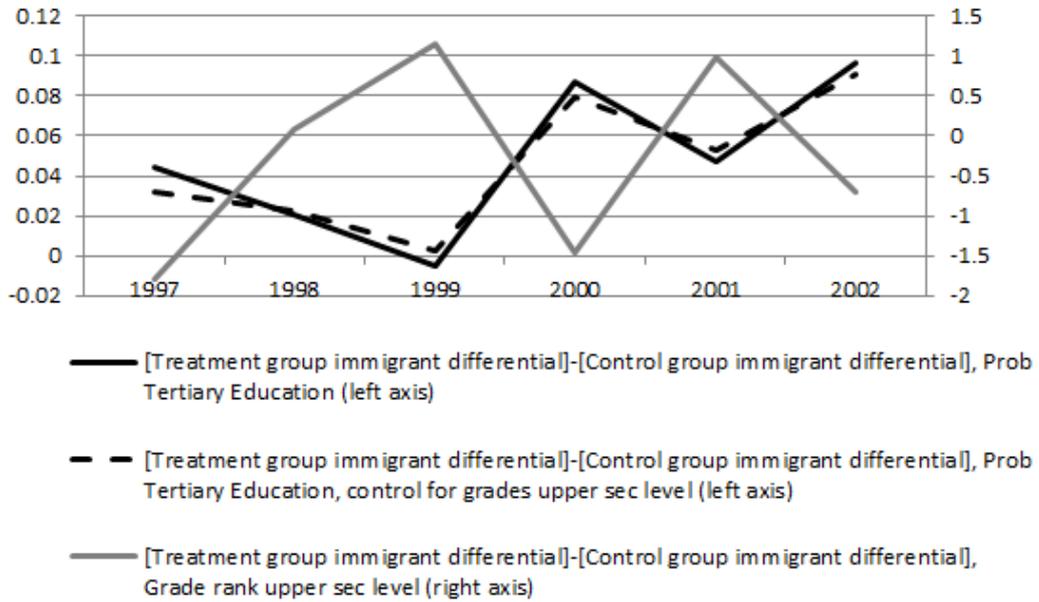
Note: Each data point represents a coefficient from a year- and treatment-group-specific OLS regression model with the following covariates included: lower-secondary level GPA rank (continuous), class background (six dummies), and an immigrant indicator (see Table 1 for definitions). The 9th-grade GPA-effect differential (Panel A) is the GPA coefficient on the outcome. The class-background differential (Panel B) is the coefficient difference between students with higher-service-class parents and students with unskilled working-class parents on the outcome. The immigrant differential (Panel C) is the coefficient difference between immigrant students with both parents born abroad and students born in Sweden with at least one parent born in Sweden on the outcome.

References

Alba, Richard, Jennifer Sloan and Jessica Sperling. 2011. "The Integration Imperative: The Children of Low-Status Immigrants in the Schools of Wealthy Societies." *Annual Review of Sociology* 37(1):395-415.

Alexander, Karl L., Martha Cook and Edward L. McDill. 1978. "Curriculum Tracking and Educational Stratification: Some Further Evidence." *American Sociological Review* 43(1):47-66.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Argys, Laura M., Daniel I. Rees and Dominic J. Brewer. 1996. "Detracking America's Schools: Equity at Zero Cost?". *Journal of Policy Analysis and Management* 15(4):623-45.

Betts, Julian R. and Jamie L. Shkolnik. 2000. "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools." *Economics of Education Review* 19(1):1-15.

Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.

Brunello, Giorgio and Daniele Checchi. 2007. "Does School Tracking Affect Equality of Opportunity? New International Evidence." *Economic Policy* 22(52):781-861.

Bygren, Magnus. 2013. "Unpacking the Causes of Segregation across Workplaces." *Acta Sociologica* 56(1):3-19.

Carrington, William J and Kenneth R Troske. 1997. "On Measuring Segregation in Samples with Small Units." *Journal of Business & Economic Statistics* 15(4):402-09.

Davis, James A. 1966. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *The American Journal of Sociology* 72(1):17-31.

Figlio, David N. and Marianne E. Page. 2002. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?". *Journal of Urban Economics* 51(3):497-514.

Gamoran, Adam. 1987. "The Stratification of High School Learning Opportunities." *Sociology of Education* 60(3):135-55.

Gornick, Janet C. and Markus Jäntti, eds. 2013. *Income Inequality : Economic Disparities and the Middle Class in Affluent Countries*. Stanford, CA: Stanford University Press.

Hattie, John A. C. 2002. "Classroom Composition and Peer Effects." *International Journal of Educational Research* 37(5):449-81.

Hattie, John A. C. 2009. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Hoffer, Thomas B. 1992. "Middle School Ability Grouping and Student Achievement in Science and Mathematics." *Educational Evaluation and Policy Analysis* 14(3):205-27.

Ireson, Judith and Susan Hallam. 1999. "Raising Standards: Is Ability Grouping the Answer?". *Oxford Review of Education* 25:343-58.

Jackson, C. Kirabo. 2009. *Ability-Grouping and Academic Inequality: Evidence from Rule-Based Student Assignments*: National Bureau of Economic Research Working Paper Series No. 14911.

Jackson, Michelle, Robert Erikson, John H. Goldthorpe and Meir Yaish. 2007. "Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to a-Level Courses in England and Wales." *Acta Sociologica* 50(3):211-29.

Kao, Grace and Jennifer S. Thompson. 2003. "Racial and Ethnic Stratification in Educational Achievement and Attainment." *Annual Review of Sociology* 29:417-42.

Karbownik, Krysztof. 2014. *Do Changes in Student Quality Affect Teacher Mobility?* Uppsala: Institute for Evaluation of Labour Market and Education Policy (IFAU) Working Paper 2014:15.

Kerckhoff, Alan C. 1986. "Effects of Ability Grouping in British Secondary Schools." *American Sociological Review* 51(6):842-58.

Lechner, Michael. 2011. "The Estimation of Causal Effects by Difference-in-Difference Methods." *Foundations and Trends in Econometrics* 4(3):165-224.

LeTendre, Gerald K., Barbara K. Hofer and Hidetada Shimizu. 2003. "What Is Tracking? Cultural Expectations in the United States, Germany, and Japan." *American Educational Research Journal* 40(1):43-89.

Marsh, Herbert W. 1991. "Failure of High-Ability High Schools to Deliver Academic Benefits Commensurate with Their Students' Ability Levels." *American Educational Research Journal* 28(2):445-80.

Mood, Carina and Jan O. Jonsson. 2008. "Choice by Contrast in Swedish Schools: How Peers' Achievement Affects Educational Choice." *Social Forces* 87(2):741-65.

Musset, Pauline. 2012. "School Choice and Equity: Current Policies in Oecd Countries and a Literature Review." Vol. *OECD Education Working Papers.* OECD Publishing.

Oakes, Jeannie and Gretchen Guiton. 1995. "Matchmaking: The Dynamics of High School Tracking Decisions." *American Educational Research Journal* 32(1):3-33.

Owens, Ann. 2010. "Neighborhoods and Schools as Competing and Reinforcing Contexts for Educational Attainment." *Sociology of Education* 83(4):287-311.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32(1):33-67.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics* 116(2):681-704.

Saleh, Mohammad, Ard W. Lazonder and Ton De Jong. 2005. "Effects of within-Class Ability Grouping on Social Interaction, Achievement, and Motivation." *Instructional Science* 33(2):105-19.

Seaton, Marjorie, Herbert W. Marsh, Florence Dumas, Pascal Huguet, Jean-Marc Monteil, Isabelle Régner, Hart Blanton, Abraham P. Buunk, Frederick X. Gibbons, Hans Kuyper, Jerry Suls and Ladd Wheeler. 2008. "In Search of the Big Fish: Investigating the Coexistence of the Big-Fish-Little-Pond Effect with the Positive Effects of Upward Comparisons." *British Journal of Social Psychology* 47(1):73-103.

Skolverket. 2006. "Vad Händer Med Likvärdigheten I Svensk Skola? [What Is Happening to Equivalence in Swedish Schools?]." Stockholm: Rapport 275.

Slavin, Robert E. 1987. "Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 57(3):293-336.

Slavin, Robert E. 1990. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 60(3):471-99.

Söderström, Martin. 2006. *School Choice and Student Achievement - New Evidence on Open-Enrollment*: IFAU Working Paper 2006:16.

Söderström, Martin and Roope Uusitalo. 2010. "School Choice and Segregation: Evidence from an Admission Reform." *Scandinavian Journal of Economics* 112(1):55-76.

Terwel, Jan. 2005. "Curriculum Differentiation: Multiple Perspectives and Developments in Education." *Journal of Curriculum Studies* 37(6):653-70.

USK. 2002. "Förändrad Elevsammansättning På Gymnasieskolan I Samband Med Ändrad Intagningsprincip [Change in Student Composition Following Change in Admission Principle]." Stockholm: Utrednings- och statistikkontoret (USK), Stockholms stad.

USK. 2006. "Har Segregationen Ökat I Stockholms Stad? [Has Segregation Increased in the Municipality of Stockholm?]." Stockholm: Utrednings- och statistikkontoret (USK), Stockholms stad.

Endnotes

---

[1] Program content is determined at the national level: the Swedish Education Act (Skollagen) is decided by the Swedish parliament and details the curricula of upper-secondary programs.

[2] The expected effects of the reform on the distribution of students across schools are dependent on a number of factors: the relationship between demographic/social groups and achievement, preferences for schools across these groups, distribution of schools across residential neighborhoods, and the degree of residential sorting across groups. School choice reforms tend to increase sorting across schools on many dimensions (Musset 2012), and the distribution of students across schools on unobserved student characteristics may have changed simultaneously with the reform.

[3] Additional important assumptions for causal inference are that the treatment does not affect the control variables, and that the treatment does not affect the pre-treatment population in the pre-treatment period, that is, through anticipatory effects (Lechner 2011). By definition, the treatment cannot affect students' background characteristics. However, the treatment might have affected student grades at the lower-secondary level, as grades became important for school placement with the reform.

[4] The linear probability model has two well-known weaknesses: it may produce probability predictions outside the interval 0 to 1, and the error term violates the linear regression assumption of homoscedasticity. As my main purpose in performing the regressions is to estimate the mean effects of the variables, I do not consider the first weakness a serious one, and the second weakness has no serious implications, because I do not conduct significance tests.

[5] As a sensitivity test, I used untransformed grades as a measure of grades at the upper-secondary level. This variable correlates .91 with actual grades, and all results are very similar using actual grades rather than grade rank in the models.

[6] I derived this measure of tertiary educational attainment from the Swedish standard classification of education. I tested the sensitivity of the results using continuous and ordinal measures of educational attainment. The same overall patterns emerged irrespective of the measure used. To make interpretations as intuitive as possible, I used a dichotomous measure.

[7] If father's class is missing for that year, I measure father's class in 1985 or 1980. I similarly measure mother's class in 1990, or 1985 and 1980 if class is missing for 1990. If father's class is missing for all years, mother's class replaces father's as the individual's class background. Also, if the mother's class is higher than the father's,

in the sense that her class position, on average, is more skilled, and both mother and father are salaried employees, the mother's class replaces the father's as the individual's class background. I have no information on class background for immigrants who came to Sweden after 1990; I assigned these and others whose class background I could not determine to a missing category to retain them in the analyses.

[8] I adjust D for the small unit upward bias (see Bygren 2013). I conceptualize segregation as systematic segregation, which is the segregation observed in excess of the segregation level we would expect given a random allocation of students to schools. Let D be the observed dissimilarity index, and let D^* be $E(D)$ generated by a random allocation of students to schools. The index of systematic dissimilarity, $\widehat{D}$, is defined as $\frac{D-D^*}{1-D}$ if $D>D^*$. I use the multigroup index (Reardon and Firebaugh 2002). For class, I use 7 categories, for grades 20 categories (based on the percentile distribution of grades), and for country of birth 28 categories.

[9] The increase in the control group aligns with previous studies documenting increased grade sorting between schools in Sweden during the period of study (e.g., Skolverket 2006).