IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12927

The Effects of Student Composition on Teacher Turnover: Evidence from an Admission Reform

Krzysztof Karbownik

JANUARY 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12927

The Effects of Student Composition on Teacher Turnover: Evidence from an Admission Reform

Krzysztof Karbownik
Emory University and IZA

JANUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Effects of Student Composition on Teacher Turnover: Evidence from an Admission Reform*

This paper examines the effects of student ability on teacher turnover using data from Stockholm high schools and an admission reform that led to the exogenous reshuffling of pupils. The results indicate that a 10-percentile-point increase in student credentials decreases the probability of a job separation by up to 10 percentage points. These effects vary somewhat across different groups of teachers and are found mainly for mobility between schools rather than out of the profession. Teachers react most strongly to direct measures of student ability, grades from compulsory school, rather than to other correlated characteristics such as immigrant origin or parental income.

JEL Classification: I2, J2, J63

Keywords: teacher mobility, student ability, school choice

Corresponding author:
Krzysztof Karbownik
Department of Economics
Emory University
Rich Memorial Building
1602 Fishburne Drive
30307 Atlanta, Georgia
USA
E-mail: krzysztof.karbownik@emory.edu

* I would like to thank Per-Anders Edin, David Figlio, Hans Grönqvist, Jon Guryan, Kirabo Jackson, Mikael Lindahl, Björn Öckert and Erik Plug for helpful comments and guidance. Special thanks go to Ijun Lai, Magda Mazurkiewicz, Jörgen Moen and Anthony Wray, whose help with editorial and data issues have been invaluable. This work has benefited from the comments of seminar participants at IFAU, Northwestern University, DIW Berlin, Upjohn Institute for Employment Research, EEA 2012, EALE 2012, WIEM 2012, ESPE 2013. I am grateful to IFAU for providing me with access to their data. I appreciate the support of Hedelius Scholarship that permitted me to conduct parts of this research while visiting Northwestern University. All errors or omissions are my own.

1 Introduction

Educational interventions such as student busing or school choice change the composition of pupils in schools. These interventions have been motivated by the idea that certain groups of students might benefit from meeting better peers. However, sometimes the policymakers simply want to put disadvantaged students into better schools. Irrespective of the motivation, the policies assume other inputs of the education production function are held constant, and thus rely heavily on their exogeneity with respect to student characteristics (Jackson 2009).[1] It is plausible, however, that changes in student composition affect other input factors such as teacher composition or school resources (Hanushek 1986). Consequently, policies aimed at improving performance may have unintended consequences.

Establishing the causal relationship between student ability and teacher composition should be of interest for two policy reasons. First, if students with lower aptitude induce teachers to leave their schools, then an inflow of less able students may be reinforced by higher teacher turnover and by unfavorable teacher sorting (Ronfeldt et al. 2013). Second, the potentially positive effects of policies aimed at reshuffling students between schools may be dwarfed by teacher mobility if highly productive teachers leave in response to an inflow of low aptitude pupils.

This paper documents how exogenous changes in student composition affect teacher turnover. In particular, I investigate whether teachers who experience an inflow of high achieving students are less likely to quit their jobs compared with teachers who face an inflow of lower aptitude students.[2] Some correlational studies suggest that pupil credentials are inversely related to teacher mobility, but we know relatively little about whether this descriptive relationship can be given a causal interpretation, and endogeneity can arise for multitude of reasons.[3] For example, teachers with strong preferences for student ability may sort into schools with high performing students. Using an ordinary OLS in such case would generate a biased negative correlation between student ability and teacher turnover.

[1]Examples of policies that lead to reshuffling of students include: increased freedom in school choice (Cullen et al. 2006); school voucher programs (Hsieh and Urquiola 2006; Epple et al. 2017); student busing (Jackson 2009); increased competition from the private sector (Jackson 2012; Hensvik 2012); changes in school admission policies (Söderström and Uusitalo 2010); or court-ordered desegregation (Reber 2005). Desegregation and busing have been shown to increase the fraction of minority students in traditionally white schools (Reber 2005; Jackson 2009) while increased competition from private sector lead to "cream skimming" of the best students from public schools in Chile (Hsieh and Urquiola 2006) but not in Sweden where it was not allowed (Böhlmark and Lindahl 2015). Crucially for this paper expanded school choice at high school level in Stockholm lead to increased segregation by ability (Söderström and Uusitalo 2010). This, particular variation can be thought of as a "first-stage" to teacher mobility question examined in the current paper.

[2]Throughout the paper I interchangeably use terms: credentials, aptitude, achievement, performance, ability or GPA. These all refer to outgoing primary school percentiled GPA which is based on examination and teacher assessment. Details on the construction of this measure are provided in Section 2.2 and in the Online Appendix.

[3]Examples of descriptive studies on teacher mobility include Hanushek et al. (2004) for Texas, Falch and Strøm (2005) for Norway, Scafidi et al. (2007) for Georgia (US), Barbieri et al. (2011) for Italy, or Karbownik (2014) for Sweden. In the US both referenced studies show strong positive correlations between higher shares of minority students and teacher turnover, a finding that was confirmed in Norwegian and Italian but to a lesser degree in Swedish data. Furthermore, both Hanushek et al. (2004) and Barbieri et al. (2011) find negative correlation between student ability and teacher turnover while in the case of Scafidi et al. (2007) this association is dominated by racial composition of schools. Since student ability is correlated with student disadvantage and other observable and unobservable characteristics, it is thus crucial to obtain quasi-exogenous variation in the variable of interest to disentangle this endogeneity.

Furthermore, since aptitude is correlated with a variety of family characteristics, it is impossible to disentangle the "effect" of ability from, say, parental education in such an analysis. Due to high potential for sorting and policy relevance of teacher turnover, it appears crucial to understand the causal channels and magnitudes behind the documented correlations.

To the best of my knowledge this is one of only a handful of papers that utilize quasi-experimental variation in school characteristics to study teacher turnover, and it is the first to document the causal effect of student aptitude on their teacher's labor supply decisions. Most of the papers utilizing exogenous variation in this context focus on accountability rules in the US (Feng et al. 2018; Dizon-Ross 2019; Shirrell 2018), UK (Sims 2016) or Norway (Gjefsen and Gunnes 2016), and even among these studies the results are inconclusive. Feng et al. (2018) and Sims (2016) suggest that schools labeled as low-performing experience higher teacher mobility but Dizon-Ross (2019) finds an opposite result. Likewise, Gjefsen and Gunnes (2016) find that introduction of accountability in Oslo increased turnover but Shirrell (2018) does not find any average effects of subgroup-specific accountability in North Carolina, and if anything, the latter paper suggests that black teachers in schools that were held accountable for black students subgroup left their schools at lower rates than black teachers in schools where this subgroup was not held accountable.

Even less is known about the effects of student characteristics on teacher turnover. The closest in spirit to this paper is Jackson (2009), who uses variation in school racial compositions due to North Carolina's busing policy. He finds that schools facing exogenous increases in black enrollment experienced adverse labor supply decisions of their teachers. It is not clear, however, if the variation due to racial desegregation generalizes beyond the U.S. labor market for teachers.

My paper extends and complements this line of work by utilizing exogenous variation in student ability rather than race, and provides estimates from a different country and institutional context. It further delineates the estimated teacher mobility effects depending on their destination (e.g., to private schools or out of schooling sector), investigates heterogeneity along multiple dimensions, and explores the role of auxiliary student characteristics that are correlated with ability.

I explore a major reshuffling of students induced by an admission reform introduced in the municipality of Stockholm, Sweden, in the fall of 2000. Prior to the reform, students applied for a program, and while they could have stated their preferred school, those living closest to a particular school had priority. Thus, although the program choice included an element of school choice, it essentially limited the possibilities of students living in less affluent neighborhoods, as these students had little chance at admission into permanently oversubscribed programs in prestigious schools.[4] The 2000 reform abolished all residence-based admission criteria and introduced a system based solely on performance in lower secondary school. The reform was intended to undo the effects

[4]Although, Stockholm has a very well developed public transportation system, its housing market is highly regulated. It is much easier to buy or rent a flat in a less affluent neighborhood and commute within the city than it is to get housing in a prestigious location and reduce transportation costs and time. This feature becomes even more salient if the school admission system is, for the most part, residence based. Söderström and Uusitalo (2010) document that the reform studied here increased the share of students attending school in an area other than where they live from 45 to 63 percent while the average commuting distance increased from 4.1 to 5.2 kilometers. Efficient public transportation system and compact size of the city likely facilitated such resorting of students once the zoning criteria have been lifted.

of residential segregation and to give all students the option of attending the most elite schools, irrespective of where they lived.

I find that a 10–percentile-point increase in incoming students' credentials decreases teacher separation rates by up to 10 percentage points (pp) or about 30 percent. This effect is driven primarily by teachers switching schools rather than leaving the profession, and it is concentrated at the bottom two-thirds of the student ability distribution. The estimated effect is statistically significant, economically meaningful and somewhat heterogeneous across different groups of teachers. Furthermore, teachers seem to react mostly to the direct measures of student aptitude. Once student credentials are taken into account, other characteristics like immigration background become unrelated to teacher mobility.

2 Background

2.1 Educational institutions in Sweden

The Swedish schooling system starts with voluntary preschool and continues with nine years of compulsory education. Lower secondary school covers grades 7-9, and school grades received in 9th grade - a combination of standardized examination and teacher assessment - determine a student's chances to advance to upper secondary (high) school. By law, Swedish municipalities are obliged to provide upper secondary schooling to all students who successfully complete compulsory education, and in the years considered in this paper about 98% of students observed in the compulsory grades were also observed in the upper secondary grades. This high continuation rate is consistent with national statistics (OECD 2019). Upper secondary school consists of different programs, lasts three years and provides eligibility for post-secondary education. Programs are subject oriented tracks that determine the amount and type of coursework (i.e., curriculum) that students need to master prior to graduation (e.g., Mathematics, English, or Marine Technology). Students applying to high school need to declare their program of interest and it is generally not possible to change this decision during the high school years. Therefore, in the Swedish setting, there is a limited scope for attempts to game the system by applying to less desirable program but in a better school. The reform in question did not alter the requirement to chose program of study but rather it allowed for freedom in student sorting by both schools and programs. Information about school quality - during the period investigated in this analysis - was not easily accessible, and was primarily based on word-of-mouth and historical reputation. Specifically, no national ranking tables, either formally assembled by the ministry or informally assembled by the newspapers, were published and parents relied on the historical value of attending one of the legacy Stockholm schools. As noted in the introduction, however, these schools were permanently oversubscribed with local students which, paired with highly regulated and inelastic real estate market, prevented access of high-performing children from poorer neighborhoods.

The teaching profession in Sweden is regulated and different qualifications are required depending on school type and subjects taught. Teaching at the secondary school level requires completion of special coursework beyond that necessary for a compulsory school teacher. Individuals from other professions (e.g., mathematicians or biologists) who want to become teachers need to supplement their professional degrees with a minimum of 1.5 years of preparation in pedagogy, didactics and teaching practice. Uncertified teachers, however, can also be hired on fixed-term contracts.

Municipalities are the formal employers of public school teachers in Sweden, the primary sample of interest in this paper (Fredriksson and Öckert 2008). They receive grants from the central government but are allowed to freely allocate the money across schools and items within the schooling budget. In practice, however, the decisions regarding teacher recruitment, selection and employment are made at the school level by a principal who has virtually unrestricted control over these decisions conditional on certain minimum standards set by Swedish law and central bargaining processes (Böhlmark et al. 2016). Teachers can be employed on either a permanent (akin to tenure in academic world) or fixed-term contracts, and the latter type is not necessarily limited to replacement teachers hired for few months only. The freedom in hiring decisions granted to the principal and expressed in the legislation is also confirmed in principal surveys (Skolverket 2009). Due to this institutional setting, unlike in other countries, there is no central allocation mechanism of teachers to schools, and job matching occurs through demand-supply equilibrium at school-individual teacher level. Teachers search for open positions, apply for the job, and undergo an interview process (generally with a principal) after which offers are made on competitive basis subject to minimum legislative standards. They are also free to terminate their contract with a school subject to its conditions, and if these are not specified then labor law or union regulations are applied. In that, depending on the length of employment, between one and three months notice is required, however, both parties can agree to different termination arrangements.

When a new teacher is hired they negotiate the salary with a principal, given the collective bargaining outcome set at the national level. Teacher pay is not legally tied to any objectively measured indicators of performance at either school or individual level such as those based on value added models. Nonetheless, teachers with better education and more experience generally earn more than those without university degrees or those who just started their careers. This system has been introduced in 1996 and was a departure from “wage scales” determined in central negotiations utilized in earlier years. Even though, theoretically, it allows for larger variation in pay Fredriksson and Öckert (2007) document that, at least in the early years overlapping with a period considered in this paper, the distribution of teacher wages remained compressed and there were no additional returns to teacher education. On the other hand, Hensvik (2012) shows that competition from private voucher schools modestly increased wages of public school teachers, and in particular those that were just entering the profession. Thus, it is conceivable to think that principals may adjust the salaries of their employees when faced with changes in ability of the students, and in particular those of newly hired teachers.

The primary empirical analysis in this paper focuses on public school teachers, however, private schooling has been growing in Sweden starting from the mid 1990s. The voucher reform (Böhlmark and Lindahl 2015) introduced both non-profit and for-profit private schools, and the municipality is obliged to pay these schools for each student they can attract with an amount corresponding roughly to average per student cost in public schools. Private schools have even more freedom in their management practices than public schools. Over the period considered in this paper, however, the private sector involvement was still limited, and these schools educated a relatively small share of students and employed only a fraction of teachers working in public schools.

2.2 The admission reform

In the fall of 1999, the municipality of Stockholm passed a regulation that changed high school admission rules. Before the 2000-01 school year, students applied only for a program, and while they could have stated their preferred school, those living closest to a school had priority. Thus, school attendance was primarily determined by child's place of residence and grades from lower secondary school mattered only in undersubscribed upper secondary schools. For example, if a school located in downtown Stockholm excelled in a science program and there were enough students living nearby who subscribed to this program, then students with better grades residing in e.g. Tensta (a relatively poor and disadvantaged district in Stockholm) would be unable to gain admission to that program and school. This restriction was particularly binding for the two most popular and broadest programs: social sciences (samhällskunskap) and natural sciences (naturvetenskap). Thus, in practice, students from low-income, disadvantaged districts had virtually no chance of attending the most popular, prestigious and thus oversubscribed inner-city schools; even if they had competitive grades.

The student cohort applying to high school in May of 2000 for the 2000-01 school year faced different admission criteria. Under the new system, students apply for a specific school and program, and applicants are ranked by schools and programs. If a student's first choice is not accepted, the second choice is considered, and so on. Thus, in line with the new regulation, all residence-based school allocation within the municipality of Stockholm was abolished and replaced by a system based exclusively on GPA from lower-secondary school (9th grade), which is the treatment variable of interest in this paper. I construct this variable in two ways. First, as an average GPA of students in each upper-secondary school in each year e.g., for school year 1999/2000 I take an average of 9th grade GPA (percentiled in graduating lower-secondary school cohort) across all students attending particular upper-secondary school in grades one to three that finished lower-secondary school up to school year 1999/2000. Note that this includes more than three cohorts of lower-secondary graduates to allow for grade repetition and returns to schooling. Second, I consider only the freshman students in each cohort e.g., freshman first-graders admitted in 1999-00 school year (pre-reform) vs. those admitted in 2000-01 (post-reform). For the most part, since I focus on the single year pre- and post-reform I use the latter measure. However, the main results are robust to using ability averaged across all students in a given high school. As noted above, student ability is based on percentiled ranking of all students in the cohort graduating from lower secondary school in the school year directly preceding their expected enrollment in upper secondary education. Further details on the sample construction are provided in the Online Appendix.

Most municipalities surrounding Stockholm do not offer all of the programs, and a student has the right to attend their chosen program in another municipality, financed by the municipality in which they reside. Cross-municipality commuting is relatively common in Sweden, and if increased school choice incentivized more students from outside of Stockholm to apply to schools in Stockholm, crowd out of students residing in Stockholm could occur. Furthermore, Stockholm schools could decide to change the number of admitted students in response to a higher demand, which would in turn lead to either higher student-teacher ratios or the need for additional hires. I address the latter issue in Section 4, and my calculations show that the fraction of students living outside of Stockholm but attending Stockholm schools is stable at around 20% over the analyzed period.

2.3 Data and descriptive statistics

This paper utilizes multiple Swedish population-wide registries. The main data source is the teacher registry that covers all teaches employed in Swedish schools during the school years from 1991-92 through 2004-05. It contains information on teachers' education, specialization, experience, certification, place of work, type of contract (permanent vs. fixed-term) and workload, and this information is gathered in October of each year. I have matched these data to background information on age, gender, immigration histories, education and employment. The pupil registries for lower and upper secondary schools are used to obtain information on students in a given upper secondary school along with their credentials from a lower secondary school. One limitation of these data is that I cannot link individual teachers to their students or grades taught, and therefore the variation needs to be analyzed at school-by-year level rather than considering changes in exposure at a finer granularity e.g., particular program, subject or subject-grade. Measures of family background were also obtained by matching students to their parents while administrative records on earnings provide information on teachers' monetary compensations. The details of the sample construction are discussed in the Online Appendix.

The reform was only implemented in the municipality of Stockholm, and thus I first compare basic descriptive statistics for Stockholm and non-Stockholm schools for the last pre-reform (1999-00) and first post-reform (2000-01) school years (Table A1). Stockholm is more affluent in many dimensions than the rest of Sweden, and schools in Stockholm admit higher-achieving students who come from richer and better educated families. At the same time, due to Stockholm's major concentration of immigrants, these schools admit more minority students. Stockholm schools also employ more teachers with university diplomas, but these teachers have less experience on average.

In the pooled sample of all secondary schools in Stockholm prior to the 1999-00 school year, there are 8 private and 21 public schools. As noted before, these private schools are smaller (mean number of students across all grades and cohorts of 591 as compared to 823 in Stockholm public schools used in the empirical sample) and employ fewer teachers (mean number of teachers of 36 as compared to 71 in Stockholm public schools used in the empirical sample). Given the timing of the reform and estimation strategy, however, I focus on secondary schools that have been in operation in Stockholm for all school years from 1994-95 to 2004-05. This restriction corresponds to the full range of years needed to perform analyses in the paper, including placebo regressions, and yields 18 schools (17 public and 1 private). It also addresses potential composition effects related to school openings and closures. In order to keep the sample more homogeneous in terms of administrative oversight, however, I exclude the one private school from the main empirical analysis. As documented in Table A2 the results are robust to including private schools as well as to various other sample definitions.

Table A3 further provides descriptive statistics for schools included in the primary empirical analysis compared with other schools in Stockholm municipality excluded from the preferred estimation sample. Included public schools have, on average, modestly higher SES students than excluded public schools but lower SES students than private schools in operation in Stockholm. Likewise, their teachers are comparable to those in other public schools but are better educated and more experienced than those in the private schools. Finally, the year-to-year turnover appears higher in the private sector but these differences are not statistically significant.

Another relevant question is whether, in Stockholm, there is variation in the characteristics of schools educating children with different achievement levels. Table A4 provides evidence on this documenting the characteristics of schools by terciles of pre-reform student achievement levels based on school year 1999-00. It illustrates two primary endogeneity concerns mentioned in the introduction. First, there is substantial sorting of teachers to schools with higher vs. lower ability students – the former set is taught by more experienced and better educated teachers. Interestingly, however, teachers in schools with high-achieving students are more likely to be employed on fixed-term contracts and earn less on average. Second, student GPA is highly correlated with other characteristics. For example, schools in the top tercile of student ability, have fewer children of immigrant origin but more of those who come from richer and better educated families.

Since implementation of the reform started during the 2000-01 school year, I present descriptive evidence for 1999-00 as the last pre-reform school year and 2000-01 as the first post-reform school year. Table 1 presents descriptive statistics based on the 1994-95 to 2004-05 panel of Stockholm schools for the immediate pre- and post-reform periods, separated by changes in student composition. In particular, for each school I calculate the difference between mean-incoming-student GPA (only first-grade students who applied to school in the same year) in the first post-reform year, 2000-01, and the last pre-reform year, 1999-00. Then, I order these seventeen differences from the schools most negatively affected to those most positively affected and divide the ranking into terciles. I call these schools downward-, middle- and upward-shocked schools, and at the bottom of the table I report the number of schools and teachers in each group.[5]

The reform reshuffled incoming first-grade pupils between schools in Stockholm. In particular, incoming student GPA in the upward-shocked schools increased from 59.2 to 62.8 percentile points, while in the downward-shocked schools, student GPA decreased from 49.0 to 45.5 percentile points.

[5] A concern here is that defining the grouping variable based on only one pre- and one post-reform year could be sensitive to choosing an atypical cohort. To explore this potential issue I have also constructed terciles based on all available years in the pre- and in the post-period separately. This variable is correlated with the one used in Table 1 at 0.89 and 0.95 for continuous measure and terciles, respectively.

This widened the gap between these two groups of schools from 10.2 to 17.3 percentile points, equivalent to over one-third of a standard deviation change in student achievement in the primary empirical sample or over 50 percent of a standard deviation change when considering all students in Sweden.

The other student characteristics correlated with student ability, such as parental income or share of minorities, also changed. For example, the gap between schools with the most and the least improvement in mean parental income increased by one-third, while the minority students' gap increased by about 20 percent. The reform also affected the composition of the teacher stock. On average, there were more teachers with university diplomas in the upward-shocked schools and more teachers on fixed-term contracts in downward-shocked schools in the post-reform period than in the pre-reform period. The gap in teacher compensation did not change substantially, and interestingly, teachers in schools with higher-achieving students earned less than those in schools with lower-performing students, suggesting descriptively the presence of compensating wage differentials in a system with a relatively flexible teacher pay scheme.

2.4 Identifying variation

Due to the reform, from one year to the next, the same set of teachers experienced changes in incoming students' ability. In particular, some teachers ended up with lower-achieving pupils, while other teachers ended up with higher-achieving pupils than in the pre-reform period. The aim of this paper is to study how teacher labor supply decisions changed in response to this unexpected change in student credentials.

Figure 1 shows the differences in average student credentials - computed over all grades in a given upper secondary school - for every school year (1996-97 to 2004-05) and for upward- and downward-shocked schools (defined as in Table 1) relative to average student credentials in the same schools in the 1995-96 school year. The differences are plotted as points, while the whiskers for each year show 95-percent confidence intervals from linear regressions, with the difference in the average student credentials compared to 1995 as the dependent variable and year dummies (one for each year between 1996 and 2004) as independent variables. Figure 1 accurately illustrates the rollout nature of the reform. Prior to the reshuffling, there are no statistically significant differences in average student ability in upward- and downward-shocked schools. Mean changes in credentials between 1995-96 and 1999-00 in these two sets of schools are 2.4 and -1.9 percentile points, respectively. Yet, post-reform, there is gradual divergence for these two groups over the first three years (the length of upper secondary education), but then subsequent years are much more comparable to school year 2002-03 (the first year where all grades were admitted under the new rules). Mean changes in credentials between 1995-96 and 2004-05 for upward- and downward-shocked schools are 20.4 and -7.2 percentile points, respectively (changes for 1999-00 to 2004-05 comparisons are 18.0 and -5.3 percentile points, respectively). These differences are more pronounced in upward- relative to downward-shocked schools which could partially be due to the fact that not all schools in Stockholm are included in the empirical sample used in Figure 1 and partially due to the fact that this reform has been found to increase average grades and the share of top-performing students in lower-secondary schools (Vlachos 2010). I address both these issues in subsequent sections of the paper, but in relation to Figure 1 it is further worth noting that school-level standard deviations in average student grades are only slightly larger in downward- as compared to upward-shocked schools. Consistent with sorting by student ability only, however, these standard deviations decline in the post-period and schools become more homogeneous. For the most part, I do not explore the changes in average characteristics in this paper, but rather, I focus on the changes in freshman cohort incoming students' credentials since this is the margin for which the shock induced by the reform was the most pronounced. Naturally, the two measures are highly correlated, and Figure A1 confirms that the largest shock in incoming student ability occurred between 1999-00 and 2000-01 school years, while the subsequently admitted cohorts mimicked the ability of the first grades from the 2000-01 school year. Furthermore, the results are robust to defining treatment based on mean characteristics of students across all cohorts, rather than freshman only, as documented in Table A5. The point estimates change in size since the underlining treatment variable has modestly different distribution, and if anything the preferred estimates using only freshman students provide more conservative effect sizes.

3 Empirical specification

The reform exogenously altered student composition starting in school year 2000-01, and thus it caused teachers to face a different set of students from one year to the next. If teachers value working with high achieving students, holding other inputs into their utility function constant, then they should be less likely to leave schools experiencing an inflow of students with relatively better credentials. Conversely, we should observe higher separation rates in schools with an inflow of low achieving students. Econometrically, the variation induced by the reform can be framed as a difference-in-differences estimator. The first difference compares schools before and after school year 2000-01, while the second difference compares schools that experienced differential changes in student ability (Figure 1). For teacher turnover to result from changes in student aptitude only, it cannot be that students selected school based on the underlying trends in teacher turnover or that teachers anticipated the inflow of students with particular characteristics. Additionally, there should be no other external factors influencing teacher mobility and student sorting simultaneously occurring with the reform. I discuss these assumptions below.

Moving to the outcomes, the nature of the outcome variable of interest - job mobility - requires two years of data to construct a single observation. That is, I use a teacher's employment status in periods t and $t + 1$ to generate a mobility indicator. Since high school education in Sweden consists of three grades, it took three years for the reform to reach full implementation (as evidenced in Figure 1). During the 2000-01 year, only a third of the student stock had been admitted under the new rules; during 2001-02 school year approximately two-thirds of students were admitted under new rules; and the reform come into full effect in the 2002-03 school year. Finally, since the application process started in May 2000, it is possible that teachers reacted to information about the ability of the incoming students and left their jobs between May 2000 and October 2000, when the employment status is measured in teacher registry. Such an anticipation effect would violate parallel trends assumption. However, lagging the dependent variable by one period mechanically purges the possibility of a reaction to student ability in advance of the policy implementation. Since the reform was not yet announced early in the 1999-00 school year, teachers could not have possibly anticipated the changes in student composition. Such a lagged specification is preferred, as it provides the most conservative econometric approach; on the downside, it requires following teachers for four years after the reform. In principle, teachers could also leave within a school year, but such situations are rare, and this type of mobility would be captured by comparing two adjacent years unless they transitioned back and forth between schools within one year. If a teacher moves from and to the same school within a single year then I cannot observe such mobility.

Taken together, these data and institutional features require studying changes in teacher mobility up to four (k) years after the reform, and the estimating equation is:

$$Y_{ist}^{k} = \beta Q_{st} + \gamma X_{is} + \delta_s + \zeta Post_t + \varepsilon_{ist} \tag{1}$$

where Y_{ist}^{k} is turnover of teacher i employed in school s, year t and for mobility lag k (k varies from one to four because of three years of high school and one anticipation effect year as explained above); Q_{st} is incoming students (freshman cohort) average ability in school s and year t; δ_s are school fixed effects; $Post_t$ takes value one for school year 2000-01; X_{is} is a vector of teacher's individual level controls including indicators for gender, marital status, immigrant origin, specialization (science, vocational or special education), university diploma as well as experience in years; while ε_{ist} is heteroskedasticity robust error term.[6]

The coefficient of interest is β, and it identifies the effect of student ability on teacher mobility under two assumptions. First, changes in student composition cannot be correlated with changes in teacher mobility in the reform's absence. Second, there cannot be any external shocks to the demand or supply of teachers that simultaneously occur with implementation of the reform. A testable implication of the former identifying assumption is that post-reform changes in student ability are not correlated with pre-reform changes in teacher mobility. This is equivalent to the

[6] The common approach in the literature is to assume independence at the level of aggregation where the variation in treatment is present (Bertrand et al. 2004). However, clustered standard errors only have asymptotic properties, and in the regressions with 17 schools, these large sample properties cannot be invoked (Angrist and Pischke 2009). For this reason, I have chosen heteroskedasticity robust standard errors for the results reported in the paper, thus imposing the assumption that teachers are independent within schools. In Table A7 I also report alternative errors computed based on the preferred specification from column 4 of Table 4: (i) robust standard errors as a reference; (ii) standard errors clustered at the school level; (iii) standard errors clustered at the school×year level, thus allowing interdependence between teachers in a school in a specific year but not across years; (iv) standard errors clustered at the teacher level to account for the fact that some teachers are observed in the dataset multiple times (v) standard errors from regressing the first-differences on the treatment variable using aggregated residualized data. Irrespective of the specification the preferred estimate is always statistically significant at least at 5% level. Furthermore, I also checked the difference between heteroskedasticity robust and school-level cluster robust standard errors for the largest possible sample of 29 schools based on column (7) from Table A2. In this setting bias due to small number of clusters should be reduced, at least as compared to the sample with 17 clusters. In the former case the error is 0.206 while in the latter it is 0.400, but the estimate remains statistically significant at 5% level.

common trends assumption, and the placebo treatment period must not overlap with the true treatment period. I present the results of these regressions in Table 2, where I lag mobility in the pre-period and regress it on the change in treatment i.e., difference in student GPA in pre-period and the first year of post-period. Note that one-year mobility estimate in this table is equivalent to estimating the anticipation effect described above. The results confirm that changes in student ability induced by the admission reform do not significantly predict pre-reform changes in teacher mobility, thus supporting the first identifying assumption. Nonetheless, to further rule out any possibility of bias due to anticipation effects, I elect for lagging the dependent variable by one year as denoted in Equation 1. Main results based on the less conservative approach are presented in Online Appendix Table A6, and yield qualitatively identical conclusions. With respect to the second identifying assumption, I am not aware of any shocks to teacher labor market or student sorting in Stockholm that occurred at the same time as the admission reform.

It is important to highlight at this point that even though I use multiple years to construct the outcomes - due to the institutional setting and the nature of outcome variable, each regression uses only one pre- and only one post-reform period. Furthermore, the treatment change is always assigned to the first year post-reform since multiple confounding factors, such as family mobility in response to changing rules, could occur in subsequent years (Calsamiglia and Güell 2018). Figure A1 documents that the largest shock at school level indeed occurred in the first year of treatment and subsequent cohorts mimicked the ability composition of students first admitted in school year 2000-01. For further clarity, in Table A8, I provide the exact details on what years are used in each comparison estimated in the main results. The preferred specification is one from the last row in this table, and, as explained in Section 2.3, I focus on a balanced panel of schools. Depending on the length of the mobility window considered, I use different samples of teachers in the pre-period but the same sample of teachers in the post-period. Balancing the sample at the teacher level in both pre- and post-periods is undesirable in this setting because it focuses on less mobile teachers and potentially leads to selection on the outcome.

In order to illustrate the logic behind the difference-in-differences strategy used in this paper, Table 3 presents changes in teacher mobility over time for schools that experienced positive or negative changes in student credentials. Mimicking Table 1, I divide schools into three groups based on their changes in incoming student credentials between school years 1996-97 (pre-reform) and 2000-01 (post-reform). I use GPA from 1996-97 rather than last pre-reform year (1999-00) to match the preferred specification defined in the fourth row of Table A8. In the first column, I show data for the one-third of schools with the most positive changes in incoming student GPA (one-third upward), while in the second column I show data for the one-third of schools with the least positive changes in incoming student GPA (one-third downward).

On average, among schools in the empirical sample, student credentials increased by more than 17 percentile points in the former set of schools and decreased slightly by almost 4 percentile points in the latter set of schools.[7] Concurrently, teacher mobility decreased by 16 percentage

[7]This does not necessarily indicate that the average student quality in Stockholm increased due to the reform as the comparison excludes the middle tercile schools and not all Stockholm schools are included in the empirical

points in upward-shocked schools, and there was small and statistically insignificant increase in mobility in downward-shocked schools. By calculating the ratio of the two changes (-21 percentage points divided by 21 percentile points), I obtain the Wald estimate of 4-year teacher mobility on incoming student credentials. The estimate implies that increasing incoming student credentials by 10 percentile points reduces teacher mobility by 9.9 percentage points. In the reminder of the paper I investigate whether these results hold up in a more formal regression analysis where the dummy variable for school shock is replaced with a continuous measure of incoming student ability.

4 Results

Table 4 presents main results of the paper. The first row compares one-year mobility in 1998-99 to one-year mobility in 1999-00, thus replicating column (1) from Table 2 in support of no anticipation effects. Subsequent rows of this table estimate the effects of student ability on teacher mobility from one to three years after the first cohort entered schools under the new rules. In columns (1) and (2) I present correlations between GPA and mobility using all Swedish high school teachers and only those working in Stockholm municipality, respectively; in column (3) I present difference-in-differences estimates without controlling for any observable teacher characteristics; and in column (4) I condition on a set of pre-determined teacher controls.

The preferred point estimate - row (4) of column (4) of Table 4 - suggests that, when the reform was fully implemented, a 10-percentile-point increase in student ability reduced the probability of a teacher leaving their school by 10 percentage points.[8] Given that the mean of dependent variable in this regression is 33.4 percent it would suggest reduction in mobility by about 30 percent. Alternatively, an increase of one standard deviation (18.6 percentiles) in incoming student credentials decreased the probability of a separation within four years by 18.6 percentage points, or approximately halved it. Estimates for shorter mobility windows, corresponding to only partial reshuffling, yield smaller coefficients and effect sizes which is expected given that post-reform incoming cohorts mimic their composition i.e., each new cohort brings similar set of students of either higher or lower ability than those attending each school prior to the reshuffling. In order to visualize how the effects

sample used in Table 3. For example, Table A3 documents that schools in the empirical sample attract, on average, higher ability students. Nonetheless, comparison between columns (1) and (2) of Table 3 makes it clear that there is variation in changes in incoming student's GPA in the post-period compared to the pre-period. In Table A2 I further document that the main results are not sensitive to including broader set of schools. Furthermore, Vlachos (2010) shows that the reform did indeed increase average grades and the share of top-performing pupils in lower-secondary schools in Stockholm, and thus part of the treatment effect could operate through pupil effort and teacher's exposure to more ambitious in addition to higher performing students. This increase in average credentials will not bias the estimates so long as selection of students into schools is not correlated with pre-trends in teacher mobility in these schools. Evidence in Table 2 supports this notion.

[8] I have also estimated models that additionally include teacher fixed effects. The estimate for fully implemented reform is -0.419 (standard error of 0.255) and is statistically significant at 10% level. Thus, 10 percentile-point increase in student ability reduced the probability of teacher leaving their school by 4.2 percentage points, an effect that is about 40 percent the size of the preferred estimate in Table 4. Teacher fixed effects specification can be viewed as very restrictive in this particular application because it requires teachers to be present in both pre- and post-reform period (i.e., potentially focusing on less mobile teachers) as well as to move in either pre- or post-period (i.e., teachers present in both periods that are moving in both periods or neither do not contribute to the identifying variation).

evolve as the reform progresses Figure A2 plots point estimates and 95% confidence intervals for specifications from column (4) of Table 4. The relationship is clearly downward sloping, starting close to zero, as there are virtually no anticipation effects. The F-test rejects the hypothesis that all four estimates are identical (p=0.033). Due to the size of standard errors, however, I cannot reject equality between the anticipation period and the first year of reform implementation (p-value of 0.316). At the same time, estimates for subsequent two periods when 2 out of 3 or all cohorts are admitted under the new rules are statistically distinguishable from the anticipation estimate with p-values of 0.050 and 0.006, respectively. This provides further confidence in the validity of the empirical design.[9] Finally, it is worth highlighting that the difference-in-differences estimate is larger than the equivalent OLS correlations using either all Swedish high school-teachers (where sample composition is a potential additional confounder) or only those working in the municipality of Stockholm where the reform was implemented. This suggests that cross-sectional associations relating teacher turnover to student ability may be severely biased.

The effect size of the preferred estimate may appear large, but as illustrated in Figure A3, there are only two schools in the empirical sample in which students improved by more than a standard deviation (mean improvements of 18.3 and 20.0 percentile points while the in-sample standard deviation is 18.1 percentile points). It is plausible that these two schools were the most oversubscribed and popular pre-reform, however, without access to applications' data I am unable to empirically verify this hypothesis. Four schools improved by more than half but less than one standard deviation, while the remaining eleven school in the empirical sample either improved or deteriorated by up to half of a standard deviation. Although Figure A3 illustrates a clear downward sloping relationship, one may be worried that this fit is driven by a particular school or set of schools - especially given the relatively small number of schools in the sample. Therefore, Figure 2 provides a robustness check testing the role of potential influential observations. Each coefficient in panels A and B comes from the preferred regression - row (4) of column (4) in Table 4 - where I exclude either one school at a time (panel A) or two schools pairwise at a time (panel B). No single school or pair of schools are driving the estimated relationship, which alleviates the worry about outliers or the relationship being due to a particular sample selection. The coefficients range from -1.3 to -0.8 in panel A and from -1.7 to -0.5 in panel B, and all are statistically significant at least at 5% level.

As explained in Section 2.3, given the timing of the reform as well as the nature of the outcome of interest, I made certain sample selection choices that resulted in the use of only a subset of Stockholm schools in the primary analysis. Table A3 documented that these schools are somewhat different

[9] In order to increase the statistical power at the cost of potentially including less stable schools, I also repeated the main analysis using the largest possible sample based on column (7) of Table A2. The four estimates corresponding to anticipation effect (mean mobility of 16.9 percent), one grade reshuffled (mean mobility of 22.8 percent), two grades reshuffled (mean mobility of 27.0 percent), and all grades reshuffled (mean mobility of 30.7 percent) are -0.072 (SE of 0.206), -0.508 (SE of 0.184), -0.683 (SE of 0.194) and -0.883 (SE of 0.206), respectively. These estimates are very similar to those presented as the main results in Table 4, and in fact the anticipation effect in this larger sample is smaller. Furthermore, the F-test rejects the hypothesis that all four estimates are identical (p=0.037), while the corresponding p-values for differences between the anticipation effect and the following three treatment effects are 0.114, 0.031, and 0.005, respectively. The conclusions thus remain unchanged.

from the excluded schools. Therefore, to better understand the generalizability and stability of my findings, in Table A2, I further explore sensitivity of the main results to alternative sample selection choices. Column (1) replicates the main coefficient of interest, columns (2) to (4) present analyses for variety of school panel assumptions, while columns (5) to (7) present results using repeated cross-section of schools complementing the main sample with either only public or only private or both types of school. This table makes it clear that initial focus on more stable public schools does not drive the results, and all coefficients are comparable irrespective of the exact set of schools used.

The evidence thus far suggests that higher student credentials reduce the probability that teachers leave their employment. To the best of my knowledge, this is the first paper that causally estimates the effect of changes in student ability - as measured by academic credentials - on teacher labor supply decisions. One alternative explanation for the findings, however, is that higher competition for slots in the best schools in Stockholm increased pupil effort in lower secondary schools, and if this effort effect persists through upper secondary school then it may be that teachers actually respond to having more ambitious rather than higher ability students (Vlachos 2010; Haraldsvik 2012). I cannot directly observe student effort in my data but there is no doubt that variation in student ability is the forcing variable that allocates students to schools, and it should be positively correlated with both cognitive and non-cognitive skills of students. This ability measure (GPA) is, however, also correlated with other observable variables such as the fraction of minority students or parental wealth. Therefore, I investigate whether the estimated effects are due to direct measures of student aptitude or to variables correlated with student GPA. I am unable to study independent variation in both student ability and other student characteristics, and thus, this analysis should be viewed as more of a descriptive exercise. To the extent that these additional characteristics explain less variation in teacher mobility than pupil GPA, it provides an indirect evidence that similar relationship may hold for achievement and effort.

The results presented in Table 5 focus on the preferred specification based on column (4) and row (4) of Table 4. The first row of Table 5 presents estimates in which the treatment is defined as a fraction of first-generation immigrants (a correlation of 0.43 with GPA), the second row presents estimates for mean parental income (a correlation of 0.82 with GPA), the third row presents estimates for mean parental education (a correlation of 0.92 with GPA), and the fourth row presents estimates for mean combined cognitive and non-cognitive assessment of fathers based on their military records (a correlation of 0.82 with GPA).[10] Column (1) presents the effects of the "alternative" characteristics, while column (2) adds student GPA in a horse race between direct and indirect measures of student ability. As noted above, I cannot distinguish between causal effect of student ability and independent causal effect of e.g. parental education as the exogenous variation exists only for the former measure.

[10]These data are available only for some fathers, and the coverage at school level increases from 24% to 51% over the time period used in this analysis. On average, I have information on fathers of 33% of the pupils. This limitation is driven by the fact that the registries are not available for individuals tested before 1970 and for immigrants. Nonetheless, I calculate the mean for all fathers with assessment information available in a given school and year, and in these regressions I also control for the share of students in a given school and year for whom I do not observe paternal cognitive and non-cognitive assessments.

The unconditional estimates in row (1) confirm that the fraction of minorities at a school correlates positively with the probability of job separation (Hanushek et al. 2004; Falch and Strøm 2005; Barbieri et al. 2011; Karbownik 2014). In row (2), the coefficient on mean yearly income in 100 000 SEK is -0.070 with a standard error of 0.035, which is small given that the mean yearly parental income in the studied group of schools is 379 817 SEK. Similarly, rows (3) and (4) indicate significant and robust negative associations of increased parental education as well as paternal aptitude and teacher job separation, which is consistent with the intergenerational transmission of education and cognitive skills (Björklund et al. 2006; Black et al. 2009).

In column (2), where the regression is augmented with student GPA, estimates for the fraction of minorities and paternal cognitive and non-cognitive military assessments become insignificant and decrease in size, and the coefficient on mean parental income actually turns positive. On the other hand, both parental education and student GPA are negative. Overall, the estimates in column (2) suggest that teachers primarily value student aptitude, but that some of the response to changes in student credentials may be driven by changes in students' socioeconomic background.[11]

Results in Table 4 indicate that teachers who face higher ability students are less likely to leave their schools suggesting their preference for working with high achieving pupils. An alternative mechanism behind such findings, however, might be a school-level equilibrium effect, e.g., in response to the reform, schools that experience an inflow of high-achieving students could simultaneously grow to accommodate an increase in demand, which would mechanically lead to reductions in teacher turnover. Thus, in Table A9 I test whether the reform affected the number of students enrolled, number of teachers in a school and student-teacher ratio. Contrary to the mobility analysis, these regressions are based on a static model in which the outcome is determined at a given point in time. Furthermore, since school composition in the pre-period was determined in September 1999, and the reform was not voted into power until later in 1999, there is no need to account for an anticipation effect in this setting. The results in Table A9 show that neither the number of students, the number of teachers nor the student-teacher ratio responded to the reshuffling. All the estimates are positive, in line with reduced mobility, but are small in magnitude and never statistically significant. Thus, it seems unlikely that post-reform changes in teacher mobility were a mechanical consequence of changes in school size or school resources.[12]

[11] Since direct (student GPA) and indirect (share of immigrants, parental income and education, paternal military test scores) measures are highly correlated, it is also plausible that models in column (2) pick up non-linear measures of student ability. When I add the square of student GPA to the estimates in column (2) it is indeed positive and statistically significant in all estimations. At the same time, the linear term in student GPA remains highly statistically significant and negative in all cases. Finally, the statistically significant negative coefficient on parental education becomes small and insignificant suggesting that indeed it was picking up non-linearity in student ability. The coefficient on parental income remains positive and statistically significant with a similar point estimate.

[12] The funding of schools in Sweden is tied to the number of enrolled students, and the reform could have forced some students to change schools as a response to changes in peer composition. I address this issue by estimating a model in which I define the outcome as the probability that I do not observe currently enrolled student in the same school in the next school year, and I construct the mean of this probability at the school level. For each exposure length I find small but statistically significant effects on student mobility pointing towards lower likelihood of leaving school if there is an inflow of higher achieving peers. Since these estimates do not appear to be quantitatively meaningful (all effect sizes are less than one percent) and I do not find any effects on the average school size, I conclude that this potential general equilibrium effect is unlikely to play a major role in a teacher's decision making process.

The results thus far focused on teachers' labor supply decisions but schools facing higher teacher retention and no apparent changes in average size of enrollment following the reform could also reduce their hiring. Likewise, if earnings and student ability both enter teacher's utility function we might expect changes in teacher's compensation. It may be implausible, however, to adjust wages of all teachers immediately following the reform, and thus earnings of newly hired teachers could serve as a better proxy for exploring the compensating wage differentials. I present these additional results for the preferred specification - corresponding to full reshuffling of students - with and without accounting for anticipation effects in Table A10.

Consistent with reductions in separations, schools that face an inflow of higher ability students reduce their hiring. This effect at -0.57 percentage points or about 1.1 percent, for fully reshuffled schools, is smaller than reductions in teacher turnover estimated in Table 4. The 95-percent confidence interval, however, covers the preferred estimate for decline in teacher turnover. Furthermore, the estimates on hires could be more sensitive to the inflow of new teachers graduating college given that I restrict the sample to teachers between ages 25 and 58. Since the interpretation of the effects on new hires is ultimately more complicated and these results are weaker quantitatively, I chose to focus on the likelihood that teachers depart their current school when faced with students of different ability – a turnover measure predominantly used in the extant literature.

As hypothesized above, I do not find any changes in average earnings of all teachers but there is a reduction in earnings among new hires in schools that experienced and inflow of higher ability students. These earnings effects are, however, relatively small at about 0.9 percent. Similar to prior results it does not appear to matter much if we account for anticipation effects or not, providing further evidence in favor of quasi-experimental nature of the studied shock.

5 Heterogeneity

The richness and completeness of Swedish registry data allows me to investigate heterogeneity in the effects of student ability, as the consequences of the admissions reform could be very different for different groups of teachers. Therefore, I analyze how the response to changes in student composition differs by teacher's education, experience, gender, specialization, type of employment, their destination, and the baseline school-level student ability.

Table 6 presents a range of heterogeneity findings based on teacher characteristics, and has the following structure: The first column reports the fraction of teachers in each group, the second column reports the mean of 4-year mobility, the third column reports the point estimate and standard error of the effect of student ability on 4-year teacher mobility using the preferred specification, and the fourth column provides implied effect sizes for ease of interpretation. I first consider teacher education and experience, as these are important predictors of student achievement (Boyd et al. 2005; Harris and Sass 2011). More than a quarter of secondary school teachers in Stockholm do not have a formal university degree, and these teachers have substantially higher turnover rates (43% vs. 30%). Because of the differences in baseline probability of changing jobs, the elasticity to changes in student's characteristics could be different in these groups. For example, teachers who are less attached to their employers may be less responsive to the changes because their expected exposure to any particular group of students is shorter. The coefficient is larger for teachers with a university degree, and the difference in effect sizes is even bigger because these teachers have lower baseline mobility. Similar conclusions apply to teacher experience - although the point estimates for the three groups are comparable, the effect sizes range from 2.0% for the least experienced to as high as 4.8% for the most experienced teachers due to differences in average turnover rates between these groups.

I also consider whether the estimated effect varies by teacher gender. Female teachers are somewhat less mobile than male teachers, and the estimate for this group is smaller (effect sizes of 3.4% and 2.6% for males and females, respectively). Science teachers are another group that receives a lot of attention in the media and in research (Edmark and Nordström Skan 2010). Here, science teachers are defined as those teaching: mathematics, physics, chemistry, biology and computer science. From the labor market perspective, these teachers provide important STEM skills (Joensen and Nielsen 2009), but they may also have more favorable outside options in the private sector. Indeed, science teachers have higher mobility rates, but the estimated effect sizes for both groups of teachers are similar. Finally, I present estimates separately for teachers on permanent and those on fixed-term contracts. Nearly 20% of teachers in Stockholm are employed on non-permanent basis, and as expected they have more than twice the rate of job separations. The estimated coefficients indicate, however, that these teachers are somewhat less sensitive to changes in average student ability, which may be related to their shorter-run employment perspective at any given school.

The models used so far pool all teacher job changes into a single dimension - leaving their current school. However, previous research indicates that the correlations with teacher characteristics may differ depending on the destination (Lankford et al. 2002). In Table 7, I investigate whether the effects of changes in student credentials are stronger along some mobility margins than others. In particular, I estimate the effect for mobility within schooling (row (1)) and out of the profession (row (2)). Furthermore, I decompose the former mobility measure into moving from one public high school to another (row (1a)), or from public high school to public compulsory education school (row (1c)), or to private schools (row (1c)). Finally, I also investigate the probability of leaving to high schools with higher baseline student achievement (row (3)). Since policymakers should be particularly interested in whether highly-educated teachers tend to leave the profession in response to such a reform, I also estimate the above specifications separately for all teachers (column (2)) and for those with a university degree (column (5)). Effect sizes are provided in columns (3) and (6) for convenience.

These estimates suggest that teachers react mostly in terms of mobility within teaching profession and in particular from one public high school to another. They also seek schools with higher achieving pupils. The estimates for the probability of leaving the profession are smaller and statistically insignificant, which is consistent with Jackson (2013), who argues that due to occupation-specific human capital teachers will adjust their match quality within the profession rather than outflow to a different occupation. The estimates are also comparable (except for private school moves) for all teachers and for those with completed university education.

The last two heterogeneity analyses investigate the distributional effects of changes in student credentials. Table A11 shows how teachers initially employed in schools from different terciles of the pre-reform student GPA distribution respond to changes in their pupils' composition, while Table A12 documents how teachers react to changes in the fraction of incoming students from different parts of the GPA distribution. For this purpose, for every school and year, I calculate the fraction of students admitted from each tercile of the GPA distribution. Then, I use these three variables in separate regressions as a substitute for the average incoming student credentials. Thus, Table A11 reports heterogeneous responses to the same treatment, while Table A12 documents responses to heterogeneous treatments.

The results in Table A11 indicate that only teachers employed in the bottom two terciles of the distribution react to changes in student credentials, and these effect sizes are comparable. When I split the sample into halves I only find significant estimate for the bottom half. It is -1.679 (SE of 0.356) while the estimate for the top half is -0.348 (SE of 0.312), implying the effect sizes for the preferred mobility measure of 6.6 and 1.5 percent, respectively. Furthermore, teachers in the bottom tercile of the student achievement distribution are the only ones who are significantly less likely to leave the profession in favor of a different occupation, and a point estimate of -0.830 implies an effect size of 7.5%. Table A12 largely confirms the findings discussed above. Teachers who experience an inflow of students coming from the bottom tercile of the achievement distribution are more likely to leave their current employment for a different school or a different occupation. On the other hand, teachers who get a positive shock are less likely to terminate their employment supporting the notion that student ability enters teacher's utility function.

6 Conclusions

A number of educational policies involve placing certain groups of students into more favorable school environments, in the hope that interacting with better performing peers will boost their own academic performance. However, the success of such policies relies on, among other things, how teachers respond to changes in their students ability. This paper provides evidence on the causal effect of student aptitude on teacher job mobility using data from Stockholm high schools and exogenous changes in incoming students' GPA following an admission reform.

The results show that an increase in student aptitude indeed leads to lower teacher mobility, and that at least in a setting studied here, an OLS estimate is severely biased. I show that this effect is robust to different model specifications, and I account for the fact that changes in student aptitude in different schools might be related to pre-existing trends in teacher mobility. The effects vary somewhat by types of teachers and are found mostly for mobility between schools rather than out of the profession. Finally, teachers seem to react mostly to direct measures of student ability rather than to characteristics correlated with student aptitude, namely immigrant status, parental income or education.

References

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

Barbieri, Gianna, Claudio Rossetti, and Paolo Sestito, "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications," *Economics of Education Review*, 2011, *30* (6), 1430–1444.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust differences-in-differences estimates," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Björklund, Anders, Mikael Lindahl, and Erik Plug, "The origins of intergenerational associations: Lessons from Swedish adoption data," *Quarterly Journal of Economics*, 2006, *121* (3), 999–1028.

Black, Sandra, Paul Devereux, and Kjell Salvanes, "Like father, like son? A note on the intergenerational transmission of IQ scores," *Economics Letters*, 2009, *105* (1), 138–140.

Böhlmark, Anders and Mikael Lindahl, "Independent schools and long-run educational outcomes: Evidence from Sweden's large-scale voucher reform," *Economica*, 2015, *82*, 508–551.

_ , Erik Grönqvist, and Jonas Vlachos, "The headmaster ritual: The importance of management for school outcomes," *Scandinavian Journal of Economics*, 2016, *forthcoming.*

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, "Explaining the short careers of high-achieving teachers in schools with low-performing students," *American Economic Review*, 2005, *95* (2), 166–171.

Calsamiglia, Caterina and Maia Güell, "Priorities in school choice: The case of the Boston mechanisms in Barcelona," *Journal of Public Economics*, 2018, *163* (20-36).

Cullen, Julie, Brian Jacob, and Steven Levitt, "The effects of school choice on participants: Evidence from randomized lotteries," *Econometrica*, 2006, *74* (5), 1191–1230.

Dizon-Ross, Rebecca, "How do school accountability reforms affect teachers? Evidence from New York City," *Journal of Human Resources*, 2019, *forthcoming.*

Edmark, Karin and Oskar Nordström Skan, "Science of success: The causal effect of high school math and science on labor market outcomes," 2010.

Epple, Dennis, Richard Romano, and Miguel Urquiola, "School vouchers: A survey of the economics literature," *Journal of Economic Literature*, 2017, *55* (2), 441–492.

Falch, Torberg and Bjarne Strøm, "Teacher turnover and non-pecuniary factors," *Economics of Education Review*, 2005, *24* (6), 611–631.

Feng, Li, David Figlio, and Tim Sass, "School accountability and teacher mobility," *Journal of Urban Economics*, 2018, *103*, 1–17.

Fredriksson, Peter and Björn Öckert, "The supply of skills to the teacher profession," 2007.

_ and _ , "Resources and student achievement - evidence from a Swedish policy reform," *Scandinavian Journal of Economics*, 2008, *110* (2), 277–296.

Gjefsen, Hege Marie and Trude Gunnes, "The effects of school accountability on teacher mobility and teacher sorting," 2016.

Hanushek, Erik, "The economics of schooling: Production and efficiency in public schools," *Journal of Economic Literature*, 1986, *24* (3), 1141–1177.

_ , John Kain, and Steven Rivkin, "Why public schools loose teachers," *Journal of Human Resources*, 2004, *39* (2), 326–354.

Haraldsvik, Marianne, "Influences on educational outcomes: Three essays on the role of parents, peers and choice." PhD dissertation, Norwegian University of Science and Technology 2012.

Harris, Douglas and Tim Sass, "Teacher training, teacher quality and student achievement," *Journal of Public Economics*, 2011, *95* (7-8), 798–812.

Hensvik, Lena, "Competition, wages and teacher sorting: four lessons from a voucher reform," *Economic Journal*, 2012, *122* (561), 799–824.

Hsieh, Chang-Tai and Miguel Urquiola, "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program," *Journal of Public Economics*, 2006, *90* (8-9), 1477–1503.

Jackson, Kirabo, "Student demographics, teacher sorting and teacher quality: Evidence from the end of school desegregation," *Journal of Labor Economics*, 2009, *27* (2), 213–256.

_ , "School competition and teacher labor markets: Evidence from charter school entry in North Carolina," *Journal of Public Economics*, 2012, *96* (5-6), 431–438.

_ , "Match quality, worker productivity, and worker mobility: Direct evidence from teachers," *Review of Economics and Statistics*, 2013, *95* (4), 1096–1116.

Joensen, Juanna and Skyt Helena Nielsen, "Is there a causal effect of high school math on labor market outcomes?," *Journal of Human Resources*, 2009, *44* (1), 171–198.

Karbownik, Krzysztof, "The determinants of teacher mobility in Sweden," 2014.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff, "Teacher sorting and the plight of urban schools: the importance of alternative labor market opportunities and non-pecuniary variation," *Educational Evaluation and Policy Analysis*, 2002, *24* (1), 37–62.

OECD, “Education Database: Enrolment by age,” March 2019.

Reber, Sarah, “Court-ordered desegregation: Successes and failures integrating American schools since Brown versus Board of Education,” *Journal of Human Resources*, 2005, *40* (3), 559–590.

Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff, “How teacher turnover harms student achievement,” *American Educationa Research Journal*, 2013, *50* (1), 4–36.

Scafidi, Benjamin, David Sjoquist, and Todd Stinebrickner, “Race, poverty, and teacher mobility,” *Economics of Education Review*, 2007, *26* (2), 145–159.

Shirrell, Matthew, “The effects of subgroup-specific accountability on teacher turnover and attrition,” *Education Finance and Policy*, 2018, *13* (3), 333–368.

Sims, Sam, “High-stakes accountability and teacher turnover: How do different school inspection judgements aggect teachers’ decisions to leave their school?,” 2016.

Skolverket, “Resursfördelning utifrån förutsättningar och behov,” Skolverket Rapport 330, Skolverket 2009.

Söderström, Martin and Roope Uusitalo, “School choice and segregation: Evidence from an admission reform,” *Scandinavian Journal of Economics*, 2010, *112* (1), 55–76.

Vlachos, Jonas, “Betygets värde: En analys av hur konkurrens påverkar betygssättningen vid svenska skolor,” *Konkurrensverket Uppdragsforskningsrapport*, 2010, (6).

Figures and Tables

Table 1: Descriptive Statistics

	Change in student credentials								
	1/3 downward			1/3 middle			1/3 upward		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pre-reform	Post-reform	p-value	Pre-reform	Post-reform	p-value	Pre-reform	Post-reform	p-value
Outcome variable									
One-year mobility	11.0	11.8	0.77	15.1	12.5	0.39	11.3	7.7	0.14
Treatment variable									
Incoming students' GPA	49.0	45.5	p<0.001	56.8	58.0	0.47	59.2	62.8	p<0.001
Teacher characteristics									
Female	56.6	53.8	0.50	54.1	54.8	0.87	45.3	46.0	0.86
Experience (years)	12.7	12.8	0.95	11.4	11.3	0.89	11.8	11.4	0.59
Has university diploma	76.2	75.3	0.81	69.2	71.0	0.64	73.7	76.8	0.38
Employed on fixed-term contract	18.1	22.2	0.23	25.8	27.2	0.70	23.7	23.2	0.87
Yearly earnings (1000 SEK)	243.2	247.1	0.59	218.8	229.8	0.10	215.8	222.7	0.32
Student characteristics (alternative treatment variables)									
First generation immigrant	18.8	21.1	p<0.001	10.4	10.4	0.86	10.8	11.6	0.17
Yearly parental income (1000 SEK)	331.4	338.8	0.13	420.8	441.8	0.01	410.0	443.3	p<0.001
Parental education	12.3	12.3	0.32	13.2	13.4	0.23	13.5	13.6	0.26
Paternal draft score	53.6	53.7	0.88	57.3	56.7	0.38	58.2	58.3	0.69
Number of schools		6			6			5	
Number of teachers		560			558			585	

Note: Columns (1), (4) and (7) present descriptive statistics for the last pre-reform year, columns (2), (5) and (8) present descriptive statistics for the first post-reform year, while columns (3), (6) and (9) present p-values for differences between pre- and post-reform years. All descriptive statistics present mean values, are based on the panel of 17 Stockholm schools in operation between 1994/1995 and 2004/2005, and refer to incoming first year students as far as aggregate school characteristics are concerned. Columns (1) to (3) describe a third of most downward shocked schools, columns (4) to (6) describe a third of middle tercile schools, and columns (7) to (9) describe a third of most upward shocked schools. Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school in the first post-reform year 2000 and mean students' credentials in the last pre-reform year 1999 in these same schools. Indicator variables are multiplied by 100.

Table 2: Placebo: Effects of post-reform changes in students' credentials on pre-reform changes in teacher turnover

	(1)	(2)	(3)
		Mobility	
	1-year	2-year	3-year
GPA	-0.176	-0.212	-0.115
	(0.215)	(0.183)	(0.193)
Mean of Y	17.4	23.7	30.7
Observations	1,867	1,998	1,977

Note: Teacher level regressions where each estimate comes from a separate regression. All point estimates are based on difference-in-differences regressions including school and year fixed effects as well as individual controls (see note in Table 4). The independent variable of interest measuring students' credentials is based on GPA from years 1999 and 2000 in column (1), years 1998 and 2000 in column (2), and 1997 and 2000 in column (3). Students' credentials are measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school. The dependent variables are lagged by one exposure-period, namely column (1) compares one-year mobility for years 1998 and 1999, column (2) compares two-year mobility for years 1996 and 1998, and column (3) compares three-year mobility for years 1994 and 1997. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

Table 3: Wald Estimator

	(1)	(2)	(3)
	Schools		
	1/3 upward shocked	1/3 downward shocked	Difference
Treatment: Student credentials - percentile ranked GPA from 9th grade in primary school.			
Year 2000 (post)	73.91	53.47	20.44***
	(13.22)	(17.28)	(1.28)
Year 1996 (pre)	56.35	57.21	-0.87
	(13.89)	(16.12)	(1.18)
Difference	17.56***	-3.74***	21.30***
	(1.10)	(1.34)	(1.73)
Dependent variable: 4-year mobility representing fully implemented reform			
Year 1999 (post)	24.09	30.28	-6.20*
	(42.84)	(46.02)	(3.68)
Year 1995 (pre)	40.46	25.49	14.97***
	(49.15)	(43.65)	(3.66)
Difference	-16.37***	4.79	-21.17***
	(3.76)	(3.60)	(5.16)
Wald estimate			
	-0.99***		
	(0.25)		

Note: Shock is defined as a difference in mean students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school in the first post-reform year 2000 and in the 1996 pre-reform year in these same schools. Based on this difference schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Only 17 schools that are present in the data in each year between 1994/1995 and 2004/2005 are included in the analysis. Dependent variable is defined as probability (multiplied by 100) of leaving a given school from school year 1995/1996 to 1999/2000 in pre-reform period and probability of leaving a given school from school year 1999/2000 to 2003/2004 in post-reform period. Independent (treatment) variable is defined as mean incoming students' credentials (only first-grade students who applied to school in the same year) in 1996 in pre-period and in 2000 in post-period. Differences report the interaction coefficients from regression of students' credentials or mobility on year dummy, upward shock dummy and their interaction. Wald estimate reports coefficient from instrumental variables regression of the probability that teacher leaves a given school on students' credentials, year dummy and upward shock dummy. Students' credentials are instrumented by interaction between year and shock. Heteroskedasticity robust standard errors and numbers rounded to second decimal.

Table 4: Main results: Effects of student credentials on teacher turnover

	(1) OLS	(2) OLS	(3) DD	(4) DD
Anticipation effect (1-year mobility)	-0.113***	-0.094*	-0.085	-0.176
	(0.019)	(0.055)	(0.226)	(0.215)
Mean of Y/Observations	13.8/35,796		17.4/1,867	
One grade reshuffled (2-year mobility)	-0.129***	-0.092	-0.360*	-0.462**
	(0.021)	(0.062)	(0.199)	(0.188)
Mean of Y/Observations	19.4/35,100		22.9/1,802	
Two grades reshuffled (3-year mobility)	-0.189***	-0.011	-0.655***	-0.748***
	(0.023)	(0.062)	(0.210)	(0.197)
Mean of Y/Observations	24.8/35,258		28.5/1,799	
All grades reshuffled (4-year mobility)	-0.210***	-0.056	-0.861***	-1.002***
	(0.024)	(0.065)	(0.220)	(0.207)
Mean of Y/Observations	29.8/35,794		33.4/1,781	
Stockholm municipality		X	X	X
School and year fixed effects			X	X
Individual controls	X	X		X

Note: Teacher level regressions where each estimate comes from a separate regression. Column (1) presents correlations conditional on individual teacher observable characteristics using data on all Swedish high school teachers. Column (2) presents correlations conditional on individual teacher observable characteristics using data on high school teachers in Stockholm municipality. Columns (3) and (4) present difference-in-differences estimates for the sample from column (2). Column (3) excludes while column (4) includes individual level controls. Control variables include: indicators for teacher's gender, nativity, marital status, having university diploma, and specialization (science, vocational or special education) as well as experience in years. The dependent variables (in rows; multiplied by 100) are defined according to columns (1) and (3) in Table A8. The treatment variables of interest measuring students' credentials are defined according to columns (2) and (4) in Table A8. Students' credentials are measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

Table 5: Correlates of student ability

	(1) Unconditional	(2) Conditional
First generation immigrant students	1.118***	0.079
	(0.421)	(0.496)
GPA		-0.980***
		(0.246)
Parental income in 100,000 SEK	-0.070**	0.344***
	(0.035)	(0.073)
GPA		-2.788***
		(0.441)
Parental education	-19.071***	-10.882**
	(4.840)	(5.303)
GPA		-0.778***
		(0.227)
Combined cognitive and non-cognitive paternal skills	-1.219***	-0.525
	(0.440)	(0.524)
GPA		-0.670**
		(0.274)
Mean of Y/Observations	33.4/1,781	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls based on the specification from row (4) and column (4) of Table 4. In column (1) I substitute students' GPA with other average school-level first grade students' characteristics: fraction of immigrants (row (1)), parental income (row (2)), parental education (row (3)) and paternal cognitive and non-cognitive military assessments (row (4)). These are correlated with student GPA at 0.43, 0.82, 0.92 and 0.82, respectively. Each row (in column (1)) and pair of estimates (in column (2)) come from separate regressions. In column (2) I keep these alternative measures but also include GPA. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

Table 6: Heterogeneity by teacher characteristics

	Characteristic	Group	(1) Fraction [%]	(2) Mean mobility	(3) Estimate	(4) Implied effect (%)
(1)	University education	Yes	72.9	29.9	-1.062*** (0.234)	-3.6
		No	27.1	42.7	-0.684 (0.439)	-1.6
(2)	Experience (years)	0-5	24.0	53.3	-1.087** (0.489)	-2.0
		6-15	36.0	33.5	-0.856** (0.400)	-2.6
		16+	40.0	21.2	-1.020*** (0.273)	-4.8
(3)	Gender	Male	47.0	35.1	-1.177*** (0.296)	-3.4
		Female	53.0	31.8	-0.822*** (0.294)	-2.6
(4)	Subject taught	Science	10.3	40.8	-1.347 (0.843)	-3.3
		Other	89.7	32.5	-0.983*** (0.215)	-3.0
(5)	Type of contract	Permanent	80.1	27.6	-0.904*** (0.219)	-3.3
		Fixed-term	19.9	56.3	-1.171** (0.515)	-2.1

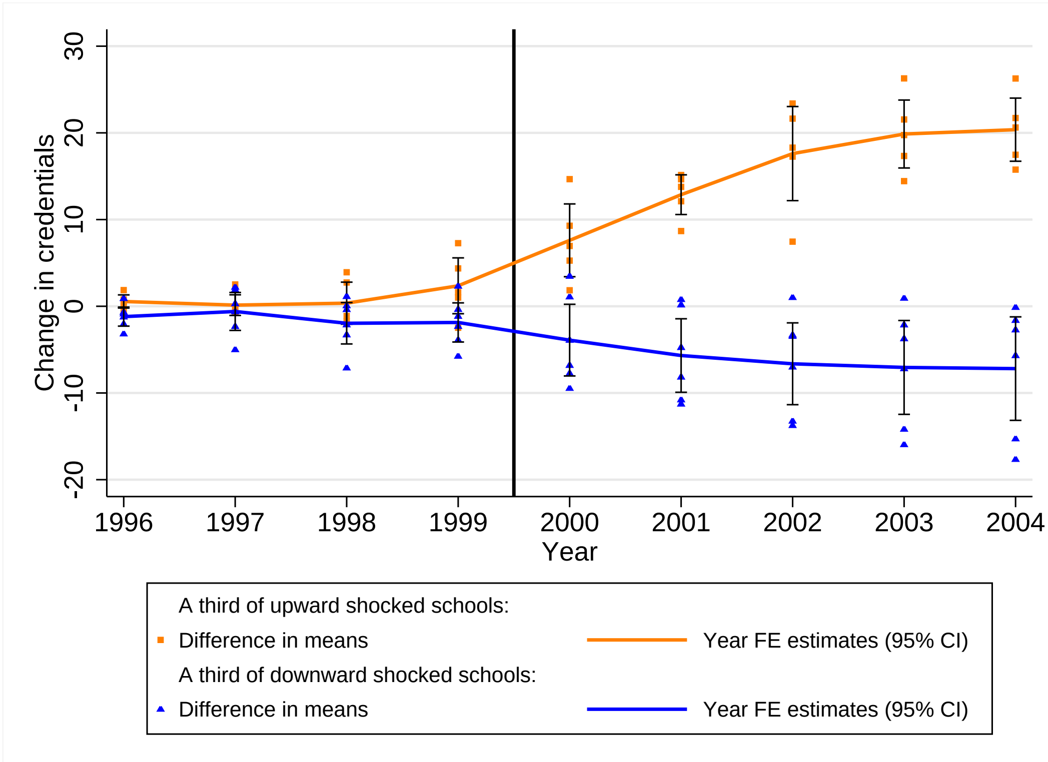
Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls based on the specification from row (4) and column (4) of Table 4. Each row reports estimates from a separate regression. Column (1) reports fraction of individuals in each group, column (2) reports mean 4-year mobility in each group, column (3) reports point estimates, and column (4) reports implied effect sizes. In row (1) a university graduate is defined as an individual graduating three, four or five year-long university education or individual with a research degree. Other forms of post-secondary education are not treated as university graduates. Science teachers are defined as those teaching: mathematics, physics, chemistry, biology and computer science. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

Table 7: Heterogeneity by teacher destinations

	(1)	(2)	(3)	(4)	(5)	(6)
		All teachers			Teachers with university degree	
	Mean mobility	Estimate	Implied effect (%)	Mean mobility	Estimate	Implied effect (%)
Mobility within schooling (1)	16.9	-0.747***	-4.4	16.5	-0.788***	-4.8
		(0.162)			(0.184)	
Mobility within public high schools (1a)	8.3	-0.520***	-6.3	7.5	-0.424***	-5.7
		(0.131)			(0.146)	
Mobility to compulsory public schools (1b)	7.5	-0.179*	-2.4	8.2	-0.266**	-3.2
		(0.098)			(0.115)	
Mobility to private school (1c)	1.1	-0.047	-4.4	0.8	-0.099*	-12.9
		(0.050)			(0.054)	
Out of schooling sector (2)	16.5	-0.255	-1.6	13.4	-0.274	-2.0
		(0.173)			(0.186)	
To a higher average student ability school (3)	4.9	-0.333***	-6.7	4.3	-0.232**	-5.4
		(0.098)			(0.109)	
Observations		1,781			1,298	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls based on the specification from row (4) and column (4) of Table 4. Each row in columns (2) and (5) reports estimates from separate regressions. Columns (1) and (4) present means of the dependent variables; column (2) presents estimates for all teachers; column (5) presents estimates for teachers with university diploma; while columns (3) and (6) present implied effect sizes in these two samples. Dependent variables are: indicator if a teacher leaves for another teaching position in a different school (row 1); indicator if a teacher leaves for another teaching position in a different public high school (row 1a); indicator if a teacher leaves for another teaching position in a public compulsory school (row 1b); indicator if a teacher leaves for another teaching position in a primary or secondary private school (row 1c); indicator if a teacher leaves for another occupation outside of teaching (row 2); and indicator if a teacher leaves for high school with higher average student GPA than in their initial employment (row 3). All indicator variables are multiplied by 100. Mean values in rows (1a) to (1c) add up to mean value in row (1); mean values in rows (1) and (2) add up to total mobility measure used in Table 4. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

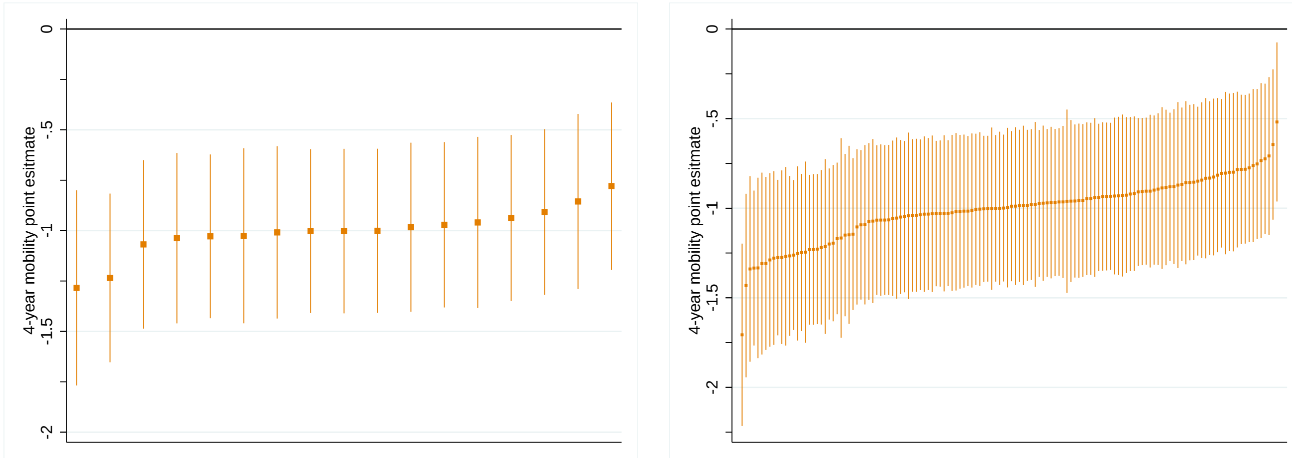
Figure 1: School-level average GPA between 1995 and subsequent years



Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school in the first post-reform year 2000 and mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Each point represents a difference between average students' credentials in these schools in a given year (1996 to 2004) and average students' credentials in these same schools in 1995. Lines plot coefficients and 95% confidence intervals from regressing these differences on year dummies (one for each year between 1996 and 2004). Black solid vertical line depicts reform implementation. Only 11 schools (top and bottom tercile) that are present in the data in each year between 1994/1995 and 2004/2005 are included in the graph. Heteroskedasticity robust standard errors.

Figure 2: Robustness of the preferred difference-in-differences estimates: Excluding one or two schools at a time

Panel A: Dropping one school at a time | Panel B. Dropping two schools at a time



Note: Estimates based on the specification from row (4) and column (4) of Table 4 with 95% confidence intervals. Panel A presents estimates dropping one school at a time while Panel B presents estimates dropping two schools at a time. Heteroskedasticity robust standard errors.

Supplemental Materials (not for publication)

Sample Construction

I construct the sample of high school teachers for the school years 1991/1992 to 2004/2005. The information about teachers comes from the teacher registry and the analysis focuses on teachers working in grades 1 to 3 of secondary education (high school) that were in operation in Stockholm municipality prior to school year 1999/2000. Teachers who are on unpaid leave of absence or whose workloads are zero hours (i.e., they do not perform any pedagogical duties) are excluded from the analysis. Such teachers are treated neutrally in terms of mobility if they come back after the absence period to the same school. Similarly, I exclude teachers who are employed as principals, study counselors etc. In each year if a teacher has multiple entries in the registry, the observation with the highest workload is selected irrespective of whether it is at the same or at different schools. The workload of teachers having multiple positions at the same school is not summed and the highest workload position is selected. The teacher registry is a high quality data set, that allows recovering information on school location (unique identifier), school ownership and type, teacher certification, workload, employment type (fixed-term vs. permanent), education, and position recorded in October of each year.

Teacher experience is not available for all years, and therefore, I use predicted experience in the analysis. In particular, since the teacher registries date back to 1979 I explore this feature to construct the "in teaching predicted experience" variable. I create a panel of all teachers between 1979 and 2006 and link it to population enlistment data between 1985 and 2006 in order to obtain teacher's birth date. I then use all this information and tenure data provided in the later registries (since 1999 onward) to construct the predicted measure of experience.

Teachers are linked (using unique identifier) to population register, which covers all individuals living in Sweden. The register includes information on gender, marital status, age, family composition (using unique family identifier), immigration history and education. The analysis is restricted to teachers aged 25-58 years in order to abstract from mobility driven by educational attainment and retirement decisions.

The students' characteristics are based on "school in" and "school out" pupil registries. The secondary school composition is based on all the students that are in a school in a given year. The ability of students in secondary school is measured based on their 9th grade GPA. I first percentile rank all students graduating lower secondary school in each cohort and for each subject, and then take the average across all subjects tested for each student. This average GPA is then percentile ranked again. Students from lower-secondary schools are then matched to upper secondary (high school) school registers. In the empirical analysis I use either mean ability in upper secondary school across all cohorts in a given school year or only across the first-grade admitted freshman in a given school year.

I match students to their parents using unique family identifier and obtain the family level socioeconomic indicators: mean parental income, mean parental education, and the cognitive and non-cognitive skill of the fathers from the military enlistment. Income is measured as a gross salary plus income from business and self-employment plus any work-related allowances. Investment losses are not included, and thus, income is lower-bounded at zero. The enlistment registry covers period 1969 to 2006 and provides information on cognitive and non-cognitive assessments. All skill measures are percentile ranked by year of draft. The data is linked to students' fathers using the unique personal identifier.

Finally, having a data set with teachers and students I match the two using the unique school identifier. I exclude schools with less than three employed teachers (in full time equivalence) and schools with less than 15 students in any given school year. I further restrict the analysis to teachers aged 25-58 years as explained above. I then select schools that operate within the municipality of Stockholm and were in operation prior to school year 1999/2000. Finally, I drop Skärholmens Gymnasium because this school did not admit any new students in school year 1998/1999. In this paper I primarily focus on a balanced panel of 17 Stockholm public high schools, i.e. I restrict the sample to schools present in the data for all years between 1994/1995 and 2004/2005, and to keep institutional homogeneity I exclude one private school that fulfills these criteria.

Tables

Table A1: Comparison of Stockholm municipality schools and other Swedish schools

	(1)	(2)	(3)	(4)	(5)	(6)
	Pre-period = 1999/2000			Post-period = 2000/2001		
	Sweden	Stockholm	p-value (1) vs. (2)	Sweden	Stockholm	p-value (4) vs. (5)
One-year mobility	12.9	14.2	0.19	11.0	14.9	p<0.001
Female teacher	48.9	55.4	p<0.001	48.4	52.9	p<0.001
Teacher experience (years)	12.0	11.0	p<0.001	11.9	11.0	p<0.001
Teacher with university diploma	65.2	71.5	p<0.001	65.0	72.0	p<0.001
Teacher employed on fixed-term contract	20.2	21.2	0.38	20.4	21.4	0.41
Yearly teacher earnings in 1000 SEK	224.0	217.3	p<0.001	225.9	225.6	0.91
Students' GPA	49.0	56.7	p<0.001	49.0	57.3	p<0.001
First generation immigrant	8.3	12.7	p<0.001	8.5	13.3	p<0.001
Yearly parental income in 1000 SEK	337.6	395.6	p<0.001	355.8	422.9	p<0.001
Parental education	12.0	13.1	p<0.001	12.2	13.2	p<0.001
Paternal draft score	51.2	57.4	p<0.001	51.6	57.2	p<0.001
Number of teachers	20,795	1,304		21,681	1,362	

Note: Columns (1) and (4) present means for all high school teachers in Sweden (excluding Stockholm municipality) in years 1999/2000 and 2000/2001 from schools that were in operation prior to school year 1999/2000. Columns (2) and (5) present means for all high school teachers in Stockholm municipality in years 1999/2000 and 2000/2001 from schools that were in operation prior to school year 1999/2000. Columns (3) and (6) present p-values for differences between Sweden and Stockholm columns in pre- and post-periods.

Table A2: Exploring stability of the preferred esitmates using alternative samples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel of schools				Repeated cross-section of schools (94-04). Expanding sample with schools:		
	94-04	91-04	94-98	91-98	Public	Private	Both
GPA	-1.002***	-0.933***	-0.887***	-0.811***	-0.885***	-0.999***	-0.883***
	(0.207)	(0.209)	(0.205)	(0.207)	(0.195)	(0.208)	(0.206)
Mean of Y	33.4	32.4	32.0	31.0	31.0	33.6	30.7
Number of schools	17	15	18	16	21	25	29
Observations	1,781	1,657	1,926	1,802	2,127	1,913	2,179

Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls based on the specification from row (4) and column (4) of Table 4 but using alternative samples. Column (1) replicates the main result from Table 4, column (2) is based on a more restrictive panel of schools in operation for all years from 1991 to 2004, column (3) is based on a panel of schools in operation for all years between 1994 and 1998, column (4) is based on a panel of schools in operation for all years between 1991 and 1998, column (5) is based on repeated cross-section of schools and adds public schools to those in column (1), column (6) is based on repeated cross-section and adds private schools to those in column (1), and column (7) is based on repeated cross-section and adds both public and private schools to those in column (1). Sample in column (7) presents the largest possible sample of schools to be included that were in operation before the reform. Heteroskedasticity robust standard errors.

Table A3: Comparison between schools included in and excluded from the main empirical sample

	(1) Included schools	(2) Excluded schools (public)	(3) Excluded schools (private)	(4) Excluded schools (public + private)	(5) p-value (1) vs. (2)	(6) p-value (1) vs (3)	(7) p-value (1) vs (4)
Outcome variable							
One-year mobility	14.3	13.4	16.3	14.2	0.53	0.35	0.92
Treatment variable							
Incoming students' GPA	54.1	45.9	65.6	50.8	p<0.001	p<0.001	p<0.001
Teacher characteristics							
Female	53.0	57.5	45.2	54.5	0.02	0.01	0.38
Experience (years)	11.8	11.5	6.7	10.3	0.25	p<0.001	p<0.001
Has university diploma	73.0	75.3	58.9	71.2	0.19	p<0.001	0.23
Employed on fixed-term contract	20.7	22.4	12.5	19.9	0.29	p<0.001	0.56
Yearly earnings in 1000 SEK	208.6	211.4	153.5	197.0	0.37	p<0.001	p<0.001
Student characteristics (alternative treatment variables)							
First generation immigrant	11.9	17.6	7.1	15.0	p<0.001	p<0.001	p<0.001
Yearly parental income in 1000 SEK	352.7	309.2	382.4	327.3	p<0.001	p<0.001	p<0.001
Parental education	12.9	12.1	13.0	12.3	p<0.001	0.31	p<0.001
Paternal draft score	54.8	50.7	56.2	52.1	p<0.001	p<0.001	p<0.001
Number of schools	17	4	8	12			
Number of teachers	4,747	796	263	1,059			

Note: Column (1) presents means for school included in the primary empirical analysis, column (2) presents means for excluded public schools, column (3) presents means for excluded private schools, column (4) presents means for excluded public and private schools together, while subsequent columns report p-values for the differences between column (1) and columns (2), (3), and (4), respectively.

Table A4: Characteristics of schools by terciles of pre-reform student ability

	Terciles of pre-reform student ability			
	(1)	(2)	(3)	(4)
	Bottom	Middle	Top	p-value
Outcome variable				
One-year mobility	13.9	10.2	13.6	0.309
Treatment variable				
Incoming students' GPA	35.9	57.7	71.9	p<0.001
Teacher characteristics				
Female	48.7	55.8	51.2	0.225
Experience (years)	11.3	11.9	12.8	0.047
Has university diploma	56.4	81.5	80.6	p<0.001
Employed on fixed-term contract	19.0	21.1	27.9	0.047
Yearly earnings (1000 SEK)	228.7	238.8	208.1	p<0.001
Student characteristics (alternative treatment variables)				
First generation immigrant	16.2	16.0	7.2	p<0.001
Yearly parental income (1000 SEK)	299.7	398.3	466.5	p<0.001
Parental education	11.8	12.9	14.4	p<0.001
Paternal draft score	49.4	57.9	61.9	p<0.001
Number of schools	6	6	5	17

Note: This table presents descriptive statistics by terciles of pre-reform (1999-00) student ability. Column (1) presents means for the bottom tercile, column (2) presents means for the middle tercile, column (3) presents means for the top tercile while column (4) presents p-values from joint equality test across columns (1) to (3). Heteroskedasticity robust standard errors for p-values computation.

Table A5: Effects of average student credentials on teacher turnover

	(1) OLS	(2) DD	(3) DD
Anticipation effect (1-year mobility)	-0.104	-0.129	-0.277
	(0.065)	(0.433)	(0.419)
Mean of Y/Observations		17.4/1,867	
One grade reshuffled (2-year mobility)	-0.077	-0.923**	-1.119***
	(0.071)	(0.402)	(0.389)
Mean of Y/Observations		22.9/1,802	
Two grades reshuffled (3-year mobility)	0.028	-1.127***	-1.231***
	(0.071)	(0.357)	(0.348)
Mean of Y/Observations		28.5/1,799	
All grades reshuffled (4-year mobility)	-0.012	-1.526***	-1.786***
	(0.075)	(0.376)	(0.366)
Mean of Y/Observations		33.4/1,781	
School and year fixed effects		X	X
Individual controls	X		X

Note: This table replicates analysis from columns (2) to (4) in Table 4 but changes treatment variable from GPA of the incoming freshman cohort of students to average GPA at the school. Heteroskedasticity robust standard errors.

Table A6: Effects of student credentials on teacher turnover without accounting for the anticipation effect

	(1) OLS	(2) DD	(3) DD
One grade reshuffled (1-year mobility)	-0.017	-0.152	-0.171
	(0.049)	(0.214)	(0.205)
Mean of Y/Observations		11.5/1,703	
Two grades reshuffled (2-year mobility)	-0.179***	-0.232	-0.365*
	(0.062)	(0.201)	(0.189)
Mean of Y/Observations		23.6/1,902	
All grades reshuffled (3-year mobility)	-0.125**	-0.604***	-0.702***
	(0.062)	(0.203)	(0.195)
Mean of Y/Observations		27.5/1,837	
School and year fixed effects		X	X
Individual controls	X		X

Note: This table presents main results (based on columns (2) to (4) in Table 4) but assuming no anticipation effects. Thus, treatment is assigned in year 2000-01 in the post-period and in years 1997-98, 1998-99 and 1999-00 for one-, two-, and three-year mobility in the pre-period. Conversely, the dependent variable in the post-period is defined as indicator leaving the school from school year 2000-01 to 2001-02 for one-year mobility, from 2000-01 to 2002-03 for two-year mobility, and form 2000-01 to 2003-04 for thee year mobility. These same mobility windows for the pre-period are defined as moves from 1999-00 to 2000-01, from 1998-99 to 2000-01, and from 1997-98 to 2000-01, respectively. All indicator variables (outcomes) are multiplied by 100. Heteroskedasticity robust standard errors.

Table A7: Testing standard errors assumptions in the main results

	(1) Anticipation effect	(2) One grade reshuffled	(3) Two grades reshuffled	(4) All grades reshuffled
Estimate	-0.176	-0.462	-0.748	-1.002
Heteroskedasticity robust standard errors	(0.215)	(0.188)	(0.197)	(0.207)
Standard errors clustered at the school level	(0.479)	(0.421)	(0.351)	(0.400)
Standard errors clustered at the school×year level	(0.336)	(0.297)	(0.247)	(0.282)
Standard errors clustered at the teacher level	(0.210)	(0.185)	(0.193)	(0.203)
Robust standard errors from first difference with aggregated data	(0.501)	(0.451)	(0.371)	(0.430)
# Teachers	1,867	1,802	1,799	1,781
# Schools	17	17	17	17

Note: This table replicates the specification from column (4) of Table 4 but with alternative standard errors. It reports (i) robust standard errors as a reference; (ii) standard errors clustered at the school level; (iii) standard errors clustered at the school×year level, thus allowing interdependence between teachers in a school in a specific year but not across years; (iv) standard errors clustered at the teacher level to account for the fact that some teachers are observed in the data multiple times (v) heteroskedasticity robust standard errors from regressing the first-differences on the treatment variable using aggregated data residualized with control variables as well as school and year fixed effects. All models, except the first difference analysis, include school and year fixed effects, and individual level controls. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions.

Table A8: Definitions of treatment and outcome variables for the main estimates

	(1) Post-period mobility	(2) Post-period GPA	(3) Pre-period mobility	(4) Pre-period GPA
Anticipation effect (1-year mobility)	99/00 to 00/01	2000	98/99 to 99/00	1999
One grade reshuffled (2-year mobility)	99/00 to 01/02	2000	97/98 to 99/00	1998
Two grades reshuffled (3-year mobility)	99/00 to 02/03	2000	96/97 to 99/00	1997
All grades reshuffled (4-year mobility)	99/00 to 03/04	2000	95/96 to 99/00	1996

Note: This table presents definitions of treatment and outcome variables as they relate to the reform implementation for different lengths of mobility window reported in rows. These definitions correspond to results presented in Table 4. First row defines mobility as teachers leaving in period t+1, second row in t+2, third row in t+3, and fourth row in t+4. Columns (1) and (2) define the post-reform period dependent and treatment variables. Columns (3) and (4) define the pre-reform period dependent and treatment variables.

Table A9: School-level equilibrium effects

	(1)	(2)	(3)
	1-year	2-year	3-year
		Panel A: Number of students	
GPA	1.178	2.349	3.645
	(1.910)	(2.459)	(2.576)
Mean of Y	749.0	758.9	760.0
		Panel B: Number of teachers	
GPA	0.168	0.103	0.159
	(0.246)	(0.349)	(0.531)
Mean of Y	59.4	61.5	63.3
		Panel C: Student-teacher ratio	
GPA	0.044	0.088	0.082
	(0.073)	(0.099)	(0.106)
Mean of Y	12.9	12.6	12.6
Observations	34	34	34

Note: School level difference-in-differences regressing number of students (panel A), number of teachers (panel B) and student-teacher ratio (panel C) on students' credentials as well as school and year fixed effects. The dependent variables are measured in 1999 in the pre-reform period and in 2000, 2001, 2002 in the post-reform period for one-, two-, and three-year exposure, respectively; while the independent variable is measured in 1999 in pre- and in 2000 in post-period. Students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the analysis. Heteroskedasticity robust standard errors.

Table A10: Effects on teacher hires and earnings

	(1) Probability of new hire	(2) Earnigs of all teachers	(3) Earnigns of new hires
	Panel A: Accounting for anticipation effects		
GPA	-0.566**	256	-1967**
	(0.281)	(381)	(776)
Mean of Y	49.6	237169	214275
Observations	1,723	1,963	854
	Panel B: Without accounting for anticipation effects		
GPA	-0.605**	-12	-1941**
	(0.283)	(532)	(941)
Mean of Y	41.6	243992	207529
Observations	1,758	1,764	731

Note: This table presents results for teacher hires and earnings corresponding to the specification from row (4) and column (4) of Table 4 in panel A and to the specification from row (3) and column (3) of Table A6 in panel B. Dependent variable in column (1) of panel A is defined as teacher hire in school year 1999-00 that was not working in this school in year 1995-96 in the pre-period and teacher hire in school year 2003-04 that was not working in this school in school year 1999-00 in the post-period while the treatment is defined as GPA in school years 1999-00 in the pre-period and GPA in school year 2000-01 in the post-period. Dependent variable in column (1) of panel B is defined as teacher hire in school year 2000-01 that was not working in this school in year 1997-98 in the pre-period and teacher hire in school year 2003-04 that was not working in this school in school year 2000-01 in the post-period while the treatment is defined as GPA in school years 1999-00 in the pre-period and GPA in school year 2000-01 in the post-period. Dependent variable in column (2) of panel A is defined as annual teacher earnings in year 1998 in the pre-period and annual teacher earnings in year 2002 in the post-period while the treatment is defined as GPA in school years 1998-99 in the pre-period and GPA in school year 2000-01 in the post-period. Dependent variables in column (2) of panel B is defined as annual teacher earnings in year 1999 in the pre-period and annual teacher earnings in year 2002 in the post-period while the treatment is defined as GPA in school years 1999-0 in the pre-period and GPA in school year 2000-01 in the post-period. Dependent variable in column (3) of panel A is defined as annual teacher earnings in year 1999 in the pre-period and annual teacher earnings in year 2003 in the post-period while the treatment is defined as GPA in school years 1999-00 in the pre-period and GPA in school year 2000-01 in the post-period. Dependent variables in column (3) of panel B is defined as annual teacher earnings in year 2000 in the pre-period and annual teacher earnings in year 2003 in the post-period while the treatment is defined as GPA in school years 1999-0 in the pre-period and GPA in school year 2000-01 in the post-period. Samples in column (3) are restricted to newly hired teachers as defined in column (1). Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the analysis. Heteroskedasticity robust standard errors.

Table A11: Heterogeneity by pre-reform average school GPA

	(1)	(2)	(3)
		Mobility	
	Total	Within	Quit
	Panel A: Bottom tercile of student GPA		
Preferred estimate	-1.387***	-0.556	-0.830**
	(0.463)	(0.375)	(0.342)
Mean of Y	22.7	11.6	11.1
	Panel B: Middle tercile of student GPA		
Preferred estimate	-1.357***	-1.285***	-0.073
	(0.438)	(0.342)	(0.326)
Mean of Y	28.4	18.2	10.3
	Panel C: Top tercile of student GPA		
Preferred estimate	-0.362	0.029	-0.391
	(0.363)	(0.129)	(0.357)
Mean of Y	20.9	8.7	12.2
p-value (A=B=C)	0.113	p<0.001	0.275

Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls based on the specification from row (4) and column (4) of Table 4. Each point estimate comes from a separate regression. Student ability at school-level is divided into terciles based on the GPA in school year 1996/1997 (pre-period). Panel A presents results for the sample of schools in the bottom tercile of GPA in the pre-period, panel B for the middle tercile of GPA in the pre-period, and panel C for the top tercile of GPA in the pre-period. Column (1) presents results for an indicator of leaving a current school (total mobility equivalent to estimates in Table 4), column (2) presents results for indicator for leaving for another position within primary or secondary schooling, and column (3) presents results for indicator for leaving for another occupation outside of teaching. All indicator variables are multiplied by 100. Bottom row presents the joint significance tests for the analyzed groups and outcomes. Sample sizes are 423, 429 and 401 for panels A through C, respectively. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

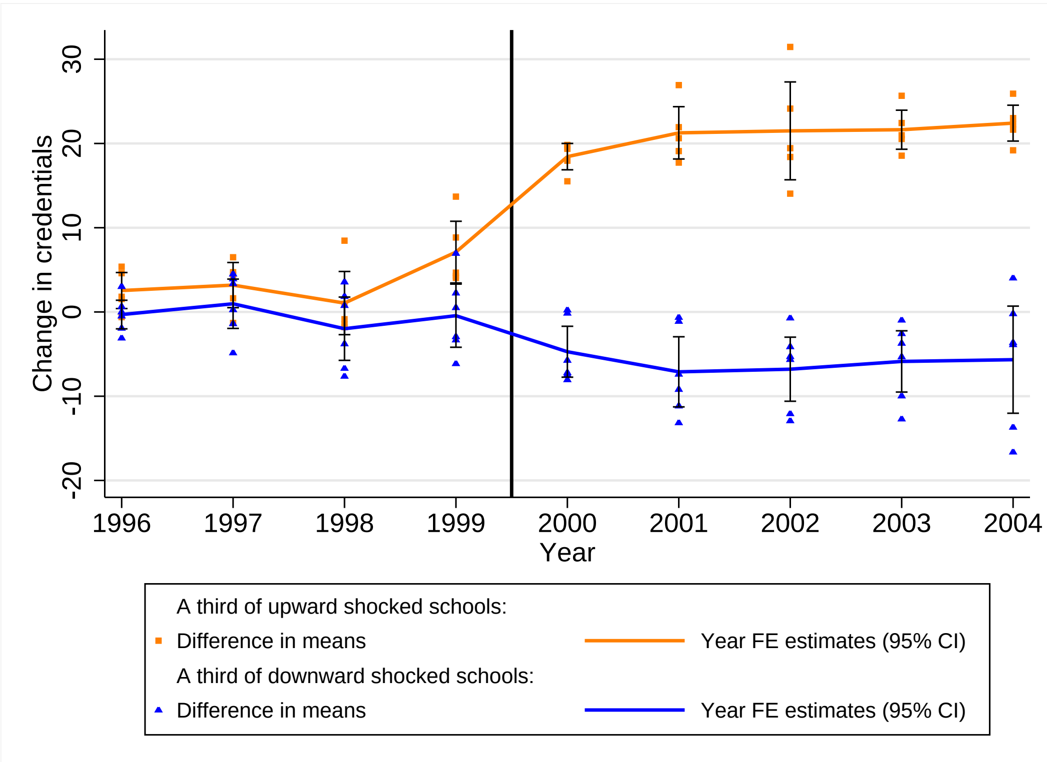
Table A12: Heterogeneity by ability of incoming students

Tercile	(1) Bottom	(2) Middle	(3) Top
	Panel A: Total turnover		
Fraction of students in k-th tercile	1.122***	-0.172	-0.354***
	(0.187)	(0.138)	(0.114)
Mean of Y/Observations		33.4/1,781	
	Panel B: Within teaching mobility		
Fraction of students in k-th tercile	0.698***	0.097	-0.358***
	(0.159)	(0.107)	(0.086)
Mean of Y/Observations		16.9/1,781	
	Panel C: Quits		
Fraction of students in k-th tercile	0.424***	-0.269**	0.005
	(0.147)	(0.117)	(0.097)
Mean of Y/Observations		16.5/1,781	

Note: Teacher level regressions controlling for school and year fixed effects as well as individual level controls and based on the specification from row (4) and column (4) of Table 4. Each coefficient comes from a separate regression. Terciles defined in columns (1) to (3). Panel A presents results for an indicator of leaving a current school (total mobility equivalent to estimates in Table 4), panel B presents results for indicator for leaving for another position within primary or secondary schooling, and panel C presents results for indicator for leaving for another occupation outside of teaching. All indicator variables are multiplied by 100. Only 17 schools that are observed in each year between 1994/1995 and 2004/2005 are included in the regressions. Heteroskedasticity robust standard errors.

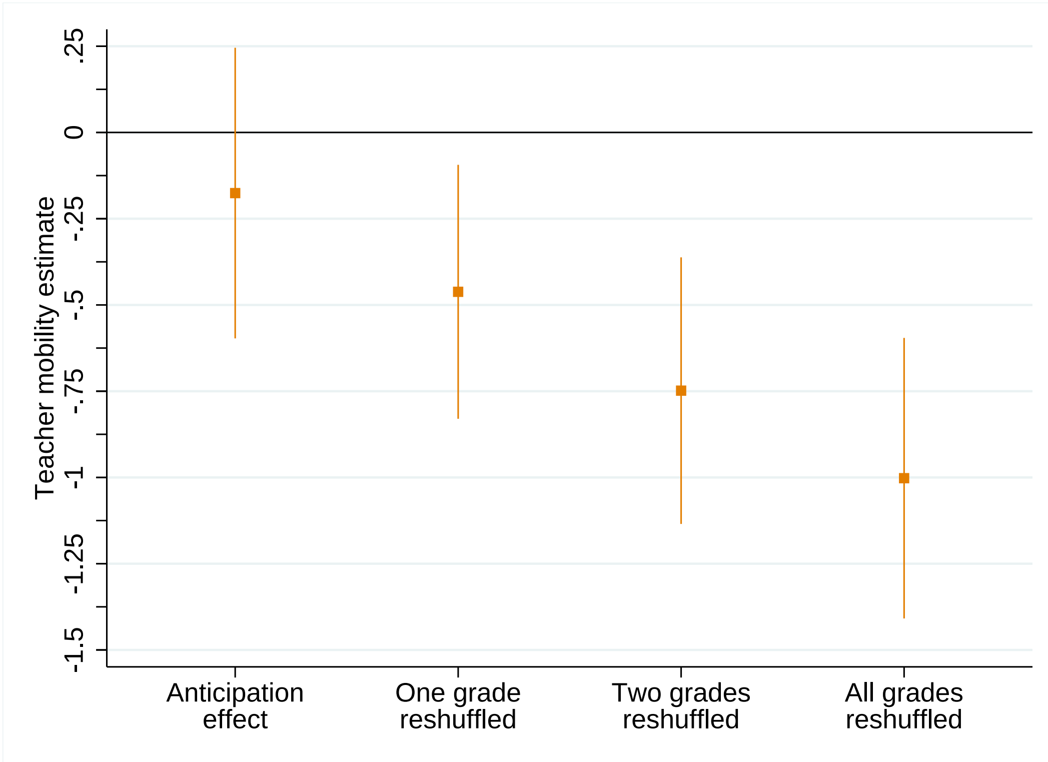
Figures

Figure A1: School-level average GPA between 1995 and subsequent years for first grade students who applied to high school in the same year as they graduated



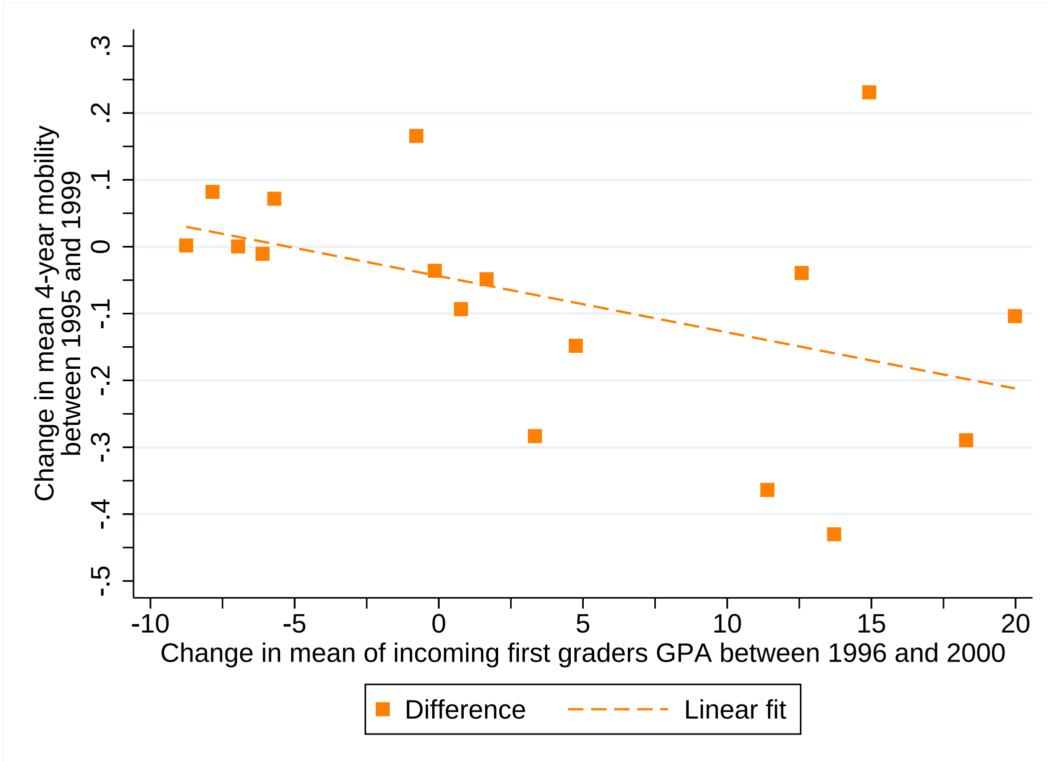
Note: Shock is defined as a difference between mean students' credentials measured by primary school 9th grade GPA (only first-grade students who applied to school in the same year) in a given high school in the first post-reform year 2000 and alike defined mean students' credentials in the last pre-reform year 1999 in these same schools. Based on the shock schools are divided into these that experience the most positive change (one-third upward shocked schools) and these that experience the least positive change (one-third downward shocked schools). Each point represents a difference between incoming students' credentials in these schools in a given year (1996 to 2004) and incoming students' credentials in these same schools in 1995. Lines plot coefficients and 95% confidence intervals from regressing these differences on year dummies (one for each year between 1996 and 2004). Black solid vertical line depicts reform implementation. Only 11 schools (top and bottom tercile) that are present in the data in each year between 1994/1995 and 2004/2005 are included in the graph. Heteroskedasticity robust standard errors.

Figure A2: Difference-in-differences estimates by reform rollout



Note: Estimates and 95% confidence intervals based on specifications from column (4) of Table 4.

Figure A3: Difference-in-differences estimates: school-level graphical representation



Note: School-level difference-in-differences analysis based on the specification from row (4) and column (4) of Table 4. Values on the vertical axis represent differences in mean 4-year mobility between 1995 (pre-reform) and 1999 (post-reform). Values on the horizontal axis represent changes in mean students' credentials between 1996 (pre-reform) and 2000 (post-reform). Student credentials are based on first grade students who applied to high schools in the same year and are measured using primary school 9th grade GPA. Line represents linear regression fit. Only 17 schools that are present in the data in each year between 1994/1995 and 2004/2005 are included in the analysis.