Quota vs Quality?
Long-Term Gains from an Unusual Gender Quota

Ursina Schaede* Ville Mankki†

March 2022

Abstract

We evaluate equity-efficiency trade-offs from admissions quotas by examining effects on output once beneficiaries start producing in the relevant industry. In particular, we document the impact of abolishing a 40% quota for male primary school teachers on their pupils' long-run outcomes. The quota had advantaged academically lower-scoring male university applicants, and its removal cut the share of men among new teachers by half. We combine this reform with the timing of union-mandated teacher retirements to isolate quasi-random variation in the local share of male quota teachers. Using comprehensive register data, we find that pupils exposed to a higher share of male quota teachers during primary school transition more smoothly to post-compulsory education and have higher educational attainment and labor force attachment at age 25. Pupils of both genders benefit similarly from exposure to male quota teachers. Several pieces of evidence suggest that the quota improved the allocation of talent due to mending imperfections in the selection process.

We are grateful to David Yanagizawa-Drott, Joachim Voth, Claudia Goldin, Ulf Zölitz, and Pekka Räihä for their guidance throughout this project. We thank David Autor, Sara Bagagli, Kelly Bedard, Anne Sofie Beck Knudsen, Augustin Bergeron, Peter Bergman, Anne Brenøe, Lorenzo Casaburi, Raj Chetty, Ana Costa-Ramón, Alessandro Ferrari, Matteo Greco, David Hémous, Andrea Hofer, Kristiina Huttunnen, Daniel Kopp, Mika Kortelainen, Rafael Lalive, Elisa Macchi, Ross Mattheis, Tatiana Mocanu, Matt Notowidigdo, Morten Olsen, Javaeria Qureshi, Claude Raisaro, Heather Sarsons, Helena Skyt-Nielsen, Hanna Virtanen and various conference and seminar participants for helpful comments and suggestions. We thank Topias Jalo and his colleagues at Statistics Finland, as well as administrative staff at the Universities of Aalto, Tampere and Zurich for their support. We are especially grateful to Kristiina Huttunen and Matti Liski for providing a home for the project at Aalto University. We gratefully acknowledge financial support from the Yrjö Jahnsson foundation.

*MIT, Department of Economics: ursina@mit.edu

†University of Turku, Department of Teacher Education

1 Introduction

Are affirmative action policies, such as quotas, inefficient? While many countries around the world are deliberating quotas to increase the representation of women and underrepresented minorities in business and politics, there is also wide-spread push-back against such initiatives (UN, 2019; Long, 2019). Universities in the United States and elsewhere face increasing judicial challenges for admissions policies alleged to advantage underrepresented groups (Green, 2022; Leman, 2021; Dhume, 2019). However, we still have a relatively limited understanding of the equity-efficiency trade-offs associated with these policies, largely due to a lack of opportunities to observe their impact on explicit measures of output in real-world contexts.

From a theoretical perspective, the effects of affirmative action and quota policies on output are ambiguous: On one hand, when quotas in educational institutions and workplaces are binding, they require these organizations to "lower the bar" by admitting less qualified applicants who would have been rejected otherwise (Welch, 1976; Lundberg and Startz, 1983; Arcidiacono and Lovenheim, 2016). Cast in this light, affirmative action policies may achieve a distributional goal only at the cost of lower productivity.

In contrast to this reasoning, affirmative action policies can raise economic efficiency when evening out unequal opportunities that are unrelated to potential ability (Becker, 1957; Coate and Loury, 1993; Hsieh et al., 2019). When decision makers rely on imperfect information of candidates' potential abilities, differential treatment can be desirable since selection criteria may not account for prior disadvantage. Recent work has found that group-neutral decision rules *de facto* disadvantage underrepresented groups in domains such as hiring, health, and the criminal justice system (Li et al., 2020; Obermeyer et al., 2019; Rose, 2021; Bohren et al., 2022). The presence of such selection imperfections raises the prospect that affirmative action policies may actually be output-enhancing.

In this paper, we study output under a quota at university admissions that changed the gender composition of an entire occupation. We document that this gender quota – despite the fact that it "lowered the bar" for candidates of the underrepresented group – led to a more efficient allocation of study slots by filling them with eventually more productive workers. Our setting is university admissions for primary school teacher studies, one of the most popular fields of study in Finland. Specifically, we analyze a quota that reserved 40% of study slots for men, thus advantaging academically lower-scoring male candidates. We document that these "male quota teachers" perform better relative to marginal female candidates: Pupils who are exposed to more male quota teachers during primary school experience gains in both educational attainment and subsequent labor force attachment. Several pieces of evidence highlight that these productivity gains materialized due to the selection criteria insufficiently accounting for male applicants' skills and intrinsic motivation in absence the quota.

Our identification strategy isolates exogenous variation in pupils' exposure to male quota teachers with a differences-in-differences instrumental variables (DiD-IV) framework that exploits the sudden termination of the quota. This policy change instantly reduced the share of men among admits to primary school teacher studies from about 40% to 20% (Uusiautti and Määttä, 2013; Räihä, 2010; Izadi, 2021). We instrument for the local teacher gender composition that pupils experience in primary school by using the lifting of the quota together with the timing of local demand shocks for new teachers. These demand shocks arise from local teachers reaching the union-mandated teacher retirement age when turning 60. The first stage employs a DiD specification that estimates the differential impact of teacher retirement between the quota and the post-quota period on the local share of male teachers. Intuitively, municipalities in which teachers turn 60 while the quota is still in place will hire new teachers from a rookie teacher market with quota men, compared to municipalities whose teachers turn 60 just after the quota was abolished.[1] The exclusion restriction requires that teacher retirements in the post-quota period do not *differentially* impact pupil outcomes except via changing the teacher gender composition. Our empirical strategy addresses this by comparing pupils who experience similar exposure to new teachers via retirements, but face a different gender composition of those rookie teachers due to the lifting of the quota.

We start by outlining a general conceptual framework of university admissions that derives conditions under which a representation quota results in lower or higher total ability of admitted candidates. When the selection criterion fully reflects candidate ability, introducing a binding quota comes at the cost of admitting less qualified candidates. In contrast, when selection relies on a noisy signal of ability, there can be efficiency gains from a quota if admissions criteria insufficiently capture expected ability of underrepresented candidates.

We then examine the efficiency effects of the quota empirically. First, we document how the lifting of the quota affected the local gender composition of teachers at the municipal level: Once the primary teacher cohorts that studied without the quota graduate and enter the market for rookie teachers in 1994, each retiring teacher is 20 percentage points less likely to be replaced with a male teacher relative to the quota period. These changes in the local teacher gender composition are accompanied by small, albeit noisily measured, increases in local teachers' average academic scores – consistent with the notion that male applicants, who are on average lower scoring, are less likely to be admitted to primary teacher studies once the quota is abolished.

We proceed to study how these changes in teacher composition affect pupils, using comprehensive register data from 1988 to 2018 to trace out pupils' education and labor market pathways until age 25. We start by analyzing pupils' application and enrollment behavior when leaving compulsory education three years after finishing primary school. We track pupils' educational trajectory with records from the nationally-organized allocation of education slots, for which pupils can put in up to five preferred institution choices. Using the timing of teacher age-based retirements as an instrument for the local teacher gender composition, we show that pupils exposed to a higher share of male teachers via the quota are more likely to directly apply to continued education. As pupils' applications are more aligned with attainable options, they are more likely to obtain one of their top two choices. These patterns translate into higher enrollment rates in post-compulsory education at age 16.

Turning to long-term impacts up to early adulthood, we examine pupils' educational attainment and labor market attachment by age 25. For pupils who were exposed to a higher share

[1] We label as "quota men" those male teachers who were only able to enter primary teachers studies because the quota was in place and would not have gotten admitted otherwise. Throughout the paper, we refer to teachers turning 60 as "retirement."

of male quota teachers, we observe a shift towards higher qualifications throughout the educational attainment distribution: For practically-oriented vocational degrees, pupils are more likely to have additional advanced qualifications instead of a basic three-year degree. For academic tracks, pupils are more likely to have obtained a university level BA degree. Consistent with acquiring more education, these pupils have higher attachment to the labor market. At age 25, they are 3 percentage points more likely to be a student or employed for a 1 SD increase in the share of male quota teachers, which corresponds to a 4% increase over the mean.[2]

As the quota exogenously changed the gender composition of an entire occupation[3], we are able to explore heterogeneous impacts by pupil gender along two dimensions that are often hypothesized to be affected from ensuring more diversity in teaching occupations specifically and in gender-segregated occupations more generally.

First, we ask whether male quota teachers had a positive impact particularly on boys. Male quota teachers could raise boys' academic aspirations and achievement by providing a same-gender role model. However, adding more male role models for boys implies that girls lose female examples to aspire to, potentially making them worse off. Examining results by pupil gender, we can rule out that girls are negatively impacted by the quota policy. Further, we do not find evidence for male teachers setting an example for boys purely via same-gender identity: Boys' educational outcomes are not more affected from exposure to male quota teachers relative to girls', and none of the other main effects differ systematically by pupil gender.

Second, we explore whether the quota mitigated the gender-segregation of occupations in the long term: Did quota men inspire boys to study education related fields, which are historically feminized? Using data from degree registers, we track pupils' field of study until age 25. Our estimates indicate that exposure to more male quota teachers makes pupils of both genders more likely to study a STEM (Science, Technology, Engineering, and Mathematics) field. However, neither boys nor girls are more likely to pick an education- or teaching-related field when exposed to a higher share of male quota teachers.

Finally, we distinguish between two main avenues through which more equal gender representation may have improved pupils' outcomes. Teacher team performance could have been higher because the quota mended selection imperfections and admitted higher performing teachers, or because diverse teacher teams were more productive due to complementarities.

We document limited scope for complementarities in production between male and female teachers. In the presence of complementarities, the marginal impact of an additional male teacher should be higher in places with a lower share of male teachers at baseline. We show that the benefits of having more male teachers via the quota are similar in magnitude between places with few male teachers and places where the share of men among colleagues is already high. This suggests limited scope for diversity *in itself* being sufficient to augment productivity.

Instead, three complementary pieces of evidence emphasize the presence of imperfect selec-

[2] 1 SD in the share of male (quota) teachers corresponds to 6.5 percentage points. The average within-municipality change in the share of male teachers over our study period is 5 percentage points.

[3] The number of available study slots for primary teacher education is determined by teacher demand projections from the government (Nissinen and Välijärvi, 2011). University admissions is thus the relevant bottleneck that determines the composition of this occupation.

tion in the unconstrained admission process and shed light on why men's lower performance on the selection criteria did not map into lower performance on the job. First, we show that the matriculation exam score as a main admission criterion emphasizes language fields, in which men perform relatively worse, with a weight of 75%, while attaching a weight of just 25% to math and science fields, in which men perform relatively better. Second, we document that teachers' performance in the matriculation exam, and in particular in language fields, is not predictive of teacher team performance. We estimate a precise zero, suggesting that the weights of this admission criterion inefficiently disadvantaged men. Third, we analyze compensating differentials to show that men who want to become primary teachers have skills that are valued in the labor market outside of teaching. Among similarly qualified applicants to primary school teacher studies, men face a wage penalty when becoming a primary teacher, whereas women obtain a wage premium. These compensating differentials also highlight men's intrinsic motivation in the teaching profession, as they forgo higher earnings when pursuing a stereotypically female occupation. Taken together, our results highlight that men's lower performance on selection criteria do not map into performance gaps on the job, such that the quota raised both representation and efficiency due to mending imperfections in recognizing talents of individual applicants.

This paper makes three main contributions. To the best of our knowledge, we are the first study to cleanly document that a quota can have positive effects that extend *beyond* its direct beneficiaries, such that the policy improved output in the relevant sector in the long run. We relate to recent work that has documented benefits from access to selective colleges for candidates admitted under affirmative action relative to their outside option (Black et al., 2020; Bleemer, 2021; Otero et al., 2021). Empirical evidence on how quotas impact output-related measures has almost exclusively focused on mandated representation of women in board rooms, documenting negative or neutral effects on firm performance in the short-run (Ahern and Dittmar, 2012; Matsa and Miller, 2013; Eckbo et al., 2021; Ferrari et al., 2021).[4]

Our second contribution relates to gender-based role models and teacher value-added. A large body of work has documented that teacher value-added, rather than teachers' certification or test scores (see Jackson et al. (2014) for a review), drive variation in academic and long run economic outcomes for pupils (Rivkin et al., 2005; Chetty et al., 2014).[5] Several studies have provided evidence that being matched with a same-identity teacher can affect academic performance and choice of field of study (Gershenson et al., 2022; Dee, 2007; Lim and Meer, 2017, 2020; Carrell et al., 2010; Kofoed et al., 2019), but others fail to find positive impacts (Holmlund and Sund, 2008; Antecol et al., 2015). We expand on prior work, which has estimated effects from reallocating a fixed set of teachers, by studying whether same-identity role model channels are at play when

[4] Peck (2017) and Cortés et al. (2021) document lower exports and higher firm exit as a consequence of a policy that required firms to hire native workers in Saudi Arabia. Several papers have studied quotas for female politicians, but do not take a stance on whether this impacts output (Chattopadhyay and Duflo, 2004; Beaman et al., 2009; Besley et al., 2017; Baltrunaite et al., 2014; Bagues and Campa, 2021). Chattopadhyay and Duflo (2004) specifically highlight that it is through the characteristic of being female herself that a political leader's preferences in India are more closely aligned with female constituents, thus moving the status quo of policies to more closely reflect preferences of the median voter.

[5] Recent work has emphasized dimensions of teacher value-added that are unrelated to pupils' test scores but matter for long-term outcomes, such as drop-out, educational attainment and crime (Jackson, 2018; Beuermann et al., 2023, Petek and Pope, forthcoming; Rose et al., 2022).

changing the actual composition of the teacher body. Our results highlight that recruiting more men as teachers may have limited potential to mitigate gendered academic achievement gaps in primary education.[6]

Third, our work underscores the importance of imperfect selection criteria in creating disparate impacts for underrepresented groups (Bohren et al., 2022). Attention has recently turned to selection criteria in the context of firms' hiring processes (Li et al., 2020; Chalfin et al., 2016), health (Obermeyer et al., 2019), and the criminal justice system (Rose, 2021; Arnold et al., 2020). We highlight a general point that extends to settings beyond our specific case study: The selection of candidates based on academic scores and interviews, among the most widely used methods to assess applicants, can miss out on important dimensions of minority talent. Adherence to group-neutral evaluation criteria that discount the talents of the minority group can thus create the false impression of selection on merit (Sethi and Somanathan, forthcoming), to the detriment of equal representation and aggregate productivity.

While this paper empirically studies the case of a specific quota policy, its features embody inherent trade-offs that are present in any context in which equal representation targets are deliberated as supply outweighs available positions. The "unusual" feature of our setting, namely the quota being in favor of a group (i.e. men) that did not suffer from widespread discrimination or stigma, is valuable for isolating the importance of these forces in absence of confounders.[7] The conceptual framework delineates the key variables to consider when applying the insights of this study to any other context in which more equal representation is a policy goal. When an underrepresented groups' lower performance on selection criteria is not mirrored in lower performance on the desired outcome a decision maker aims to maximize, addressing the mismatch between selection criteria and this outcome can be beneficial both for equity and efficiency. In essence, a promising path to more equitable representation lies in carefully considering how different evaluation criteria impact minority representation and actual performance. In academia, with current experimentation on test requirements for US college applications, as well as in the private sector, where companies are starting to use balanced candidate lists across groups, such avenues are increasingly being explored.

The paper is structured as follows: The next section details the Finnish education and teacher training system. We outline a brief conceptual framework in Section 3. Section 4 explains our data sources and sample, followed by the empirical design in Section 5. In Section 6, we first examine the effects of the quota on teacher gender composition at the municipal level, before turning to a pupil panel. This is followed by examining long-term role model impacts of the policy in Section 7. We turn to mechanisms in Section 8 and robustness checks in Section 9; the final section concludes.

[6] This is perhaps because studies that show the existence of role model effects examine settings where the potential beneficiaries are underrepresented (e.g. girls in STEM fields), which is typically not the case for boy pupils in compulsory schooling.

[7] In the presence of factors such as stigma against the underrepresented group, one would likely *underestimate* the efficiency effects of more equal representation, because that group would likely face additional discriminatory hurdles in the labor market. It would then be unclear if muted impacts of more equal representation were due to lower candidate quality or other confounders. Because of its long historical tradition, the male teacher quota had broad acceptance in Finland (Mankki et al., 2020).

2 Context

2.1 Primary School Teachers in Finland

Finland has been among the top scoring countries for multiple rounds of international student assessments, leading to considerable international attention paired with efforts to adopt best practices from the Finnish education and teacher training system (Malinen et al., 2012; Niemi et al., 2016).

Due to being one of the most competitive degrees in university admissions, primary school teachers enjoy high social status (Finnish National Agency for Education, 2018). While salaries are on par with the OECD average, active teaching hours are comparatively low (Sahlberg, 2021). Primary school teachers are municipal employees who are hired by local schools, and are part of a powerful teachers' union that fixes both salary schedules, and – for the relevant period in this study – a retirement age of 60 in collective bargaining agreements (Kivinen and Rinne, 1994; Valtiokonttori, 1988). A national curriculum outlines broad learning goals. Under the supervision of municipal education authorities, teachers within and across schools collaborate in designing detailed learning plans (Sahlberg, 2021; Sahlberg et al., 2019).

In contrast to the United States, primary school teachers are assigned to a cohort as their main classroom teacher covering all subjects in the respective grade, and may spend several years with that class. However, primary school teachers are also actively embedded in their work environment through extensive collaboration with their colleagues, both in curriculum design, preparing lessons and school wide activities, as well as in active teaching (Sahlberg, 2021). Pupils in our setting are thus exposed to and interact regularly with the teacher body of their entire school. Conducting the analysis at the municipal level takes into account any spill-overs that arise from teachers collaborating within and across schools.[8]

2.2 Primary School Teacher Training and the Quota Reform

2.2.1 Historical Context

The first teacher training institutes in Finland were founded in the mid-1800s, and offered training separately by gender. In 1881, new education decrees allowed for co-education for children attending municipal primary schools as long as sufficient instruction in handicrafts could be guaranteed, de facto leading to "differentiation between male and female elementary school teachers and a quota system in teacher training" with men constituting about 40% of primary school teachers in the first half of the twentieth century (Sysiharju, 1987).

In the context of educational reforms in the 1970s, primary teacher education was transferred to universities and elevated to a master's level degree (Niemi et al., 2016). With an acceptance rate fluctuating around 10%, primary school teaching has been and still is among the most competitive degrees in the country and applicants often apply multiple years in a row until they are successfully admitted (Tirri, 2014; Uusiautti and Määttä, 2013).

Admissions throughout our study period closely followed the main principles established in

[8]The median population size of the 461 municipalities in our sample is 5000 inhabitants. See also Section 4.

those reforms, including that "the Ministry of Education maintained the sex quota system for the training of classroom teachers" (Sysiharju, 1987): In a first step, applicants were ranked in a centralized system according their score in the matriculation exam (the nationally graded high school exit exam), with a few additional points given for candidates' extra-curricular activities. The highest ranked candidates were then invited to an in-person second round, in which a faculty board evaluated them on essay tasks, exercises, and interviews, with the exact procedure varying across departments (Izadi, 2021; Räihä, 2010; Uusiautti, Määttä, et al., 2013). The highest ranked candidates in the second step were admitted to study primary teacher education according to the number of available study slots.

During the quota period, the Ministry of Education jointly with the education departments ensured that around 40% of candidates invited to the second round were men (Liimatainen, 2002; Mankki and Räihä, 2022). Documentary evidence suggests that universities followed the Ministry of Education's requested gender mix also in the decentralized second step of the selection process by ranking candidates within their specific gender, and allocating 40% of final slots to men (Sysiharju, 1987; Liimatainen, 2002).[9] The quota was abolished for the cohort applying to university in the fall of 1989 (thus graduating from primary school teacher studies in 1994), as it was not in compliance with a broad anti-discrimination law passed by parliament in 1987 (Tasa-Arvovaltuutetu, 1987). Since its lifting, politicians and the general public have repeatedly argued for the quota's reinstatement, motivated by the fact that boys are increasingly lagging behind academically and that a growing number of children raised by single mothers may lack a father figure (Etelä Suomen Sanomat, 1988; Liiten, 2012).

2.2.2 Summary Statistics: Admissions and Teachers' Characteristics

Using aggregate statistics issued by the Ministry of Education, Figure 1 displays the share of men among those applying to primary teacher studies, and among those being invited to the second round of the selection process. While there is a sharp drop from 40% to 20% for second round invitees in 1989, the share of men who apply evolves smoothly around the time of the reform.[10]

As the quota did not only change the gender composition of incoming teachers but also advantaged academically lower scoring men, Figure 2 plots future teachers' grade in the matriculation exam for the first attempt of the exam, against the last year in which they ever took this exam.[11] While the quota was in place, men on average scored about .45 grade points, or

[9]Statistics on pass rates and scores for the decentralized second step of the selection process for this time period were not collected. Uusiautti and Määttä (2013) cite a statistic from the University of Lapland in 1978, where about three applicants were invited per available study slot.

[10]While the lifting of the quota was widely discussed in policy and media reports at the time, we have found no documentary evidence that either application numbers (see Appendix Figure A2) or the composition of applicants would have drastically changed with the lifting of the quota. Figure 2 shows no unexpected discontinuity in the matriculation exam scores of those admitted post-quota.

[11]The final matriculation exam score is averaged across the grades obtained in four mandatory fields of the exam, discussed in detail in Section 8.2.1. Grades are assigned on a curve and range from 2-6, with 6 being the highest grade and 0 assigned for a failing exam (Kupiainen et al., 2018). We plot the grade obtained in the *first* attempt at the exam in order to get at a measure of cognitive ability that is not influenced by repeated test taking. We plot the exam grade against the date of someone's *last* exam take to most closely approximate the point of entry to university studies.

a bit more than half a standard deviation (.8) lower. Once the quota was lifted, the score gap narrowed to about .33 grade points for the cohorts displayed. We will return to the changes in primary teacher gender composition and academic scores more formally in Section 6.1.

Teacher gender in our setting is correlated with a bundle of other characteristics that may matter for teaching. Table 1 presents summary statistics on male and female teachers who are active in the profession before the lifting of the quota (i.e. before 1994 as the year in which the first non-quota cohort graduates from teacher studies), and thereafter. In Panel A, we can observe that male teachers are somewhat more likely to come from rural areas and to live in their region and municipality of birth when compared to female teachers, but differences are small. Regarding educational trajectories in Panel B, there is no difference in having obtained a high school degree and being a certified teacher.[12] In Panel C, statistics on the matriculation exam show no difference in having passed the exam, but again illustrate that male teachers have significantly lower exam scores, even when considering the best grades obtained across repeated attempts. High school students had some flexibility to choose either mathematics or a combination of other natural and social sciences ("Reaali") in the matriculation exam. Male teachers are about 9 percentage points more likely to have taken the mathematics exam compared to female teachers, and 11 percentage points more likely to have chosen advanced level mathematics rather than the basic level exam.[13]

3 Conceptual Framework

This section develops a conceptual framework assessing the trade-offs between teacher ability and gender representation. In our setting, the admissions office is the decision maker who would like to maximize future teacher ability. However, true teacher ability is unobserved by the admissions office so that it has to select candidates based on scores.[14] The goal of this section is to conceptually highlight *when* different admission rules are costly for output. We differentiate between the two most prominent admission rules highlighted in current affirmative action debates: group-blind admissions vs. affirmative action, which we model as a binding representation target that allows for group-specific admission thresholds. We show that in a world in which scores fully reflect ability, a binding quota rule is costly for output. In a world in which scores are a noisy signal of ability, however, group-blind admission can miss out on talent when gaps in test performance are not mirrored by gaps in underlying ability. We discuss the relationship between group-specific signals and ability using the example of our setting, but its conclusions apply more broadly to any context in which an under-represented groups' lower performance on evaluation

[12]Male teachers are on average a year older when being awarded their teaching degree, likely due to mandatory military service for men. We illustrate how the timing of military service interacts with taking the matriculation exam in Section 8.2.1.

[13]Each exam field and level of difficulty is graded on a curve within that group. We discuss gender differences in exam scores in detail when turning to mechanisms in Section 8.

[14]Without loss of generality, we make a simplification in the model: We consider a one-stage selection process based on scores (that can be any combination of academic score and evaluator score) to illustrate the main forces at play. In Section 8.2.3, we document in detail that post-quota admission probabilities conditional on exam score do not differ by gender. This suggests that focusing on academic scores captures similar dimensions of ability as are evaluated in the second step, since those criteria are not differentially overturning the academic score signal by gender.

criteria may not reflect ability differences of a similar scope.

3.1 Set-Up

Consider an admissions office that seeks to select a fixed mass of candidates c from a pool of applicants.[15] Candidates belong to one of two groups g, with $g \in \{M, F\}$ for Male and Female. The admissions office would like to select teachers with the highest teaching ability a, but it can only observe candidates' scores, s, with group-specific density $h_g(s)$, $s \in [\underline{s}, \bar{s}]$. The mass of candidates of each group that is admitted above a cutoff score s_g^* is given by $m = \int_{s_M^*} h_M(s)ds = 1 - H_M(s_M^*)$ and $f = \int_{s_F^*} h_F(s)ds = 1 - H_F(s_F^*)$.[16] The admissions office solves:

$$\max_{s_M, s_F} \int_{s_M} s\, h_M(s)ds + \int_{s_F} s\, h_F(s)ds \tag{1}$$

$$s.t. \;\; m + f = c$$

In the following, we discuss two iterations of this set-up that vary according to how well scores reflect underlying ability. We first examine a case in which scores fully reflect ability, and then move to analyze a case in which scores are a noisy signal of ability, allowing for differential noise by group. In both iterations, we evaluate total teacher ability relative to unconstrained optimization for two admission rules: i) affirmative action by imposing a group-specific representation goal in the form of a quota, and ii) group-blind admissions based purely on scores. Under all rules, we assume that the admissions office is required to set score thresholds above which every candidate within a specified group is admitted, such that randomization is not allowed. Under affirmative action, the admissions office solves equation (1) with the additional constraint of admitting at least mass q of the M group:

$$s.t. \;\; m \geq q \tag{2}$$

3.2 Case 1: Scores Fully Reflect Ability

We start by evaluating a benchmark case in which scores fully reflect teaching ability:

$$a = s$$

Unconstrained optimization: When maximizing total ability of admits, the optimal cutoff score s^* when solving the admissions office's problem in equation (1) is the same for both groups.

[15] The admissions office's problem here is similar to Chan and Eyster (2003), but we add an explicit distinction between observable scores and unobserved ability. Chan and Eyster (2003) make the theoretical point that forbidding universities that value diversity from using race as a selection criteria may result in lower quality of admits. As noted by Ray and Sethi (2010), the optimal admissions rule under race-blind admissions is generically non-monotone, with lower scoring candidates admitted and higher scoring candidates screened out within each group. Ellison and Pathak (2021) bring this reasoning to the data to evaluate the efficiency of a place-based affirmative action rule in two Chicago Public Schools. Instead of assuming an explicit preference parameter for diversity, their model defines students' outcomes as a trade-off between an optimal level of school diversity and academic match. Our set-up does not assume a taste parameter for diversity.

[16] Upper limit of integral ($\bar{s}$) suppressed here and in the following for better readability.

i) Affirmative action: A binding quota will require the admissions office to deviate from the optimal cutoff by allowing the M group to enter with lower scores:

$$s_M^*(q) = s_F^*(q) - \delta \tag{3}$$

where δ is the Lagrange multiplier of the quota constraint and indicates the shadow price of admitting a male candidate at the margin. This case illustrates one of the main concerns frequently raised against affirmative action: Requiring the admissions office to forgo its preferred allocation of slots will lower total candidate quality, as the ability of the additional men admitted under the quota is lower than the ability of women who must be rejected to satisfy it:

$$\int_{s_M^*(q)}^{s^*} s\, h_M(s)ds < \int_{s^*}^{s_F^*(q)} s\, h_F(s)ds$$

ii) Group-blind: Under group-blind admissions based purely on scores, the admissions office sets the same cutoff score s^* for both groups, with s^* a function of score densities and capacity. This corresponds to the unconstrained solution.

3.3 Case 2: Group-Specific Noise

We next consider a world in which scores are a noisy signal of true ability a. In particular:[17]

$$s_g = a_g + e_g \tag{4}$$

$$\textit{where}\;\; a_g \sim \mathcal{N}(\mu_{a_g}, \sigma_a^2) \;\; \textit{and} \;\; e_g \sim \mathcal{N}(\mu_{e_g}, \sigma_e^2)$$

with a group-specific error, e_g, that is independent of a_g. The admissions office now maximizes total *expected* ability given the scores it observes:

$$\max_{s_M, s_F} \int_{s_M} E(a_M \mid s_M)\, h_M(s)ds + \int_{s_F} E(a_F \mid s_F)\, h_F(s)ds \tag{5}$$

$$s.t.\;\; m + f = c$$

Unconstrained optimization: The optimal cut-off scores for the M and F group take into account both variance-weighted mean ability differences between groups, as well as mean differences in noise:

$$s_M^* = s_F^* + \frac{\sigma_e^2}{\sigma_a^2}(\mu_{a_F} - \mu_{a_M}) - (\mu_{e_F} - \mu_{e_M}) \tag{6}$$

[17]The assumption that the variance of ability and noise do not differ by group is based on the matriculation exam score distributions between men and women exhibiting a mean shift, but similar variance in our setting (see also Section 8.2.1). This is similar to group differences in test taking in other contexts: In the General Aptitude Test Battery (GATB), a job placement test used for decades by the US Employment Service, minorities' score distribution exhibits a mean shift, but similar variance (Hartigan and Wigdor, 1989). In contrast, Phelps (1972) assumes actual differences in mean ability by racial group that arise from pre-market disadvantages.

Because scores are a differentially noisy signal of ability by group, optimal scores are group-specific and correct for group differences in test taking that are not mirrored in ability gaps. In our setting, an unconstrained admissions office would like to set a lower bar for the M group if the mean gap in noise $(\mu_{e_F} - \mu_{e_F})$ outweighs the variance-weighted mean ability gap $(\mu_{a_F} - \mu_{a_M})$.[18] Total expected teacher ability is maximized when men's cutoff score is lower: $s^*_M = s^*_F - \kappa$, with $\kappa = (\mu_{e_F} - \mu_{e_M}) - \frac{\sigma^2_e}{\sigma^2_a}(\mu_{a_F} - \mu_{a_M})$.

i) Affirmative action: The admissions office's unconstrained solution already takes into account differential test performance. This makes the quota less likely to bind, since the admissions threshold is already being relaxed for group M. When a quota rule is binding, it adds an additional test score wedge to the unconstrained maximization: $s^*_M(q) = s^*_F(q) - \kappa - \delta$. Total expected teacher ability declines as the wedge between marginal male and female admission scores grows beyond κ with the mass of male candidates required by the quota.

ii) Group-blind: When the admissions office is required to set equal cutoff scores at s^*, it is forced to ignore men's score penalty on the test. In this case, it is the lack of differential score thresholds that is costly, as lower-scoring, high expected ability men are replaced with higher-scoring, but lower expected ability women:

$$\int_{s^*_F - \kappa}^{s^*} E(a_M \mid s_M)\, h_M(s) ds > \int_{s^*}^{s^*_F} E(a_F \mid s_F)\, h_F(s) ds$$

Taking stock: The framework highlights two main take-aways. When scores fully reflect underlying ability, binding representation targets come at a quality cost. This is a strong assumption, as in practice admissions committees must rely on noisy signals of unobserved ability. When such signals disproportionally discount an underrepresented group's ability, forcing admissions to set equal thresholds across groups is both costly in aggregate and detrimental for equal representation.[19]

In our setting, admission moves from a binding quota that thus allows for differential score cutoffs between men and women to a setting where admissions are constrained to use the same score for any applicant post-quota.

4 Data and Sample

Our main data source is register data maintained by Statistics Finland which span the years 1988 - 2018, and contain detailed yearly information on all residents in Finland. We compile two main data sets that correspond to the respective parts of the analysis.

Teachers: We construct a panel of active primary school teachers from 1990 - 2000 for all

[18] Men perform worse at the matriculation exam when considering the full population of test takers, discussed in detail in Section 8.2.1.

[19] This insight generalizes to other settings in which under-represented groups' worse test performance is documented (or at least assumed) to not mirror equally sized ability differences. In their evaluation of the General Aptitude Test Battery (GATB), a job placement test that had been in use for decades by the US Employment Services, Hartigan and Wigdor (1989) write "there is not so great a difference in average job performance between minority and majority applicants as there is in average test performance."

individuals whose occupation at any point in time between 1990 - 2005 is classified as a primary school teacher by Statistics Finland's occupation classification system in the employment register. Since occupation categories are first available in 1990 and are not reported in every year, we use a combination of workplace, industry, salary, degree and career information to infer active teacher status in any given year [data sets referenced in brackets: FOLK employment, basic, and degree]. We can match teachers' matriculation exams scores and exam dates for all cohorts born after 1952 [YTL moduuli], but we do not observe university applications in this sample as these registers were not maintained at the time.

In order to examine future teachers' application success and outside options, we thus rely on a sample of adjacent cohorts. We use data on the universe of all applicants to primary teacher studies for the years 2000 - 2005, with 2000 as the first year in which data on field of application was collected in the centralized university application system [EDU-HAREK]. We link this sample of primary teacher applicants to their performance in the matriculation exam [YTL moduuli], as well as their degree attainment and earnings in the year 2020 [FOLK degree and basic]. We use this sample to shed light on compensating differentials in Section 8.2.3.

Pupils: We observe the universe of children living in Finland who turn seven years old (and therefore start school in that calendar year) between 1988-2000, reaching age 25 until 2018 as the last year of our data in which all outcomes of interest are available. We assign children to a municipality (and teacher gender composition during grades 1-6) based on their place of residence in the year in which they start school. We further match pupils to their parents which allows us to observe a rich host of variables related to families' socio-demographic characteristics at age seven [FOLK family]. We use a variety of registers, available on a yearly basis after age 16, to measure pupils' outcomes:

Intermediate outcomes: We merge pupils to registers on post-compulsory education applications that occur in the last year of middle school, i.e. the year in which pupils turn 16 [EDU-THYR]. This allows us to observe when pupils apply, their preference ranking of up to five degree and institution choices, as well as which option they are allocated in the centralized admissions process. For the school starting cohorts from 1990 on-wards, we can additionally observe enrollment in post-compulsory education [EDU-OPISK].

Early adulthood: We measure pupils' labor force status as recorded in the last week of the calendar year in which they turn 25 years old [FOLK employment]. Regarding educational outcomes, we observe pupils' highest degree achieved, and we construct their field of education using information on their latest degree [FOLK degree]. We also examine fertility patterns up to early adulthood with yearly data from the population register [FOLK basic].

We measure all of the treatment variables at the municipal level since data to link pupils and teachers to classrooms or schools do not exist. As our main goal is to estimate the impact of a quota per se, and not the impact of having a teacher of a particular gender, aggregating the data to a level higher than the classroom is consistent with both the research question and a setting in which collegial collaboration is widely practiced. The median population size among the 461 municipalities in 1990 is 5061 inhabitants, with 438 children and about 22 primary school teachers across the six cohorts attending primary school.

To comply with data disclosure regulations by Statistics Finland, we exclude municipalities that contain fewer than three teacher observations in a given year from our analysis. Once we move to a pupil level panel, we restrict the sample to municipality*year cells for which we are able to observe at least six teacher observations (i.e. the teaching staff for grades 1-6).[20]

5 Empirical Strategy

We want to study whether and how output is affected when the gender composition of teachers changes via a quota. Lifting the quota at the point of university admissions will impact the gender composition among active primary school teachers only gradually over time, but the changes in the flow of incoming teachers are sharp and immediate. In the estimation strategy, we therefore use shocks to the demand for new teachers that arise from idiosyncratic local teacher retirement. Since teacher retirement could respond endogenously to the policy reform itself, we only use variation from teachers reaching the union-mandated retirement age of 60. We use the term "retirement" exclusively to refer to teachers turning 60 throughout the paper.

An ideal experiment given the aggregation level of our data would consist in randomly removing some teachers from municipalities, and deciding with a coin flip whether replacement teachers are drawn either from a pool of male quota teachers, or from a pool of marginal female teachers.[21] Our DiD-IV estimation strategy closely approximates this experiment, taking into account that changes in quota teachers materialize via the inflow of rookie teachers and that we cannot directly observe quota male and marginal female teachers in the data. Municipalities in our setting are randomly assigned more quota men – and thus more male teachers in general – via the *timing* of their open positions arising from teacher retirement. We thus estimate a local average treatment effect for complier municipalities, exploiting variation from those municipalities that are induced to hire more vs less quota men among their teachers via the timing of their retirements. This gives the policy-relevant parameter of interest: What happens when we change the composition of an occupation via a quota that operates through the inflow of incoming candidates?

Figure A1 outlines the timeline of our reform: The primary school teacher students who enter university before 1989 are selected via the quota rule. As the time to complete the degree is five years (Nissinen and Välijärvi, 2011), the quota and non-quota cohorts of new teachers will leave university around the year 1994 and will be hired by municipalities for their local schools. If municipalities have open positions during the time when quota cohorts enter the teacher market, they will be more likely to hire candidates from a pool with relatively more male rookie teachers compared to municipalities that have to fill open positions once new teacher cohorts selected without the quota are entering the teacher market.

[20] Results are qualitatively similar, but more noisily estimated, when keeping the 7,154 pupils for which we have incomplete teacher composition information in the sample.

[21] We label as "male quota teachers" those male teachers who were only able to enter primary teachers studies because the quota was in place and would not have gotten admitted otherwise. We refer to "marginal female teachers" as those female teachers who were able to be admitted to primary teacher studies once the quota was abolished and would not have gotten in if the quota were still in place.

5.1 Municipal Level: Changes in Teacher Composition

We first document the first stage relationship. Local retirement interacted with the timing of abolishing the quota changes the local gender composition of teachers. Consider the following specification:

$$\textit{share male}_{mt} = \pi_0 + \pi_1 \, \textit{total share 60}_{mt} + \pi_2 \, \mathbb{1}_{t=post} \, \textit{total share 60}_{mt} + X_{mt}\delta + \eta_{rt} + \gamma_{mp} + \zeta_{mt} \quad (7)$$

with $\textit{share male}_{mt}$ the share of male teachers in municipality m in a given year t, and $\textit{total share 60}_{mt}$ the cumulative teacher retirements in a municipality up to that point in time in the sample.[22] The indicator function $\mathbb{1}_{t=post}$ switches on once non-quota teacher cohorts graduate and start entering the teacher market in 1994. The coefficient of interest, π_2, measures how additional retirements in the post-quota period affect the share of local male teachers *relative* to when the quota was still in place. We add region-by-year fixed effects η_{rt} to control for time-varying shocks whose impacts may vary regionally, with a total of 19 regions comprising on average 24 municipalities. We can also include controls for time-varying municipal characteristics X_{mt}. The municipality-by-period fixed effects, γ_{mp}, 'reset' the measure of total retirements once the post period starts to separately estimate how retirements affect the local share of male teachers in the post period.[23]

5.2 Pupil Level: Does the Quota Shift in Teacher Gender Affect Outcomes?

Structural equation: We are interested in how increasing the share of male teachers via the quota affects pupils' outcomes. To isolate variation in the share of quota men that pupils face, we use the timing of local retirement shock as an instrument for the average share of male teachers in the following structural equation:

$$y_{im,t+x} = \beta_0 + \beta_1 \, \overline{\textit{share male}}_{mt} + \beta_2 \, \overline{\textit{total share 60}}_{mt} + X_i\delta + \gamma_m + \eta_{rt} + u_{imt} \quad (9)$$

with $y_{im,t+x}$ the outcome of interest at time $t+x$ for pupil i who at age seven lived in municipality m, and X_i individual level controls for socio-economic status, also measured at age seven.[24] We add municipal fixed effects γ_m, as well as region-by-cohort fixed effects, η_{rt}.

In the two-stage least-squares set-up, $\hat{\beta}_1$ is the coefficient of interest and measures how increasing the average share of male teachers via the quota affects pupils' outcomes, with $\overline{\textit{share male}}_{mt}$

[22] The fixed effects specification of equation 7 uses the stock of the dependent variable (the share of male teachers) and the independent variable (the cumulative share retiring teachers over time). The corresponding first difference equation uses flows on both sides of the equation by regressing the year-on-year changes in the share of male teachers within a municipality on the share retiring teachers in each year, dropping the municipal fixed effects:

$$\Delta \textit{share male}_{mt} = \pi_0 + \pi_1 \, \textit{share 60}_{mt} + \pi_2 \, \mathbb{1}_{t=post} \, \textit{share 60}_{mt} + X_{mt}\delta + \eta_{rt} + \zeta_{mt} \quad (8)$$

We report first stage results for both equations, and use equation 8 when thinking in flows is more intuitive for robustness checks on hiring patterns.

[23] The reset is necessary so as to properly net out any effect of the quota-period retirements from the post-quota estimate. I.e. the effect of retirements on the gender composition in the post-quota period is independent of how much retirement the municipality faced in the quota period.

[24] The controls we include are pupil gender, language (Swedish, Finnish, other), foreign origin, single parent household, and highest level of education in the household (Compulsory, Secondary, Tertiary, n/a).

the average of the share of male teachers across the years we observe pupils in primary school.[25]

Our empirical strategy isolates variation in the share of male quota teachers from gender changes in the inflow of recently graduated teachers that is caused by retirements. Rookie teachers may differ from older teachers along various dimensions: they have less experience, but they might also be differentially motivated to teach. We account for pupils' exposure to rookie teachers via retirement by controlling for the average aggregate share of teacher retirements during a pupils' time in primary school, $\overline{\textit{total share 60}}_{mt}$, and we discuss its construction in more detail below. Schools' hiring decisions, and thus the impact of being exposed to retirements, could change due to the quota. However, we do not find evidence that schools changed their hiring practices as a response to the reform in Section 9.

First stage: We instrument for $\overline{\textit{share male}}_{mt}$ with the following first stage equation on the pupil level that closely mimics the municipal level first stage in equation 7. Since every time period t corresponds to the start of school for a particular cohort, we refer to t as a cohort identifier in the following:

$$\overline{\textit{share male}}_{mt} = \pi_0 + \pi_1 \overline{\textit{total share 60}}_{mt} + \pi_2 \mathbb{1}_{t=post} \overline{\textit{total share 60}}_{mt} + X_i\delta + \gamma_m + \eta_{rt} + \epsilon_{imt} \quad (10)$$

Variation in treatment intensity arises from how much teacher retirement different cohorts of pupils across different municipalities experience in the post-quota relative to the quota period. π_2 measures how the share of male (quota) teachers a pupil experiences is affected by retirements in the post-quota relative to the quota period. By measuring the differential impact of retirements, we compare the causal effect of being exposed to new teachers against the causal effect of being exposed to new teachers with a changed gender composition due to the lifting of the quota.

In the structural equation, β_1 as the main coefficient of interest thus estimates the causal effect of being exposed to relatively more male teachers via incoming quota men as opposed to marginal female teachers.

Exposure to retirement: We measure pupils' exposure to retirements during their time in primary school $\overline{\textit{total share 60}}_{mt}$, by taking the average of cumulative retirements a pupil is exposed to during their six grade levels g in primary school:

$$\overline{\textit{total share 60}}_{mt} = \frac{1}{6}\sum_{g=1}^{6} R_{mtg} \quad (11)$$

with $R_g = \textit{share 60}_g + R_{g-1}$ and $R_1 = \textit{share 60}_{-2} + \textit{share 60}_{-1} + \textit{share 60}_1$.[26]

Each retirement is thus weighted by the number of years it affects the teacher gender composition a pupil experiences in school: R_1 is the sum of retirements that occur just before a pupil enters the first grade and the two years before that. R_2 adds retirements that occur before a pupil enters second grade to the sum of retirements experienced up to that point: $R_2 = \textit{share 60}_2 + R_1$, and so

[25] Our pupil panel spans 13 cohorts that are starting school in the years 1988-2000, and thus experience teachers who we can observe from 1990 - 2000. For some cohorts of pupils, we observe the teacher composition for each year that pupils are in school, while for others, we only know it for their starting or ending years. Appendix Figure A23 depicts the cohorts over time observed in our data.

[26] Subscript mt omitted for better readability in this definition (i.e. $R_{mtg} = R_g$) and in the following paragraph.

forth up to grade 6. We construct $\overline{\textit{total share 60}}_{mt}$ in this way to reflect the fact that retirements that happen later in the pupils' school career will have an impact on the teacher composition for relatively fewer years compared to retirements when pupils start school. In the empirical analysis, we report grade level results for the first stage that directly motivate the construction of this measure.

5.3 Discussion of Identifying Assumptions

We revisit explicit and implicit identifying assumptions of our setting in more depth. To start with, our identification strategy needs to satisfy the two main IV assumptions. Relevance requires that the relative size of teacher retirements in the post-quota period decisively impacts the local share of male teachers, which we can assess directly in the first stage regressions. The exclusion restriction, briefly touched on above, warrants more discussion: We require that teacher retirements affect pupils' outcomes only via changes in the share of male teachers, and thus changes in male quota teachers. However, retirements themselves, by triggering teacher turnover, may have a direct effect on pupils. We tackle this by measuring relative changes in outcomes between cohorts that experience similar exposure to retirements, but with different timing. The underlying assumptions here are twofold: First, we need to assume that there are no other policy changes that happen simultaneously with the quota that have effects on students *via the channel of retirements*. To the best of our knowledge, there are no such policies. Secondly, we assume that exiting patterns and hiring practices to replace retiring teachers do not differentially change as a response to the quota.[27] We test for such patterns in Section 9 and do not find evidence for differential changes in the post-quota period.

Implicit in our empirical design is the further assumption that the local timing of retirements is idiosyncratic, and therefore uncorrelated with any other shocks that could affect pupil outcomes. We address such concerns by only using variation arising from teachers turning 60 (instead of actual exits), by controlling for a rich host of pupils' socio-economic characteristics at age 7, and by including region-by-cohort fixed effects. As such, we are only comparing cohorts in municipalities within the same region and year, with the notion that relevant economic shocks (in the past and currently) will similarly affect neighboring places.

Finally, while our regressions are measuring the effect of having more male quota teachers, we see teacher gender not just as a biological distinction, but as something that proxies for a bundle of characteristics that may differentiate quota male and marginal female teachers.

6 Main results

6.1 Municipal Level: Effects on Teacher Composition

Teacher gender: We start by documenting the effects on teacher gender composition at the municipal level after the quota was lifted. We first examine teacher exit patterns. Figure 3

[27] This includes the monotonicity assumption that rules out defiers in a LATE framework. In our case, these would be municipalities that would not want to hire male teachers while the quota is in place when facing retirements, but start hiring differentially more male teachers for retiring teachers in the post period.

plots the exit probability by age for all primary school teachers in our sample. We report the probability of a primary teacher not teaching at a given age, conditional on having been an active teacher in the previous year. There is a large spike in exits exactly at the union-mandated retirement age of 60. In our estimation, we are only using variation from teacher exits that is due to teachers turning 60 years old.[28]

We start by illustrating the intuition of the first stage using raw data. Figure 4 displays the relationship between teacher retirement in a municipality (on the horizontal axis) and changes in the share of male teachers by separately plotting the period in which quota cohorts enter the teacher market (1991-93) and a period of similar length in the post-quota period (1994-96). Teacher retirement has a small, positive effect on the local share of male teachers in the quota years. In the post-quota period, higher shares of teachers retiring are associated with substantial local drops in the share of male teachers.[29]

Figure 5 formalizes this intuition by running the first stage Equation 7 as an event study, estimating separate coefficients year-by-year, relative to 1993 as the last quota-period year. Teacher retirements in the years in which the quota was still in place do not differentially affect the local share of male teachers relative to the year 1993, while retirements in the post-quota period lead to a sizeable drop of about 20 percentage points. Table 2 summarizes this result for both the first difference and fixed effects specifications, estimating separate coefficients for the quota and post-quota period. Results are quantitatively similar across specifications: While retirements in the pre-period have a small positive effect on the local share of male teachers, the coefficients of interest on retirements in the post-quota period are consistently negative. We document robustness to negative weights arising in two-way fixed effects estimation in the presence of heterogeneous treatment effects in Section 9 following De Chaisemartin and d'Haultfoeuille (2020).

The magnitude of reported coefficients corresponds to the new steady state and measures the reduction in the share of male teachers if (eventually) all teachers in a municipality retire post-quota: In this scenario, the local share of male teachers would drop by between 16-20 percentage points. These magnitudes closely match the drop in incoming male teachers reported by the literature and observed in teacher admissions (Figure 1). For any given post-quota retirement, each retiring teacher is 20 percentage points less likely to be replaced by a male teacher relative to the quota period.[30]

Teacher academic ability: While the quota targeted the gender composition of incoming primary school teachers, it simultaneously affected overall academic ability among teachers by giving preferential access to men with lower academic scores on average. In Table A1, we report the first stage with the municipal average of teachers' overall matriculation exam grades as the outcome. While coefficients are noisily estimated due to exam grades only being available for teacher cohorts born after 1952, retirements in the post period lead to an increase of about .08

[28] Appendix C reports municipal level statistics on the share of teachers turning 60. In any given year, about 45% of municipalities have any retirement. We also examine teachers' likelihood of changing jobs across municipalities in Appendix Figure A3. Less than 1% of teachers in the age bracket above 55 are changing the location of where they teach across all years of our panel.

[29] Since our teacher panel spans 1990-2000, the first year for which we can calculate the share of teachers turning 60 that determines re-hiring for the upcoming academic year is for 1991 (i.e. the 1991/92 academic school year)

[30] Over the ten years of our municipal panel, the share of male teachers drops by on average 5 percentage points. Close to 25% of teachers within a municipality retire over the same period (see Appendix Figure A6).

grade points in the local teacher body, relative to the quota period (column 1). This magnitude is consistent with replacing approximately 20% of teachers with an on average .45 grade point higher score in the post-quota period (see Figure 2) and corresponds to about a .3 SD increase in local teacher teams' exam grades.[31] We next turn to examine how these changes affect pupils.

6.2 Pupil Level: First Stage

Our pupil-level panel spans the cohorts that enter primary school between the years 1988 to 2000. We start by documenting the first stage relationship: Are children who experience more teacher retirement post-quota exposed to fewer male quota teachers? As we observe pupils at fixed points in time after having completed primary education, we would like to relate pupils' overall exposure measure to male teachers, i.e. the average share of male teachers across the six years a pupil spends in primary school, to their overall exposure to teacher retirements.

We begin by documenting grade-level patterns to trace the dose-response function between exposure to male (quota) teachers and retirements. Figure 6a shows the first stage results if we regress the average share of male teachers on the share of retirements pupils experience just before they start each grade level, starting up to two years before they enter school and until grade six.[32] Figure 6a depicts coefficients separately for the quota period (grey) and the post-quota period (green), while Figure 6b shows the effect of retirements in the post-quota period *relative* to the quota period. Teacher retirements in the early grades have a large and significant impact on the average share of male teachers pupils experience during their time in primary school. At higher grade levels, this effect gradually peters out. This pattern clearly shows that retirements in early grades, which affect the teacher composition during the entire six years a pupil spends in primary school, contribute more to explaining the average share of male teachers a pupil faces across their entire time in primary school. Similarly, retirements that happen just before a pupil enters grade six will only impact the share of male teachers for one year, and therefore contribute less to moving the average share of male teachers over all six years. This pattern, as described in Section 5.2, informs our construction of the instrument when measuring a pupil's exposure across all grades. We define a pupils' exposure to retirements as the average cumulative share of teachers retiring in each grade level, which weighs retirements proportional to the number of grades they impact the teacher composition that a pupil experiences.

In Table 3, the first three columns report results for the pupil level first stage. Due to the cumulative nature of the explanatory variable, we can interpret this coefficient as 'how much does the average share of male (quota) teachers change if all teachers were to retire just before a pupil starts school'. The magnitudes closely match the municipal level regressions and the drop in the share of men in admissions to primary teacher studies (Figure 1).

[31] We use teachers' grade in their first attempt at the exam. We observe grades for 59% of the total teacher sample (as grades are only available for cohorts born after 1952) and restrict the sample to municipalities for which we observe at least one teacher with a score in any year. We repeat the first stage regression with the restricted sample to ensure comparability in Appendix Table A2, with results unchanged.

[32] Appendix Figure A23 shows which cohorts are exposed to quota years in which grade levels.

6.3 Intermediate Outcomes: Applications and Enrollment for Post-Compulsory Education

Turning to outcomes, we start by tracking pupils' application choices to higher education options that take place after compulsory schooling at age 16.[33] After primary school (grades 1-6) and middle school (grades 7-9), pupils in Finland have the option to apply to upper secondary education, which typically takes three years to complete, is provided free of charge, and is divided into vocational and academic tracks. In grade 9, the final year of middle school, pupils apply for their desired institution, and in the case of the vocational track also their desired field. While further education is not mandatory after age 16, raising completion rates of upper secondary education is a policy priority as a post-compulsory degree is deemed crucial for labor force attachment: Finns with only compulsory education have significantly lower employment rates in adulthood and are four times more likely to be out of the labor force altogether (Virtanen, 2016; Niemi et al., 2016).[34] In the centrally-organized application process, each pupil can submit up to five choices for institution (and field), and a student-proposed deferred acceptance (DA) algorithm allocates available study slots.[35] As applications take place before pupils obtain their final grades that are used to allocate slots, and with the popularity of institutions and fields varying over the years, students face uncertainty. The number of available slots per degree is centrally regulated and about 4% of a cohort end up without a study slot in the fall after finishing middle school.

We start by examining the dose-response function of the reduced form: How do retirements affect application decisions? Rather than establishing results for impacts at particular grade levels, the goal of this exercise lies in examining the similarity of dose-response patterns between the first stage and the reduced form. Figure 7 shows the grade-level reduced form for whether pupils apply to post-compulsory education directly in their last year of middle school, with the upper panel reporting separate coefficients for the quota and post-quota period and the lower panel showing the relative difference. As documented in the upper panel, exposure to new teachers via teacher retirements during the quota period has small positive, but insignificant impacts on pupil's likelihood of applying. Post-quota retirements in the earlier grades of pupils' primary school attendance have larger and negative impacts on applications, similar to the patterns observed in the first stage (Figure 6). As factors other than male quota teachers may impact application decisions, the grade-level coefficients in the reduced form are more noisily estimated than the more mechanical relationship in the first stage, with idiosyncrasies present in particular grade levels. Overall, however, the patterns between first stage and reduced form are reassuringly synchronous.

[33] Virtually everyone (99.7% of a cohort) successfully graduates from compulsory education (Virtanen, 2016). See Appendix Figure A24 for more details on the Finnish education system

[34] Prior research with Finnish data has shown that slot allocations in upper secondary education matter for degree completion: With an RDD design, Virtanen (2016) shows that failing to obtain a preferred choice or a study slot at all results in a lower probability of graduation. Huttunen et al. (2023) document that admission to any post-secondary education reduces crime among young men.

[35] For an infinite number of choices, the algorithm would be strategy-proof. Since students can only submit five choices, some may choose to enter a 'safe' option to make sure they get a spot. See Virtanen (2016) for an in-depth description of the allocation process of slots for upper secondary schooling.

Considering pupils' exposure to male (quota) teachers and retirements over their entire time in primary school, Table 3 reports the first stage, reduced form and IV for the main outcome for this section, gradually adding controls. Our preferred specification includes region-by-cohort fixed effects, thus comparing pupil cohorts in close-by municipalities, and we subsequently report results for this specification choice.[36] While teacher retirements that pupils experience during the quota period have a small positive, but insignificant impact on the share of male (quota) teachers (column 3) and their application likelihood (column 6), there is a significant negative impact of retirements in the post-quota period on their exposure to male (quota) teachers and applications to post-compulsory education.

Column 9 reports the corresponding IV estimates. Being exposed to more male teachers via the quota during primary school results in higher likelihood of pupils applying to post-compulsory education. The coefficients report the effect size associated with an increase of male quota teachers from zero to all of the teaching staff being male quota teachers. When scaling effect sizes by a 1 SD increase in the share of male (quota) teachers (.065), pupils have a .027 percentage points higher likelihood of applying, corresponding to a 3% increase over the mean of the dependent variable.[37] [38]

Table 4 reports IV results on the full set of outcomes regarding pupils' application timing and choices after compulsory schooling.[39] Having more male quota teachers makes pupils more likely to apply directly in their final year of middle school and less likely to postpone applying to up to five years later. When considering the allocation of slots, pupils are more likely to get one of their top two choices.[40] These patterns translate into higher enrollment rates in upper secondary education in general, and significantly so in the year in which students turn 16. Why are pupils who are exposed to more male quota teachers more successful in obtaining their preferred choice? In Appendix Table A5, we document that having more male quota teachers makes pupils apply more in line with attainable options.[41]

[36] Appendix Table A3 reports IV estimates for the main application outcomes in this section without region-by-cohort fixed effects, documenting that results are not sensitive to this choice.

[37] A 1 SD increase in male (quota) teachers corresponds relatively closely to the average within municipality change in the share of male (quota) teachers over our study period of 5 ppt. Standardized effect sizes are reported in the bottom row of Table 4.

[38] We can also ask how many pupils in a school this corresponds to. An increase in the share of male teachers of .065 corresponds roughly to switching out 1 in 15 teachers from marginal female to quota male at a local school. As the average class size is 20 pupils, this place would have a total of 300 pupils, and therefore about 9 pupils switch their application status.

[39] Appendix F.1 reports the full set of corresponding reduced form results and Appendix F.2 displays grade-level reduced form estimates.

[40] Results for the full set of mutually exclusive categories regarding which slot pupils obtain are reported in Appendix Table A4.

[41] Specifically, we check whether pupils are more sophisticated in their applications with effects reported directly by pupil gender. Male pupils are more likely to include any vocational training option among their choices (column 2), while refraining from applying exclusively to academic high schools (column 3). For girls, the effect goes in the opposite direction. When examining which track pupils obtain, the margin for boys shifts from no slot (column 4) towards a vocational spot (column 5), while girls are more likely to obtain an academic rather than a vocational spot (column 6). As such, boys adjust their aspirations downwards, which prevents them from ending up without a slot, and girls correctly have high aspirations as they get into academic high schools.

6.4 Long-Term Outcomes: Labor Force Attachment and Educational Attainment

Higher exposure to male quota teachers has positive impacts on pupils' continuation of education beyond compulsory schooling at age 16, but do these patterns translate into longer-term gains? This section explores the impacts of male quota teachers for outcomes in young adulthood. We examine whether positive impacts on applications and enrollment translate into higher human capital and labor market attachment. As obtaining post-compulsory education in Finland is considered a pre-requisite to prevent social exclusion and to successfully transition into the labor market (Virtanen, 2016; Niemi et al., 2016), these are particular relevant outcomes from a policy perspective.

Educational attainment: As pupils show a higher attachment to education after middle school, we first trace whether pupils have obtained more human capital as young adults. After compulsory education, the Finnish education system has two tracks: vocational and academic. Standard three-year vocational degrees offer training in occupation-specific skills, but can be augmented with an additional year for further specialization or academic high school course work ("vocational plus"). The typical study path for the tertiary level is at polytechnics. The academic path leads from a three-year high school degree to a Bachelor's (3 years) and Master's degree (2 years) at university. Appendix Figure A24 shows the organization of the Finnish education system in detail.

Table 5 presents IV results for educational attainment by examining the highest degree achieved by age 25 using mutually exclusive education categories.[42] We observe a shift towards higher attainment both in vocationally-oriented as well as in academic education paths. As such, we observe a shift away from remaining with compulsory education or a standard three year vocational degree only, towards a "vocational plus degree". Turning to academically oriented degrees, we similarly observe a shift away from high school degrees towards having completed a university bachelor level degree.

Labor market attachment: We next examine pupils' labor market attachment at age 25. As many youths are still studying at this age, but are classified as employed due to part time work, we combine the categories of being a student and being employed into one measure that reflects not sitting idle.[43] For this age group, this metric is considered relevant to measure the propensity to successfully integrate into the labor market (Eurostat, 2021; OECD, 2021).

Table 6 reports effects for mutually exclusive labor market status categories. Being exposed to more male quota teachers during primary school results in higher likelihood of being either employed or a student at age 25. For a 1 SD increase in the share of male (quota) teachers, pupils have a 0.03 percentage point higher likelihood of working or studying, which corresponds to a 4% increase over the mean. While we observe no effect on unemployment, pupils are somewhat less likely to be on a disability pension, and significantly less likely to be out of the labor force

[42] Reduced form results are reported in Appendix F. Appendix Figure A10 shows the reduced form for the main long-term outcomes grade by grade. As longer-term outcomes are increasingly impacted by a variety of factors other than male quota teachers, the estimated coefficients are noisier when compared to patterns at age 16, but patterns generally mirror the first stage dose-response function.

[43] At age 25, 40% of pupils in our estimation sample are still enrolled in post-compulsory education.

for reasons other than disability.[44] Table 7 shows the first stage, reduced form and IV results for gradually adding in controls for the main outcome (Employed/Student) of this section.

We can also examine realized fertility up to age 26. Consistent with our finding that pupils invest more in education and have a higher attachment to the labor force when exposed to more male quota teachers, we document in Appendix Section D.3 that female pupils are less likely to have given birth by age 26, which is indicative of delaying fertility.[45]

7 Long-Term Role Model Impacts of the Quota

Having established male quota teachers' positive impact on pupil outcomes, we turn to examine whether the quota had impacts along two dimensions of role model effects: Same-gender match effects between teachers and pupils in schools, and impacts on the occupational choices of the next generation more broadly. Our setting provides us with the rare opportunity to gauge such role model effects with an exogenous shift in the composition of an entire occupation, which allows us to overcome two key limitations of prior work. Studies examining teacher match effects typically estimate impacts based on a fixed set of teachers, which does not to take into account general equilibrium effects when changing the composition of teachers. Similarly, our understanding of the effects of quotas on the occupation choices for future generations has been constrained by the fact that many recent quota policies, such as those for female board room members, affect only a small set of workers and potential aspirants.

7.1 Same-Gender Role Model Effects in Academic Outcomes

While the main effects clearly demonstrate that the overall impact of the quota was positive, this could mask heterogeneous effects by pupil gender. In the presence of same-gender role model effects, boys would benefit more from having more male teachers relative to girls. This further raises the question whether girls were made worse off when having fewer female role models to aspire to.

Figure 8 and Appendix Table A10 report heterogeneity by pupil gender for the main application outcomes at age 16. We run our main specification (Equation 9) with separate treatment effects for boys and girls while estimating controls and fixed effects jointly.[46] Girls' outcomes are not negatively impacted from exposure to male quota teachers. We test whether boys benefited more from male quota teachers, with p-values reported in the bottom row of Appendix Table

[44]With a simple back-of-the-envelope calculation, the increase in the propensity to apply to post-compulsory education can account for close to a quarter of the labor force attachment effects (the raw difference in labor force attachment for pupils who directly apply against those who do not is 25.4 percentage points).

[45]In Finland, there is low prevalence of teenage pregnancies and the average age at first birth increased from 27.2 in 1995 to 29 in 2016 and is close to the OECD average (OECD, 2019). For male pupils there is a small increase in the likelihood of having a first child by age 26, but it is statistically not significant and economically small (results not displayed).

[46]Estimating heterogeneity by pupil gender requires taking a stance on how to account for controls and common shocks that are absorbed by municipality and cohort-by-region fixed effects. When estimating these fixed effects jointly, the underlying assumption is that e.g. we expect time-varying region-specific economic shocks to affect the choice of whether to apply for boys and girls to a similar extent. We report split sample results for all main outcomes in Appendix E.2.

A10: For educational outcomes at age 16, we cannot reject the null hypothesis of the coefficients being the same for boys and girls for any outcome at the 5% level.

We report impacts on long-run outcomes by gender in Appendix Section E. While some coefficients for highest degree achieved differ, these are the ones where boys are not benefiting as much as girls. There are significant differences by pupil gender for labor market outcomes at age 25, but the gendered pattern is sensitive to the choice of whether to estimate fixed effects jointly or separately. We cannot reject that coefficients by gender are the same for specifications estimated with separate controls and fixed effects in Appendix E.2.

Taken together, we do not detect main effects that differ systematically and significantly by pupil gender, in particular when considering the main educational achievement outcomes. This finding suggests that simply recruiting more men to primary schools may have limited potential to allow boys lagging behind to catch up – possibly so because same-gender role model effects may be more limited in scope when the matched group is not in a minority role.[47]

7.2 Occupational Choices and Role Model Effects

We next turn to study whether exposure to more male quota teachers inspires pupils to pursue different fields of education. Male teachers could be setting an important example of men working in an occupation that is otherwise female-dominated. As such, they may inspire primarily boy pupils to pursue a teaching-related field. On the other hand – and separate from a classical same-gender role model effect – male teachers could also more broadly motivate pupils to pursue different education fields. This could be e.g. via male teachers' skills in particular subjects. As documented in Section 2.2, male teachers are on average more likely to have chosen math as one of their matriculation exam fields, and may thus be more skilled or motivated to teach mathematically-oriented topics.

In order to investigate these hypotheses, we measure pupils' choice of educational field at age 25. We classify their career choices via their field of education rather than their occupation because many youths at this age are still studying. For each pupil in our sample, we pick the field of the highest degree acquired if they are no longer a student and the field of their current degree if they are still studying. We define fields as primarily female- or male-dominated based on the generation prior to our sample, i.e. the 13 cohorts who are seven years old during the years 1975-87. If a group constitutes more than 60% within a field and degree level cell, we define the field as male or female leaning, and gender neutral otherwise.[48] This results in 30% of pupils being in "Male" fields, 43% in gender-neutral, and 27% in "Female" fields. We also report results on STEM and STEM-M (STEM plus Medical) fields as well as teaching-related fields in

[47] In line with this notion are studies that do not find evidence for same-gender match effects across a range of OECD countries (Cho, 2012) and Sweden (Holmlund and Sund, 2008). Most work that has documented positive same-identity match effects has been in settings in which the matched group was in some sense disadvantaged or underrepresented, i.e. Gershenson et al. (2022) document positive match effects for black pupils in the US. Lim and Meer (2017) focus on girl pupils in Korea, where gender norms are more conservative.

[48] I.e. we define share female based separately for a vocational degree in business vs. an academic degree in business. For the group that has never finished a degree beyond compulsory education and is currently not a student (9.8% of the sample), we assign the gender share of compulsory education, which is categorized as a gender-neutral field based on the previous generation.

general and primary school teacher in particular.[49]

Appendix Section D.4 reports results on the choice of education field. We observe a somewhat noisy shift away from gender neutral towards both more male- and female-dominated fields. Turning to STEM and STEM-M, pupils are significantly more likely to take up such fields when exposed to more male quota teachers, with effect sizes corresponding to a .08 and .09 SD increase for a 1 SD increase in the share of male quota teachers, respectively. In Appendix Tables A8 and A9, the STEM shift is similarly pronounced for both pupil genders.[50] Regarding teaching fields overall and primary teacher education specifically, we fail to reject a null effect.

We can do a back-of-the-envelope calculation to gauge the extent to which labor market outcomes could be explained from field choices. The shift towards STEM fields can account for about 5.8% of the total increase in labor force attachment measured in Section 6.4, as pupils with a STEM field have a five percentage point higher attachment to the labor force (0.88 vs 0.83 for non-STEM pupils).[51] In summary, we find limited scope for the quota having been able to diversify teaching occupations for the next generation, but the shift towards STEM fields is consistent with pupils aiming for higher earning fields in the labor market.

8 Mechanisms

Our results indicate that teacher teams with more male quota teachers performed better, analogous to having had higher "value-added": Conditional on socioeconomic background, pupils who experienced a higher share of male teachers via the quota have better outcomes. Turning to mechanisms, we differentiate between two potential ways in which selection without the quota may fail to capture benefits from a more diverse teacher workforce.

First, gains from diversity may arise when selection criteria do not properly account for a candidate's teaching ability at an individual level. Selection on scores unconditional of gender can miss out on high quality male teachers if men's lower scores do not map into equally lower teaching ability as illustrated in the conceptual framework in Section 3. Second, if there are complementarities in production between male and female teachers, i.e. through specialization according to comparative advantage, overall teacher quality may be lower when fewer men are in the pool of available teachers.

In the following, we document limited scope for complementarities and present three pieces of evidence that underscore the presence of imperfect selection of talent as a driving force: We first document that the weighting of fields in the final score of the matriculation exam attaches considerably less weight to mathematics and natural science fields, in which men perform

[49] STEM fields are defined based on the three-digit classification of Statistics Finland in one of the following fields: Agricultural Sciences (incl. Forestry and Fishery), Biology, Engineering, Environmental Sciences, ICT, Mathematics and Statistics, Physical Sciences, Veterinary Science, and the 4-digit category related to Materials Sciences (glass, paper, plastic and wood). STEM-M in addition includes the 3-digit field Health.

[50] In all of the regressions on field choices, we do not estimate joint fixed effects for both genders, but report separate regressions by splitting the sample. We do this since for these gendered outcomes, the assumption that shocks would affect boys and girls similarly does not seem justified (i.e. a shock that raises demand for health care workers is likely to have quite different effects on young women vs. young men).

[51] Increasing pupils exposure to quota men results in an increase of 0.6 of switching into STEM fields. Multiplied by a 0.05 higher participation rate results in a 0.03 increase in labor force attachment. This corresponds to about 5.8% of the estimated increase in labor force attachment measured in section 6.4.

relatively better. We then show that performance on the exam is not predictive of teacher team performance, suggesting that the score criterion inefficiently disadvantages men. Lastly, we document patterns of compensating differentials in the labor market for men and women: Among primary school teacher applicants, we show that similarly qualified men suffer a wage penalty when entering teaching as they forgo higher earning outside options. This is in contrast to women, who enjoy a wage premium. These patterns underline both that (marginal) men's skills are valued in the labor market, and that men in teaching may have high intrinsic motivation as they forgo potential earnings.

8.1 Complementarities in production

We test for the presence of complementarities in production by assessing marginal returns to male quota teachers along the distribution of the share of male teachers at baseline (i.e. in 1990). If male and female teachers are complements, adding an additional male teacher at a place with mostly female teachers should have larger marginal returns compared to adding an additional male teacher in an environment that is close to gender parity. We split the sample by the median share of male teachers in a municipality. The first group has initially a lower share of male teachers (average: 29%), and the second group a relatively higher share of male teachers (average: 43%).[52] Appendix Table A18 shows the reduced form for the main outcomes.[53] The magnitude of coefficients across places with high and low share of male teachers is qualitatively similar and we cannot reject that they are the same. These patterns suggest limited scope for complementarities in production driving the positive impacts of the quota.

8.2 Imperfect selection of individual talent

We present three pieces of evidence that taken together explain why men's lower performance on the selection criteria in the teacher admission process do not result in lower performance on the job.[54]

8.2.1 Gender differences in matriculation exam scores

Weighting of fields in the matriculation exam: During the time of our study, applicants to primary school teaching were ranked based on their average performance across the four mandatory fields of the matriculation exam. These four fields consisted of a test in the mother

[52]We can split the sample at other points in the distribution, with results qualitatively similar across different splits. Outside of complementarities in municipal-level teacher collaboration across schools (e.g. for curriculum design), this interpretation presumes that the gender composition in schools within the same municipality corresponds to the average municipal gender composition.

[53]We focus on the reduced form here since splitting the sample across municipalities renders relatively noisy IV estimates due to loss of power in the first stage. The IV results are displayed in Appendix Tables A19 and A20 and lead to a similar conclusion.

[54]We can observe data for matriculation exam scores used in the first step of the admission process directly. The patterns in Section 8.2.3 show that how this exam score ties into admission probability overall does not differ by applicant gender post-quota (see in particular Table A22), highlighting that the second round similarly fails to recognize or make up for skills of male applicants that exam scores may not fully account for. As pointed out by several studies of the teacher selection process in Finland, multiple components of the second selection step in admissions also emphasize cognitive performance, with some departments directly giving additional points for grades (Malinen et al., 2012; Izadi, 2021; Mankki and Räihä, 2020).

tongue (Finnish or Swedish), the respective second national language, a foreign language, and the student's choice of either a mathematics exam or a science exam ("Reaali"), the latter containing a test battery in both natural and social sciences (Kupiainen et al., 2018).

The score that was used to rank primary school teacher applicants in the first admission step thus put a weight of 75% on performance in language fields, and a weight of 25% on performance in either mathematics or sciences. As women perform markedly better in languages, the relative importance of language performance in determining university admissions across a vast number of fields has been a source of criticism (Kupiainen et al., 2018).

We examine gender differences in performance across exam fields during the quota period in Figure 9, which displays the difference in average test scores by field for the full population of exam takers for three cohorts in the mid-1980s. The overall exam score is composed of the average across the four mandatory fields, with men scoring about .1 grade points lower relative to a 4.2 grade point mean for women.[55] In the full population of exam takers, men score substantially lower on all three language fields compared to women, but higher both in absolute and relative (about .35 grade points) terms on mathematics and sciences.

While we do not observe applicants to primary school teachers studies for this time window, we can repeat the analysis for those who eventually become teachers in Appendix Figure A11. Among primary teachers, men perform worse on all fields compared to women, but the relative patterns are similar to the overall population. Men score about .25 grade points higher on mathematics and sciences compared to language fields.

Timing of exam: A further circumstantial factor that may contribute to men's lower performance on the matriculation exam is that men start compulsory military service usually directly after the end of high school and having sat the exam for the first time. Any repeats of the exam – either because of a failed first attempt or to raise prior grades – will thus typically occur during the time of men's full-time military training, as shown in Appendix Figure A12.

Taken together, multiple dimensions of the matriculation exam score, that served as a main criterion for primary teacher admissions, suggest that men's skills may be discounted when ranking primary teacher applicants. However, the differential weighting of fields could be justified and even desirable if teacher performance on languages is paramount for pupil outcomes. We therefore continue by examining the relationship between pupil outcomes and teacher test performance across the different fields of the matriculation exam.

8.2.2 Impact of teacher scores on pupil outcomes

In Appendix Table A21 we examine the relationship between teachers' matriculation exam scores and "teacher team value-added" at the cohort-by-municipality level.[56] We estimate a precise zero

[55] Exam scores range from 2-6, with 0 assigned for failing the exam. If exam takers took both mathematics and sciences, the better of the two scores is considered for the overall score. About two thirds of test takers take both fields.

[56] We use the term "teacher team value-added" to refer to teacher team fixed effects that vary at the cohort-by-municipality level. We thus exploit variation in teacher team test performance that different pupil cohorts within the same municipality are exposed to. Estimating teacher team value-added by residualizing pupil outcomes with family background characteristics (see Chetty et al., 2014) is equivalent to controlling for pupil characteristics in this set up.

for the impact of teachers' matriculation exam score on pupils' outcomes. For a 1 SD increase in a teacher team's overall score (0.2), we can rule out effects on pupils' application likelihood that are larger than .002 (or .2 of a percent over the mean) as the upper and smaller than -0.003 (or -.3 of a percent over the mean) as the lower bound. Splitting the overall exam score by language and grades in math or sciences in column 2 and 3 show similarly negligible impacts. Teacher test scores not explaining teacher (team) value-added is not unique to our setting, with many studies having documented similarly negligible relationships between teacher test performance and pupil outcomes (Kane et al., 2008; Angrist and Guryan, 2008; Harris and Sass, 2011; see Jackson et al. (2014) and Hanushek and Rivkin (2006) for a review).

Since teacher's matriculation exam score is not informative of teacher team value-added, candidate selection on these scores thus inefficiently constrains the pool of male teachers, who on average perform considerably lower in particular on the language sections of this exam. In other words, a re-weighting of exam fields with more emphasis on math and sciences would diversify the pool of admits without detrimental impacts on output.

8.2.3 Compensating differentials

Do the selection criteria used in primary teacher admissions fail to capture valuable skills of male teachers more generally? We examine the returns to male and female teachers' skills by analyzing compensating differentials in the labor market across marginal candidates. For this exercise, we compare similarly scoring applicants for primary school teacher studies by gender in order to understand the outside options of admits. Our sample contains the universe of applicants to primary school teacher studies in the years 2000 - 2005 (i.e. post-quota, with the pool of male applicants containing men who would have been accepted during the quota), linked to candidates' performance on the matriculation exam, as well as their adult wage earnings and degree obtained in 2020.[57] University study slots in Finland are allocated in a centralized university application system by deferred acceptance. Candidates in our sample send the lion's share of their applications to primary school teaching, highlighting their desire to become a primary teacher (76% of the average male and 72% of the average female applicant's applications).

We start by documenting the relationship between matriculation exam score and the likelihood of obtaining a primary teaching degree by gender in Appendix Figure A13. For both male and female applicants, the likelihood of obtaining a primary teaching degree is increasing in score. Appendix Table A22 formalizes these patterns and confirms that there is not differential score premium by gender: Conditional on exam score, men and women face similar application success. This clearly shows that men no longer obtain a score premium post-quota. In addition, these patterns highlight that the selection criteria in the second admissions step do not differen-

[57] The year 2000 is the first time that data on application fields in the national application system is available. While the analysis in this subsection is based on data after the main study period, we think that it is reasonable to extrapolate from the described patterns to adjacent years. Data on field choice for 2003 is missing and we exclude applicants in this year from the analysis. On average, about 10% of the full population of applicants apply to primary school teacher education in a given year. Since actual admissions are unobserved, we rely on whether applicants ever obtain a primary teaching degree by 2020. Completion rates for primary teaching are at 90% (Nissinen and Välijärvi, 2011).

tially reverse the impact of exam scores for men.[58] These patterns are similar in columns 3 and 4 where we restrict the sample to those applicants who have obtained any university degree by 2020.[59]

In order to understand whether admission criteria miss out on male applicants' skills more generally, we turn to measuring compensating differentials by comparing labor market earnings among similarly qualified applicants within gender:

$$Wage_i = \beta_0 + \beta_1\, Teacher_i + \beta_2 Male_i * Teacher_i + \beta_3\, Male_i + X_i + u_i \tag{12}$$

with $Wage_i$ yearly wage earnings, $Teacher_i$ an indicator for having obtained a primary teaching degree, $Male_i$ an indicator for being male, and X_i a vector of controls for exam score, application year, age and labor market experience. Appendix Table A23 documents these compensating differentials, with women who got rejected from primary teaching as the omitted group. Women who obtain a primary teaching degree earn a wage premium of 3.400 EUR (approximately 3.400 USD) relative to their female counterparts who got rejected. For men, the teaching degree comes with a wage penalty, both in relative and absolute terms. Male degree holders have a 5.000 EUR penalty relative to women's premium, and an absolute penalty of 1.600 EUR when compared to men who do not become a primary teacher.[60] These wage penalties are more pronounced when restricting the sample to applicants who ever obtain a university degree in columns 3 and 4.[61] Taken together, these patterns suggest that among similarly scoring applicants, men possess skills that the labor market values, but that do not seem to receive sufficient consideration in the teacher admission process. At the same time, men who are becoming primary teachers are willing to accept a wage penalty, hinting at their intrinsic motivation when choosing this field.[62]

In Appendix Figures A14 and A15, we document men and women's outside options that underpin this result. When examining alternative fields that male and female primary teacher applicants apply to, we see that close to 60% of women's alternative applications go to other education fields compared to 35% for men. Men are instead more likely to pick higher earning fields, such as natural sciences and business administration or law. This maps into alternative fields obtained by 2020, depicted in Figure A15. As men are not accepted into primary teaching,

[58] I.e. the interview round does not disproportionally recognize or make up for skills of male applicants that exam scores may not fully account for (such as, for example, a stronger background in mathematics).

[59] Among applicants in this sample, men are about 10ppt less likely to obtain any university degree compared to women and instead complete vocational training and specializations.

[60] Our primary focus for this exercise lies in comparing differences between those holding a teaching degree and those who do not within each gender group. Differences in labor market participation between men and women may explain why male teaching degree holders have somewhat higher earnings relative to women, but this is not our focus.

[61] In Appendix Figure A16a, we show that these wage penalties are present throughout the exam score distribution.

[62] A potential concern with this interpretation of earnings patterns may be that men going into primary teaching could be negatively selected within the pool of male applicants (i.e. if they put in primary teaching as their lowest ranked "safe option"), whereas women would be positively selected. The data suggest that this is unlikely: If anything, men put in more applications to primary teaching than women (76% of the average male applicant's applications go to primary school teaching vs. 72% of the average female applicant's applications). Using data in which we can observe preference rankings in applications, 78% of men and 72% women rank primary teaching among their top two choices in 2017 and 2018. Since primary school teaching is among the fields that receive the most applications, listing such a competitive field as a low ranked "safe option" would make admission particularly unlikely.

they thus obtain fields with higher earnings as their next best alternative.

9 Robustness

9.1 Do Schools Change Hiring Practices due to the Reform?

Our treatment coefficients measure the effect of the quota policy, and thus include any impacts that may be due to schools responding endogenously to the policy, for example by changing their hiring patterns and recruiting more experienced teachers in lieu of rookies. While this is not a direct threat to identification, assessing these aspects helps to understand the underlying drivers of our effects. Table 8 reports municipal level regressions, with all specifications assessing changes in flows for consistency (see Equation 8). Our goal is to understand whether teacher retirements in the post-quota period differentially affect teacher exit or entry margins.

We start by assessing the effect of teachers turning 60 on the share of teachers leaving their current job in columns 1 and 2. Teachers turning 60 has almost a 1:1 impact on the share of teachers leaving, but not differentially so in the quota period. This effect is not driven by turnover of relatively younger teachers (column 2), and rather reinforces the observation that teachers reaching age 60 corresponds to actual exits from the teaching profession. In column 3 and 4, we examine how retirements affect proxies of experience of the local teacher body and do not detect a sizeable or significant change in the post-quota period. Column 5 shows (noisily estimated) that retirements in general result in a higher share of new entrants among newly arriving teachers at a municipality, but this does not change differentially in the post-quota period. Taken together, we fail to find corroborating evidence for changed teacher exit or re-hiring strategies as a response to the quota reform.

9.2 Teachers on Parental Leave

Apart from hiring patterns, the lifting the quota coincides with bringing more young female teachers to schools, who may have a higher propensity to go on leave when giving birth. The positive effects we detect from having more male quota teachers could then simply arise from pupils having less teacher turnover. During the 1990s, Finland provided 6.5 months of entirely shareable parental leave taking effect after three months of birth-related maternity leave (Kamerman and Moss, 2009). To check whether any changes related to leave taking of teachers becoming mothers (or fathers) could affect pupils, we repeat the municipal first stage regressions. Table A24 (in Appendix H.1) shows that teachers turning 60 in the post period do not have a differential impact on either female or male teachers having a birth in their household. The share of a female teachers leaving the teacher force subsequent to becoming a mother is also not differentially affected by retirements in the post-quota period (column 4). In these specifications, the variation used stems from such patterns arising immediately as a response to teacher retirements. We therefore also document that, conditional on municipal and region-by-cohort fixed effects, higher exposure to female teachers having a newborn child does not impact pupil outcomes (Appendix Table A25). We conclude that differential leave taking patterns due to maternity from more female teachers post-quota are unlikely to drive our results.

9.3 Placebo Test and Randomization Inference

To further assess robustness, we perform a placebo test in Appendix Figure A17, by estimating reduced form treatment effects for additional "fake grades", i.e. years in which pupil cohorts have already left primary school. We show that any teacher retirements in those "fake grades" do not affect pupils' application outcomes.

In addition, Appendix Figure A18 plots the distribution of reduced form treatment coefficients for applications (Equation 10), estimated from randomly re-assigning the share of retiring teachers across municipality*year cells 500 times for the post-quota period. This allows to assess significance without making parametric assumptions on the structure of the error term. The vertical line indicates the reduced form estimate (-0.071, see Appendix Table A15) and shows that our estimated treatment effect on applications would have been expected to occur zero times under the null hypothesis of no effect.

9.4 Heterogeneous Treatment Effects in Two-Way Fixed Effects Designs

An active literature has documented that in the presence of heterogeneous treatment effects, the coefficient of a two-way fixed effects (TWFE) regression, $\hat{\beta}_{fe}$, may be a biased estimate of the treatment effect and in severe cases exhibit the opposite sign.[63] If treatment effects are heterogeneous, such bias arises when already treated units are used as a control group in later periods. In a two stage least squares (2SLS) set-up, potential issues would arise from residualized treatment assignment in the first stage (which is then used to generate predicted values of the endogenous variable for the second stage), *if* treatment effects are heterogeneous. In our setting, however, the first stage portrays a relationship between local retirements and teacher gender composition that should be purely mechanical, and for which – given our knowledge about the quota reform – we have a clear ex ante prior on sign and magnitude. While the TWFE literature to date has not tackled settings with continuous treatment variables, we follow the reasoning outlined in De Chaisemartin and d'Haultfoeuille (2020) to discuss negative weights and potential heterogeneity in treatment effects in Appendix H.5. We further probe whether first stage coefficients are driven by particular years, regions, or levels of treatment assignment in Appendix H.6 and Appendix Figure A20. We conclude that treatment effect heterogeneity leading to sign reversal in $\hat{\beta}_{fe}$ is not a major concern in our setting.

9.5 Further Robustness

Appendix H.3 documents further sensitivity checks, showing that results are not driven by selective attrition in the pupil sample, the capital or large cities in general. While our exposure measure is defined as the share of male (quota) teachers and thus keeps the likelihood of exposure comparable across places of different size, we also report reduced form results separately for small municipalities with 12 teachers or fewer, i.e. maximally two classes per grade level. The coefficients are unchanged. We further discuss the main macro-economic shocks in Finland

[63] See, among others: De Chaisemartin and d'Haultfoeuille, 2020; Arkhangelsky et al., 2021; Athey and Imbens, 2021; Sun and Abraham, 2020; Imai and Kim, 2021; Goodman-Bacon, 2021.

during our study period.

10 Conclusion

In this paper, we document that a quota that advantaged academically lower scoring men to obtain a study slot for primary teacher education has positive effects on output as measured by their pupils' intermediate and long-run educational and labor market outcomes. Using comprehensive register data, we show that male quota teachers impact consequential application patterns to post-compulsory education: Pupils are more likely to apply to continue education directly after middle school, to obtain their preferred study slots and to enroll. We find that pupils who were exposed to a higher share of male quota teachers during their time in primary school are more likely to be either employed or studying at age 25, and have higher educational attainment as measured by their highest degree achieved.

We do not find evidence that our main effects are more pronounced for boy pupils and we show that boys in particular are not more likely to choose education or teaching related fields. These findings point towards same-identity role model effects being less pronounced in contexts where the targeted group (i.e. boys in primary school) is not a minority and highlight promising avenues for further research with respect to *when* role model interventions can make a difference.

Instead, multiple pieces of evidence suggest that the quota in our setting fixed selection imperfections. While selection criteria in absence of the quota were group-neutral 'on paper', in practice they discounted skills of male applicants. The quota thus helped to off-set such discounting.

Our study evaluates the effects of a quota policy with fixed parameters and does therefore not aim to take a stance on the design of an optimal policy. However, the key trade-offs it highlights generalize to any setting in which more equitable representation is a policy goal. When a main criterion for choosing candidates discounts the abilities of an underrepresented group, conscious re-weighting of criteria can help to overcome such misalignment irrespective of the chosen policy instrument. Our results suggest that this may pay off not only in terms of achieving more equitable representation, but also in terms of economic efficiency.

References

Ahern, K. R., & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *The Quarterly Journal of Economics*, *127*(1), 137–197.

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? evidence from state certification requirements. *Economics of Education Review*, *27*(5), 483–503.

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics*, *33*(1), 63–89.

Arcidiacono, P., & Lovenheim, M. (2016). Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, *54*(1), 3–51.

Arkhangelsky, D., Imbens, G. W., Lei, L., & Luo, X. (2021). Double-robust two-way-fixed-effects regression for panel data. *arXiv preprint arXiv:2107.13737*.

Arnold, D., Dobbie, W. S., & Hull, P. (2020). *Measuring racial discrimination in bail decisions* (Working Paper). National Bureau of Economic Research.

Athey, S., & Imbens, G. W. (2021). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*.

Bagues, M., & Campa, P. (2021). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics*, *194*, 104315.

Baltrunaite, A., Bello, P., Casarico, A., & Profeta, P. (2014). Gender quotas and the quality of politicians. *Journal of Public Economics*, *118*, 62–74.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: Does exposure reduce bias? *The Quarterly Journal of Economics*, *124*(4), 1497–1540.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago Press.

Besley, T., Folke, O., Persson, T., & Rickne, J. (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from Sweden. *American Economic Review*, *107*(8), 2204–42.

Beuermann, D. W., Jackson, C. K., Navarro-Sola, L., & Pardo, F. (2023). What is a good school, and can parents tell? evidence on the multidimensionality of school output. *The Review of Economic Studies*, *90*(1), 65–101.

Black, S. E., Denning, J. T., & Rothstein, J. (2020). *Winners and losers? The effect of gaining and losing access to selective colleges on education and labor market outcomes* (Working Paper). National Bureau of Economic Research.

Bleemer, Z. (2021). Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209* [qjab027]. *The Quarterly Journal of Economics*.

Bohren, J. A., Hull, P., & Imas, A. (2022). *Systemic discrimination: Theory and measurement* (tech. rep.). National Bureau of Economic Research.

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, *125*(3), 1101–1144.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124–27.

Chan, J., & Eyster, E. (2003). Does banning affirmative action lower college student quality? *American Economic Review*, *93*(3), 858–872.

Chattopadhyay, R., & Duflo, E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, *72*(5), 1409–1443.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–79.

Cho, I. (2012). The effect of teacher–student gender matching: Evidence from oecd countries. *Economics of Education Review*, *31*(3), 54–67.

Coate, S., & Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 1220–1240.

Cortés, P., Kasoolu, S., & Pan, C. (2021). *Labor market nationalization policies and exporting firm outcomes: Evidence from Saudi Arabia* (Working Paper). National Bureau of Economic Research.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528–554.

Dhume, S. (2019). India's ethnic quotas are a cautionary tale [2019-01-24]. *The Wall Street Journal*.

Eckbo, B. E., Nygaard, K., & Thorburn, K. S. (2021). Valuation effects of Norway's board gender-quota law revisited. *Management Science*.

Ellison, G., & Pathak, P. A. (2021). The efficiency of race-neutral alternatives to race-based affirmative action: Evidence from Chicago's exam schools. *American Economic Review*, *111*(3), 943–75.

Etelä Suomen Sanomat. (1988). Koulutuspaikkojen lisäzksellä eroon opettajapulasta [1988-08-25]. *Etelä Suomen Sanomat*.

Eurostat. (2021). Education and training in the EU - facts and figures [Online; accessed 21-Sep-2021].

Ferrari, G., Ferraro, V., Profeta, P., & Pronzato, C. (2021). Do board gender quotas matter? Selection, performance and stock market effects. *Management Science*.

Finnish National Agency for Education. (2018). *Finnish teachers and principals in figures*. Finnish National Agency for Education.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C., & Papageorge, N. W. (2022). *The long-run impacts of same-race teachers* (tech. rep.).

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Green, E. (2022). The Inherent Contradictions in the Affirmative-Action Debate [2022-10-31]. *The New Yorker*.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, *2*, 1051–1078.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798–812.

Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery.* National Academy Press.

Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics*, *15*(1), 37–53.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and US economic growth. *Econometrica*, *87*(5), 1439–1474.

Huttunen, K., Pekkarinen, T., Uusitalo, R., & Virtanen, H. (2023). Lost boys? secondary education and crime. *Journal of Public Economics*, *218*, 104804.

Imai, K., & Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, *29*(3), 405–415.

Izadi, R. (2021). *Teacher selection in Finland* (Unpublished manuscript).

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

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), 801–825.

Jakiela, P. (2021). Simple diagnostics for two-way fixed effects. *arXiv preprint arXiv:2103.13229*.

Kamerman, S., & Moss, P. (2009). *The politics of parental leave policies: Children, parenting, gender and the labour market.* Policy Press.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from New York City. *Economics of Education review*, *27*(6), 615–631.

Kivinen, O., & Rinne, R. (1994). The Thirst for Learning, or Protecting One's Niche? The shaping of teacher training in Finland during the 19th and 20th centuries. *British Journal of Sociology of Education*, *15*(4), 515–527.

Kofoed, M. S. et al. (2019). The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at West Point. *Journal of Human Resources*, *54*(2), 430–467.

Kupiainen, S., Marjanen, J., & Ouakrim-Soivio, N. (2018). Ylioppilas valintojen pyörteissä: Lukio-opinnot, ylioppilastutkinto ja korkeakoulujen opiskelijavalinta [a student in a whirlwind of choices: High school studies, high school diploma and student selection of universities].

Leman, N. (2021). Can affirmative action survive? [2021-07-26]. *The New Yorker.*

Li, D., Raymond, L. R., & Bergman, P. (2020). *Hiring as exploration* (Working Paper). National Bureau of Economic Research.

Liimatainen, S. (2002). Opettajankoulutuksen valintayhteistyö. In P. Räihä & J. Kari (Eds.), *Opettajaksi soveltuvuuden moni-ilmeisyys. Opiskelijavalintavaltakunnallisesti puntaroituna.* (pp. 24–31). Jyväskylän yliopisto.

Liiten, M. (2012). Opetus- ja sosiaalialan koulutukseen harkitaan mieskiintiöitä [2012-10-06]. *Helsingin Sanomat.*

Lim, J., & Meer, J. (2017). The impact of teacher–student gender matches random assignment evidence from South Korea. *Journal of Human Resources*, *52*(4), 979–997.

Lim, J., & Meer, J. (2020). Persistent effects of teacher–student gender matches. *Journal of Human Resources*, *55*(3), 809–835.

Long, H. (2019). 80 nations set quotas for female leaders. Should the U.S. be next? [2019-03-29]. *Washington Post*.

Lundberg, S. J., & Startz, R. (1983). Private discrimination and social intervention in competitive labor market. *The American Economic Review*, *73*(3), 340–347.

Malinen, O.-P., Väisänen, P., & Savolainen, H. (2012). Teacher education in finland: A review of a national effort for preparing teachers for the future. *Curriculum Journal*, *23*(4), 567–584.

Mankki, V., Mäkinen, M., & Räihä, P. (2020). Teacher educators' predictability and student selection paradigms in entrance examinations for the Finnish primary school teacher education programme. *European Journal of Teacher Education*, *43*(2), 151–164.

Mankki, V., & Räihä, P. (2020). Opettajankoulutuksen opiskelijavalintojen uudistamisen jatkumo: Tausta-agendoja ja sivuvaikutuksia [reforming student admissions in teacher education: Background agendas and side effects.] *Kasvatus & aika*, *14*(3), 4–19.

Mankki, V., & Räihä, P. (2022). The hidden admission agendas in finnish primary teacher education in the 1990s, 2000s and 2010s. *History of Education*, 1–14.

Matsa, D. A., & Miller, A. R. (2013). A female style in corporate leadership? evidence from quotas. *American Economic Journal: Applied Economics*, *5*(3), 136–69.

Niemi, H., Toom, A., & Kallioniemi, A. (2016). *Miracle of education: The principles and practices of teaching and learning in Finnish schools*. Springer.

Nissinen, K., & Välijärvi, J. (2011). Opettaja-ja opettajankoulutustarpeiden ennakoinnin tuloksia [forecasts for teacher demand and teacher training needs in finland].

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

OECD. (2019). Society at a glance 2019: Oecd social indicators [Online; accessed 21-Sep-2021].

OECD. (2021). Oecd employment outlook [Online; accessed 21-Sep-2012].

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358–369.

Otero, S., Barahona, N., & Dobbin, C. (2021). *Affirmative action in centralized college admission systems: Evidence from brazil* (tech. rep.). Working paper.

Peck, J. (2017). Can hiring quotas work? The effect of the Nitaqat program on the Saudi private sector. *American Economic Journal: Economic Policy*, *9*(2), 316–47.

Petek, N., & Pope, N. G. (forthcoming). The multidimensional impact of teachers on students. *Journal of Political Economy*.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659–661.

Räihä, P. (2010). Koskaan et muuttua saa! Luokanopettajakoulutuksen opiskelijavalintojen uudistamisen vaikeudesta. *Tampere University Press*.

Ray, D., & Sethi, R. (2010). A remark on color-blind affirmative action. *Journal of Public Economic Theory*, *12*(3), 399–406.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Rose, E. K. (2021). Who gets a second chance? effectiveness and equity in supervision of criminal offenders. *The Quarterly Journal of Economics*, *136*(2), 1199–1253.

Rose, E. K., Schellenberg, J. T., & Shem-Tov, Y. (2022). *The effects of teacher quality on adult criminal justice contact* (tech. rep.). National Bureau of Economic Research.

Sahlberg, P. (2021). *Finnish lessons 3. 0: What can the world learn from educational change in Finland?* Teachers College Press.

Sahlberg, P., Johnson, P., & Strauss, V. (2019). What Finland is really doing to improve its acclaimed schools. *Washington Post*.

Sethi, R., & Somanathan, R. (forthcoming). Meritocracy and representation. *Journal of Economic Literature*.

Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Sysiharju, A.-L. (1987). Women school employees in Finland. In P. A. Schmuck (Ed.), *Women educators: Employees of schools in Western world countries* (pp. 21–41). Suny Press.

Tasa-Arvovaltuutetu. (1987). Tasa-Arvovaltuutetun toimisto 126/1987, lausunto kiintiöiden käytöstä [Office of the High Commissioner of Equality (Paavo Nikula): Opinion on quotas 126/1987, obtained by private correspondence with the Office of the Equality Ombudsman].

Tirri, K. (2014). The last 40 years in Finnish teacher education. *Journal of Education for Teaching*, *40*(5), 600–609.

UN. (2019). Progress on the sustainable development goals: The gender snapshot 2019.

Uusiautti, S., & Määttä, K. (2013). Significant trends in the development of Finnish teacher education programs (1860-2010). *Education Policy Analysis Archives*, *21*, 59.

Uusiautti, S., Määttä, K. et al. (2013). Who is a suitable teacher? The over-100-year long history of student selection for teacher training in Finland. *International Journal of Sciences*, *2*(03), 109–118.

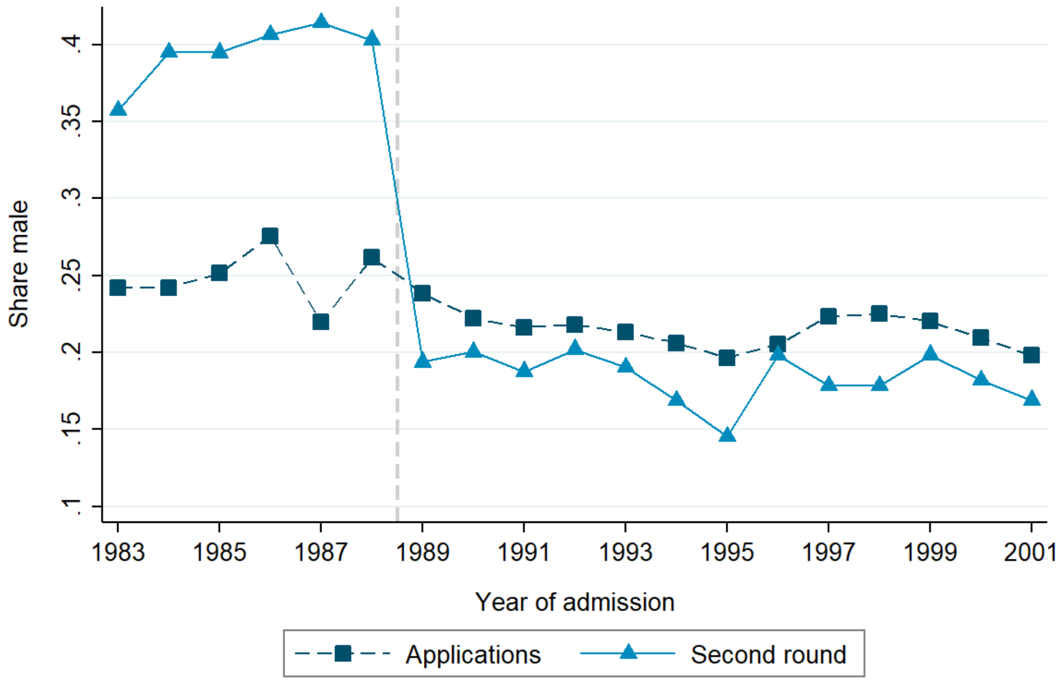
Valtiokonttori. (1988). Valintaoikeus: Valitsijan opas.

Virtanen, H. (2016). Essays on Post-Compulsory Education Attainment in Finland.

Welch, F. (1976). Employment quotas for minorities. *Journal of Political Economy*, *84*(4, Part 2), S105–S141.

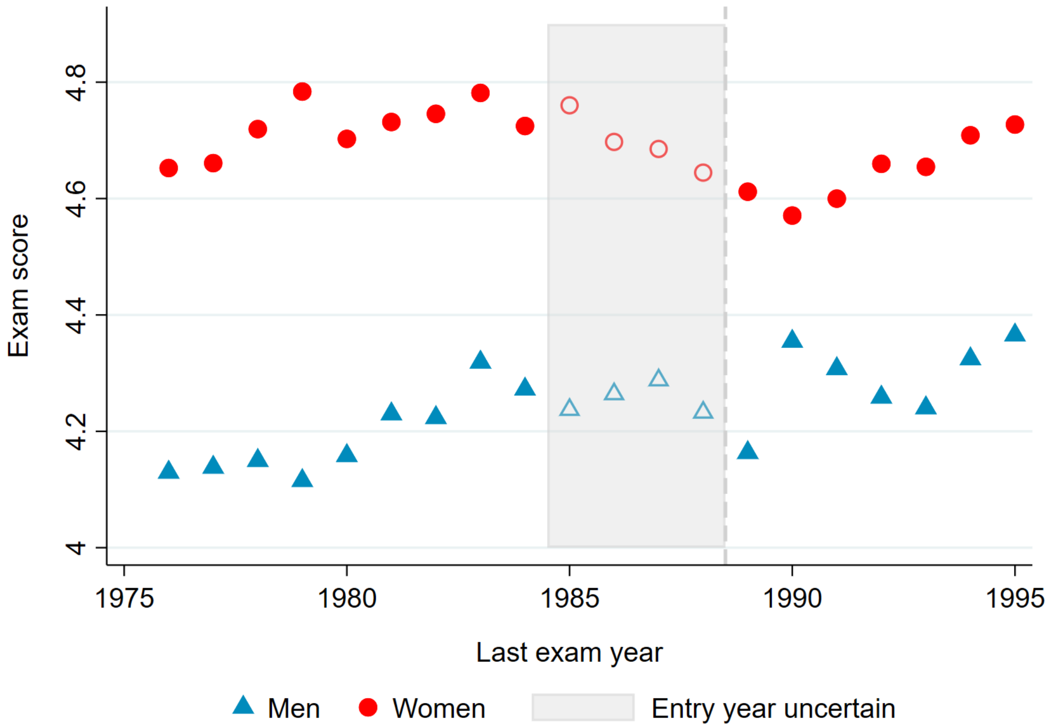
Tables and Figures

Figure 1: Share Male in Applications to Primary School Teacher Studies



Note: Share male among applicants (dark blue squares) and among invitees (light blue triangles) to the second round of admissions to primary teacher studies by year of admission. *Source:* Liimatainen (2002). (back)

Figure 2: Matriculation Exam Score among Primary School Teachers



Note: Overall score in the matriculation exam among primary school teachers, by gender and the last year in which they took the matriculation exam. The last year of taking the exam serves as a proxy for year of admission to university, which is unobserved. Exam takers in 1989 (dashed grey line) and thereafter will have studied after the male quota was abolished, but there is uncertainty with respect to start date of exam takers in the years prior to 1989.

This figure plots the final score obtained in the *first* attempt at the exam in order to get at a measure of cognitive ability that is not influenced by repeated test taking. We plot the exam grade against the date of someone's *last* exam take to most closely approximate the point of entry to university studies.

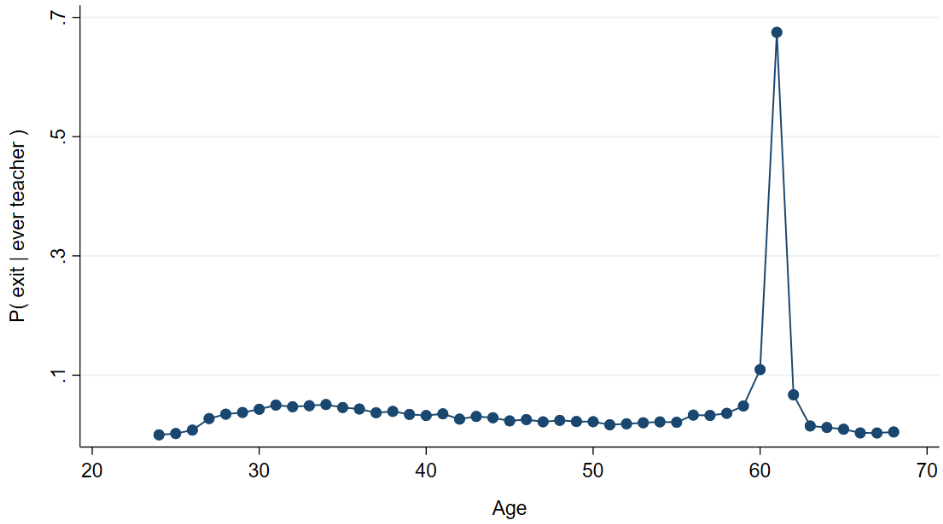
The final matriculation exam score is averaged across the grades obtained in four mandatory fields of the exam, discussed in detail in Section 8.2.1. Grades are assigned on a curve and range from 2-6, with 6 being the highest grade and 0 assigned for a failing exam (Kupiainen et al., 2018). (back)

Table 1: Summary Statistics of Teachers by Gender

	Quota			Post-Quota		
Variable	Female	Male	Difference	Female	Male	Difference
Background and current place of living						
Urban residence at birth	0.586	0.562	-0.023***	0.612	0.590	-0.022***
			(0.007)			(0.007)
Rural residence at birth	0.371	0.392	0.021**	0.358	0.376	0.018**
			(0.007)			(0.006)
Born on Russian territory	0.043	0.046	0.003	0.028	0.033	0.005*
			(0.003)			(0.002)
Finnish mother tongue	0.922	0.947	0.025***	0.916	0.940	0.024***
			(0.004)			(0.004)
Lives in region of birth	0.457	0.488	0.031***	0.473	0.496	0.023***
			(0.007)			(0.007)
Lives in municipality of birth	0.229	0.268	0.039***	0.249	0.277	0.028***
			(0.006)			(0.006)
Total observations	14,995	7,298		18,074	7,887	
Education path (born after 1952)						
High school degree	0.980	0.979	-0.002	0.983	0.980	-0.003
			(0.003)			(0.003)
Teaching degree	0.894	0.889	-0.005	0.920	0.910	-0.010
			(0.007)			(0.006)
Age at high school degree	19.25	19.47	0.23***	19.23	19.46	0.23***
			(0.02)			(0.02)
Age at teacher degree	25.54	26.50	0.96***	25.48	26.48	1.00***
			(0.09)			(0.09)
Academic performance (born after 1952)						
Matriculation exam	0.986	0.984	-0.003	0.982	0.976	-0.006*
			(0.003)			(0.002)
Overall Exam Score, first take	4.78	4.22	-0.56***	4.72	4.22	-0.49***
			(0.02)			(0.01)
Overall Exam Score, best take	4.85	4.35	-0.51***	4.79	4.35	-0.45***
			(0.02)			(0.01)
Mathematics exam	0.741	0.826	0.086***	0.768	0.842	0.074***
			(0.009)			(0.007)
Advanced mathematics exam	0.281	0.387	0.106***	0.268	0.390	0.122***
			(0.010)			(0.008)
Total observations	7,053	3,273		11,701	4,614	

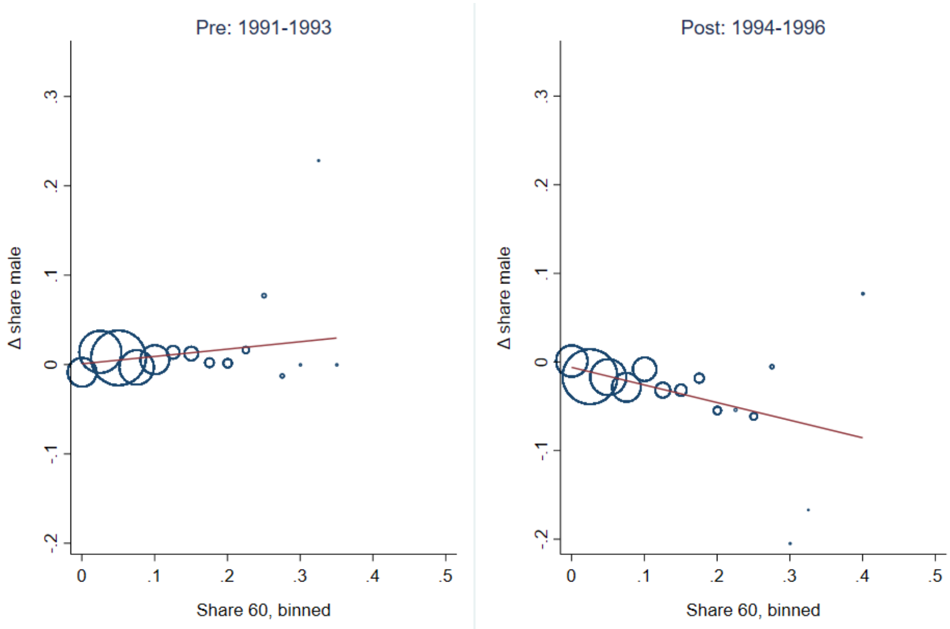
Note: Characteristics of male and female primary school teachers who are active teachers for at least one year in the quota period (1990-93) or in the post-quota period (1994-2000) and who are between 24 and 60 years old.(back)

Figure 3: Probability of Teacher Exit by Age



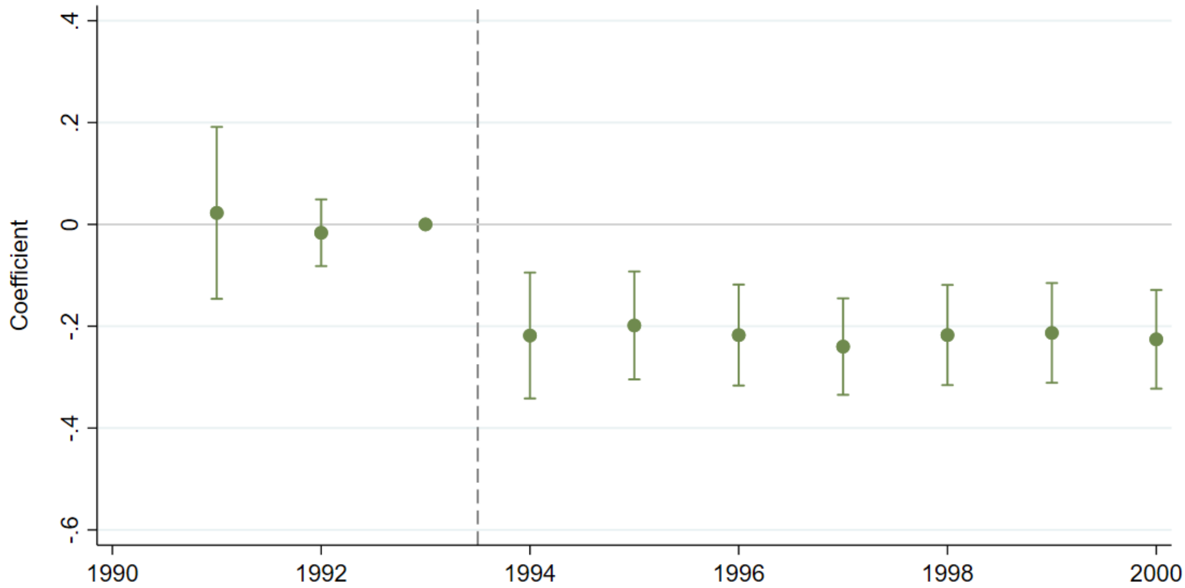
Note: Share of primary school teachers not working as a primary school teacher at a given age, conditional on having worked as a primary school teacher in the previous year. Data for all active primary school teachers in the years 1990-2000. Multiple exits per teacher possible.
(back)

Figure 4: First Stage Intuition: Changes in Share Male Teachers by Local Retirements, Raw Data



Note: Municipality level data, binned: Change in the share of male primary school teachers for a period of similar length in the quota (1991-93) and post-quota (1994-96) period against total share of teachers turning 60. Linear fit, weighted by the number of municipalities per bin.
(back)

Figure 5: First Stage: Municipal Level Event Study



Note: Year-on-year estimates of π_2 for the first stage Equation 7, showing impact of primary teachers turning 60 on the local share of male teachers (relative to 1993 as last year of the quota period). Standard errors clustered at the municipality level. Population weighted. (back)

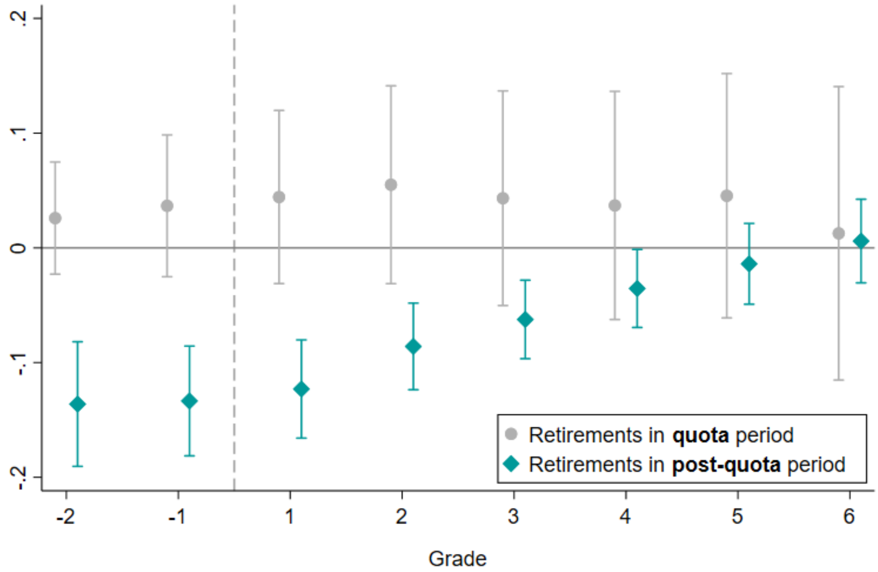
Table 2: First Stage at the Municipal Level

	First Differences				Fixed Effects		
	Δ Share Male				Share Male		
Share 60	0.062	0.062	0.070*	0.072*			
	(0.038)	(0.039)	(0.041)	(0.039)			
Share 60 * Post-Quota	–0.165***	–0.170***	–0.175***	–0.161***			
	(0.044)	(0.044)	(0.046)	(0.044)			
Total Share 60					0.068	0.099**	0.078*
					(0.043)	(0.045)	(0.043)
Total Share 60 * Post-Quota					–0.218***	–0.243***	–0.194***
					(0.049)	(0.054)	(0.049)
Municipal*Post-Quota FE					X	X	X
Year FE		X		X	X		X
Region*Year FE			X			X	
Municipal controls				X			X
Adj. R^2	0.017	0.022	0.018	0.025	0.869	0.867	0.869
Obs	4448	4448	4448	4448	4443	4443	4443
Dep mean	.0007	.0007	.0007	.0007	.3601	.3601	.3601

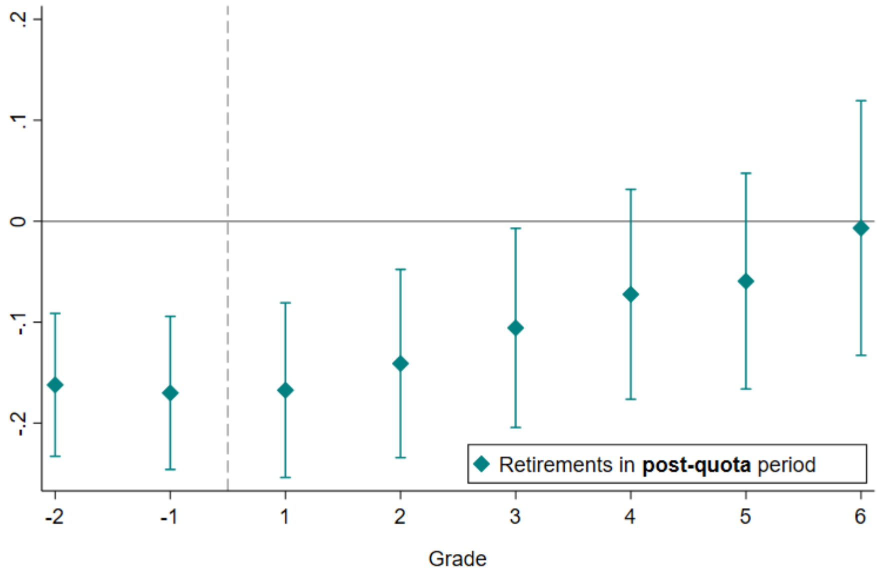
Note: Estimates for Equation 8 (columns 1-4): Year-on-year changes of the share of male teachers (Δ Share Male) on the share of teachers reaching retirement age (Share 60), and the corresponding fixed effects specification in Equation 7 (columns 5-7) of local share of male teachers on cumulative teacher retirement (Total Share 60). Observation counts between specifications change due to municipal consolidation. Standard errors clustered at the municipality level. Regressions weighted by population, means unweighted. Time-varying municipal controls include log population, log household income, share unemployed, share of families in single parent HH, share of adult population with compulsory, secondary and tertiary education. * $p<0.1$, ** $p<0.05$, *** $p<0.01$ (back)

Figure 6: First Stage by Grade: Average Share Male Teachers

(a) Separate Estimation of Quota and Post-Quota Coefficients



(b) Post-Quota (Relative to Quota Coefficients)



Note: Grade level estimation of pupil level first stage (c.f. Equation 10). Outcome is the average share of male teachers a pupil is exposed to during their time in primary school (Grades 1-6), regressed on the share of teachers turning 60 just before a pupil enters the respective grade in school (Grades 1-6), starting two years prior to a pupil entering school (Grades -2 and -1). Panel (a) estimates absolute coefficients for effect of retirement pupils experience by grade in the quota and the post-quota period. Panel (b) depicts coefficients for the post-quota period *relative* to the quota period (i.e. it shows the difference between quota and post-quota estimates depicted in Panel (a)). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

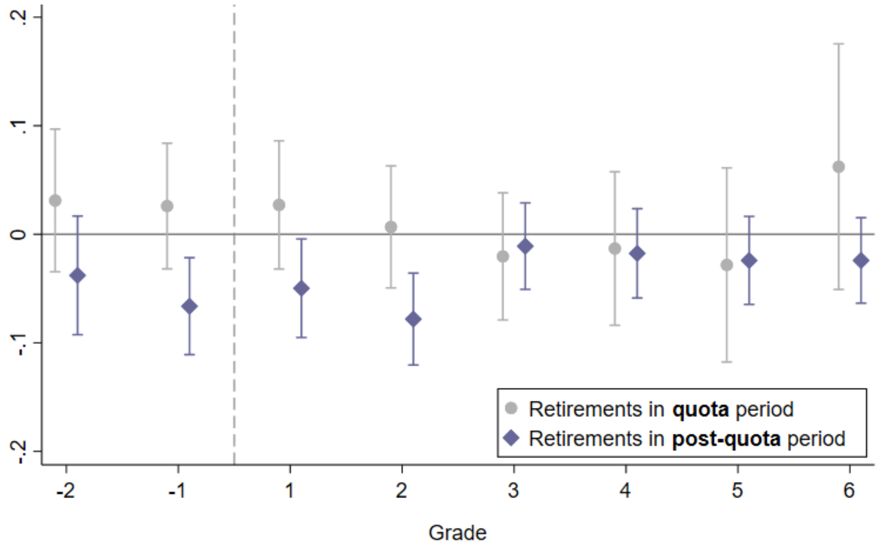
Table 3: First Stage, Reduced Form and IV: Applications for Post-Compulsory Education

	First Stage			RF			IV		
	Avg Share Male			Apply			Apply		
Total Share 60	0.033	0.033	0.043	0.048**	0.048**	0.018	0.031	0.025	−0.001
	(0.036)	(0.036)	(0.037)	(0.020)	(0.021)	(0.021)	(0.025)	(0.027)	(0.020)
Total Share 60 *	−0.176***	−0.176***	−0.168***	−0.096***	−0.122***	−0.071***			
Post-Quota	(0.042)	(0.042)	(0.043)	(0.026)	(0.031)	(0.027)			
Avg Share Male							0.544***	0.696***	0.424**
							(0.202)	(0.234)	(0.197)
Municipal FE	X	X	X	X	X	X	X	X	X
Cohort FE	X	X		X	X		X	X	
Region*Cohort FE			X			X			X
Ind. controls		X	X		X	X		X	X
MOP F^{eff}							17.66	17.64	15.28
Adj. R^2	0.916	0.916	0.922	0.038	0.069	0.070			
Obs	825,094	825,094	825,094	825,094	825,094	825,094	825,094	825,094	825,094
Dep mean	.313	.313	.313	.9106	.9106	.9106	.9106	.9106	.9106

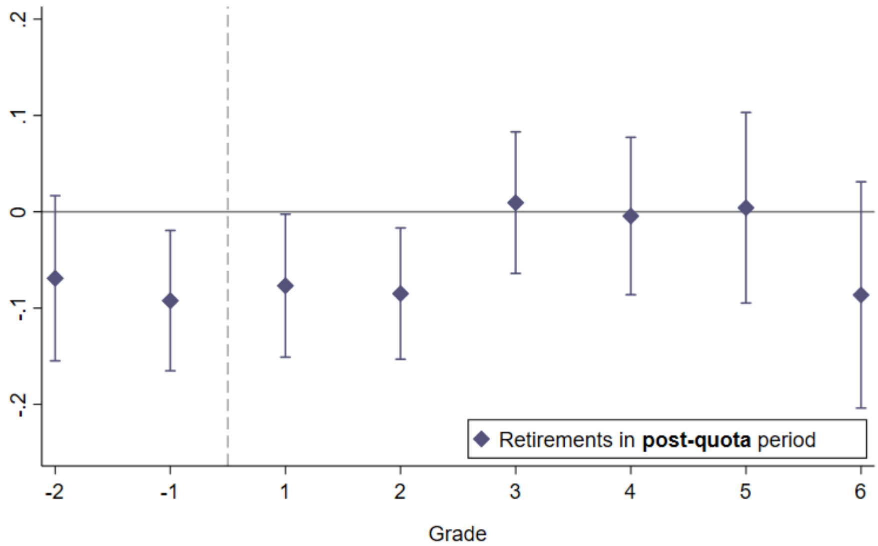
Note: Columns 1-3 show estimates for Equation 10 with the average share male teachers pupils are exposed to during primary school as the outcome. Columns 4-6 show reduced form estimates (corresponding to Equation 10), and Columns 7-9 show IV estimates of Equation 9, with a pupil applying directly in the spring of the year they turn 16 (i.e. the last year of middle school) as the outcome. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. MOP F^{eff} is Olea and Pflueger (2013) effective F-statistic. * p<0.1, ** p<0.05, *** p<0.01 (back)

Figure 7: Reduced Form by Grade: Applications for Post-Compulsory Education

(a) Separate Estimation of Quota and Post-Quota Coefficients



(b) Post-Quota (Relative to Quota Coefficients)



Note: Grade level estimation of pupil level reduced form (c.f. Equation 10). Outcome is binary indicator for pupils applying to post-compulsory education directly after middle school, regressed on the share of teachers turning 60 just before a pupil enters the respective grade in school (Grades 1-6), starting two years prior to a pupil entering school (Grades -2 and -1). Panel (a) estimates absolute coefficients for effect of retirements pupils experience in the quota and the post-quota period. Panel (b) depicts coefficients for the post-quota period *relative* to the quota period (i.e. it shows the difference between quota and post-quota estimates depicted in Panel (a)). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Table 4: IV Estimates: Applications and Enrollment for Post-Compulsory Education

	Apply directly	Apply late	Apply never	Pref. choice	Enrolled at 16	Enrolled ever
Avg Share Male	0.424**	−0.353**	−0.071	0.547**	0.608**	0.124*
	(0.197)	(0.178)	(0.073)	(0.244)	(0.309)	(0.074)
Municipal FE	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X
MOP F^{eff}	15.28	15.28	15.28	15.28	13.00	13.00
Obs	825,094	825,094	825,094	825,094	695,340	695,340
Dep mean	.911	.067	.022	.862	.861	.98
Std effect	.095	-.09	-.031	.101	.11	.055

Note: IV estimates for Equation 9.

Applications: Outcomes in columns 1-3 are mutually exclusive categories of applications to upper secondary education: Pupils apply directly in spring of the year in which they turn 16 (Apply directly), they apply up to four years after they have turned 16 (Apply late), or they apply never or later than five years after having turned 16 (Apply never).

Allocation: (column 4) Pupils obtain one of their first two preferred choices in the application (Pref. choice).

Enrollment: (columns 5-6): Pupils are enrolled in upper secondary education in the fall of the year in which they turn 16 (Enrolled at age 16), and ever enrolled in upper secondary education up to age 25 (Ever enrolled).

Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table 5: IV Estimates: Highest Degree Achieved at Age 25

	Compulsory schooling	Vocational Sec	Vocational Sec Plus	Vocational Tert	Academic Sec	Academic Tert: BA	Academic Tert: MA
Avg Share Male	–0.169	–0.055	0.426**	–0.079	–0.438*	0.386***	–0.070
	(0.154)	(0.260)	(0.208)	(0.211)	(0.228)	(0.146)	(0.093)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.36	15.36	15.36	15.36	15.36	15.36	15.36
Obs	810,065	810,065	810,065	810,065	810,065	810,065	810,065
Dep mean	.127	.316	.108	.146	.211	.054	.038
Std effect	-.032	-.008	.088	-.014	-.068	.108	-.023

Table 6: IV Estimates: Labor Market Attachment at Age 25

	Employed/ Student	Un- employed	DI/ Pension	Other out of LF
Avg Share Male	0.512**	–0.038	–0.124	–0.327**
	(0.243)	(0.153)	(0.076)	(0.137)
Municipal FE	X	X	X	X
Region*Cohort FE	X	X	X	X
Ind. controls	X	X	X	X
MOP F^{eff}	15.37	15.37	15.37	15.37
Obs	811,392	811,392	811,392	811,392
Dep mean	.842	.086	.017	.053
Std effect	.089	-.009	-.061	-.093

Note: IV estimates of Equation 9.

Highest Degree: Outcomes are mutually exclusive categories of pupils' highest degree achieved at age 25, from left to right: Compulsory education only. Vocational track: Basic three year secondary degree (Secondary), additional qualifications or high school coursework beyond a basic degree (Sec. Plus), tertiary degree from a polytechnic (Tertiary). Academic track: Three year high school degree (Secondary), university BA degree (Tert: BA), university MA degree (Tert: MA) or higher.

Labor Market Attachment: Outcomes are mutually exclusive categories of pupils' labor market status measured at age 25: Being in employment or a student, unemployed, on disability insurance (DI) or receiving pension payments, or being out of the labor force for other reasons. This table and all other labor market attachment results at age 25 do not report estimates for the separate category of "conscripts/community service", which contains a total of 1185 observations.

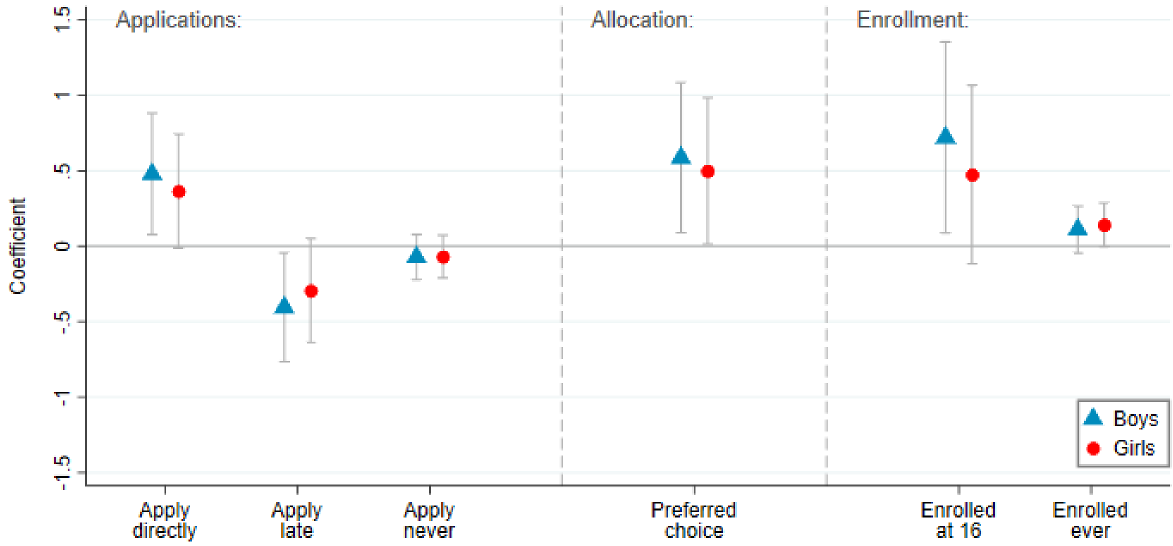
Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table 7: First Stage, Reduced Form and IV: Employed/Student at Age 25

	First Stage			RF			IV		
	Avg Share Male			Employed/Student			Employed/Student		
Total Share 60	0.034	0.034	0.044	0.031	0.032	0.042	0.021	0.019	0.020
	(0.036)	(0.036)	(0.037)	(0.028)	(0.028)	(0.029)	(0.026)	(0.027)	(0.027)
Total Share 60 *	–0.177***	–0.177***	–0.168***	–0.053	–0.071**	–0.086***			
Post-Quota	(0.042)	(0.042)	(0.043)	(0.033)	(0.032)	(0.033)			
Avg Share Male							0.297	0.403*	0.512**
							(0.215)	(0.216)	(0.243)
Municipal FE	X	X	X	X	X	X	X	X	X
Cohort FE	X	X		X	X		X	X	
Region*Cohort FE			X			X			X
Ind. controls		X	X		X	X		X	X
MOP F^{eff}							17.96	17.94	15.37
Adj. R^2	0.916	0.916	0.921	0.008	0.025	0.025			
Obs	811,392	811,392	811,392	811,392	811,392	811,392	811,392	811,392	811,392
Dep mean	.3133	.3133	.3133	.8423	.8423	.8423	.8423	.8423	.8423

Note: Columns 1-3 show estimates for Equation 10 with the average share male teachers pupils are exposed to during primary school as the outcome. Columns 4-6 show reduced form estimates (corresponding to Equation 10), and Columns 7-9 show IV estimates of Equation 9 with being either employed or a student at age 25 as the outcome. Standard errors clustered at the municipality level. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Figure 8: IV Estimates: Applications and Enrollment for Post-Compulsory Education



Note:

IV estimates for Equation 9.

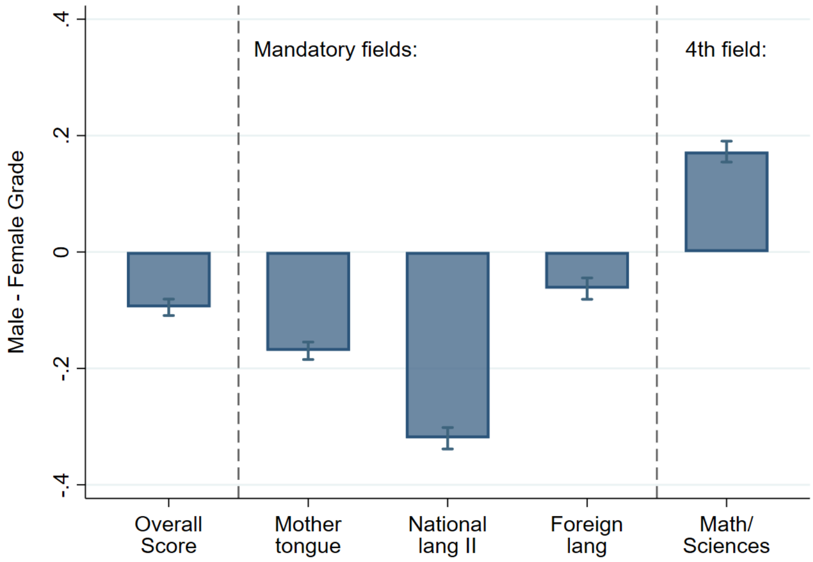
Applications: Outcomes in columns 1-3 are mutually exclusive categories of applications to upper secondary education: Pupils apply directly in spring of the year in which they turn 16 (Apply directly), they apply up to four years after they have turned 16 (Apply late), or they apply never or later than five years after having turned 16 (Apply never).

Allocation: (column 4) Pupils obtain one of their first two preferred choices in the application (Pref. choice).

Enrollment: (columns 5-6): Pupils are enrolled in upper secondary education in the fall of the year in which they turn 16 (Enrolled at age 16), and ever enrolled in upper secondary education up to age 25 (Ever enrolled).

Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Figure 9: Performance Differences by Gender on Matriculation Exam by Field, Full Population



Note: Performance differences in the matriculation exam by gender for the full population of exam takers. Figure displays the gender difference in exam score ("Overall Score") and in each of the exam fields in the matriculation exam that count towards the overall score. Each displayed field receives equal weight of 25% in the calculation of overall score. The score for Math/Sciences is based on the best grade received if an exam taker chose both fields. Sample based on the full population of first time matriculation exam takers in high school track (lukion opiskelija) under age 22 for the years 1983 - 1985. (back)

Table 8: Exit and Hiring Patterns in Municipalities

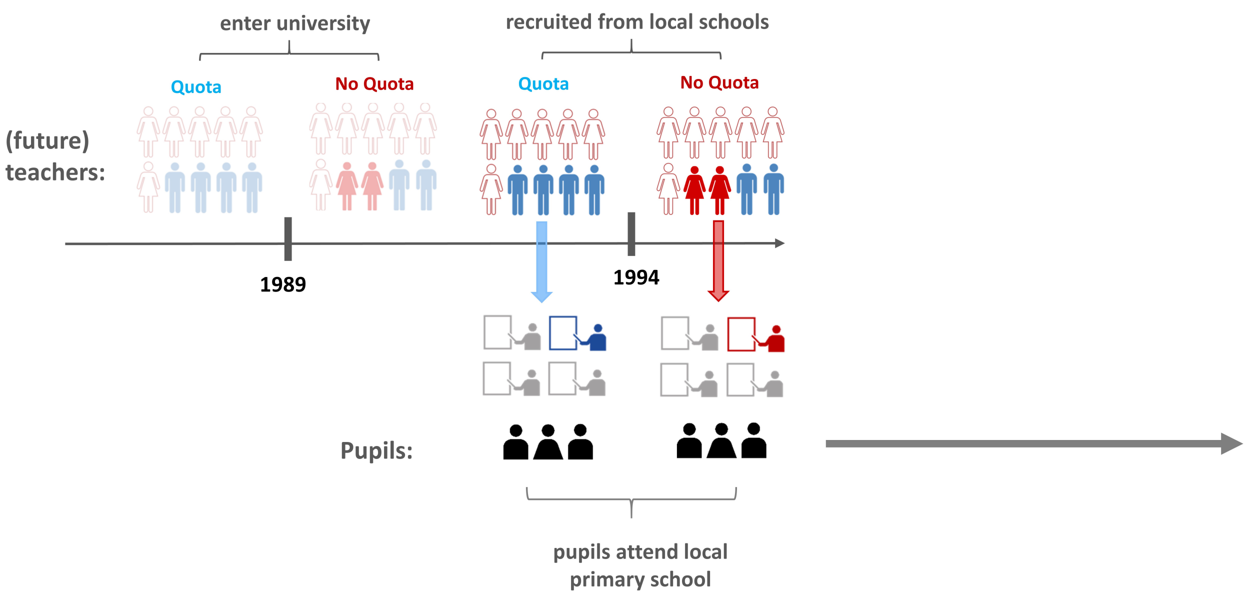
	Leave	Leave ≤ 55	Δ Age	Δ Time since degree	First entrants
Share 60	0.873***	–0.049	–16.088***	–15.599***	0.335
	(0.061)	(0.057)	(1.215)	(1.154)	(0.315)
Share 60 * Post-Quota	0.041	–0.000	–1.300	–1.051	–0.043
	(0.067)	(0.063)	(1.387)	(1.323)	(0.375)
Year FE	X	X	X	X	X
Adj. R^2	0.176	0.011	0.222	0.211	0.038
Obs	4448	4448	4448	4448	3746
Dep mean	.1	.07	-.21	-.2	.35

Estimates for Equation 8. Outcomes from left to right are: Share of teachers exiting, share of teachers below age 55 exiting, year-on-year changes in average age of all local teachers (Δ Age), average time since obtaining a teaching degree of all local teachers (Δ Time since degree). The share of new teacher arriving that are first entrants defined as not having taught before and being below age 28 (column 5). Standard errors clustered at the municipality level. Regressions weighted by population, means unweighted. * p<0.1, ** p<0.05, *** p<0.01 (back)

Appendix Tables and Figures

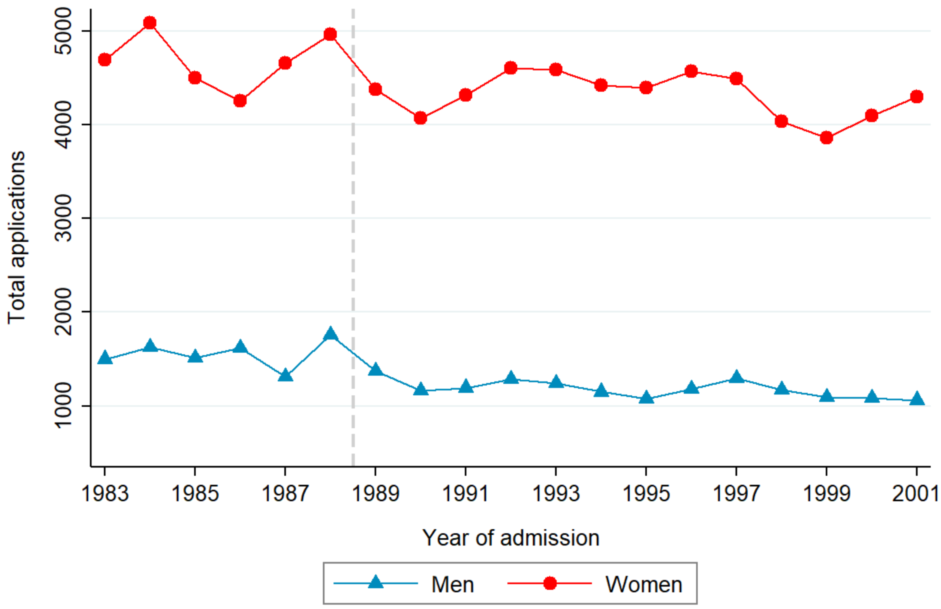
A Reform Context: Timeline, Applications and Graduates

Figure A1: Timeline of the Reform



Note: Future primary school teachers enter university with the quota (pre-1989) and without the quota (1989 and thereafter), and graduate from the five-year primary school teaching degree before 1994 (quota), and thereafter (post-quota). Primary teacher graduates get hired by municipalities to teach in local schools. Pupils will experience differential exposure to quota teachers, described in detail in Section 5 of the paper. (back)

Figure A2: Total Applications by Gender



Note: Total number of male and female applicants to primary teacher studies. *Source:* Liimatainen (2002). (back)

B Teachers' Matriculation Exam Scores

Table A1: First Stage: Teachers' Matriculation Exam Scores

	Overall Score	Language	Best of Math or Sciences	Math	Sciences
Total Share 60	–0.08	–0.05	–0.13	–0.31	–0.02
	(0.22)	(0.24)	(0.26)	(0.45)	(0.30)
Total Share 60 * Post-Quota	0.08	0.05	0.15	0.10	0.11
	(0.25)	(0.27)	(0.29)	(0.51)	(0.33)
Municipal * Post-Quota FE	X	X	X	X	X
Year FE	X	X	X	X	X
Adj. R^2	0.84	0.85	0.82	0.81	0.82
Obs	4311	4311	4311	4300	4301
Dep mean	4.5	4.41	4.78	4.34	4.52

Note: Estimates for Equation 7 with average of local teachers' grades in first attempt of matriculation exam as the outcome. Overall score is the average grade obtained across the four mandatory exam fields. Language score includes grades for the three mandatory language fields in column 2 (mother tongue (FI/SE), second national language (SE/FI) and foreign language). The fourth mandatory field consists in a candidate's choice of either Mathematics or Sciences. The best score of Math or Sciences (column 3) is considered in the calculation of overall score if candidates took both fields. Math (column 4) and Sciences (column 5) report grades irrespective of whether they are counted towards overall score. Grades are assigned on a curve and range from 2-6, with 6 being the highest grade and 0 assigned for a failing exam (Kupiainen et al., 2018. For more detail regarding the matriculation exam fields and overall scores, see detailed discussion in Section 8.2.1. Data is available for teacher cohorts born after 1952. Sample is restricted to municipalities with at least one teacher with observed score in any year. Standard errors clustered at the municipality level. Regressions weighted by population. * p<0.1, ** p<0.05, *** p<0.01 (back)

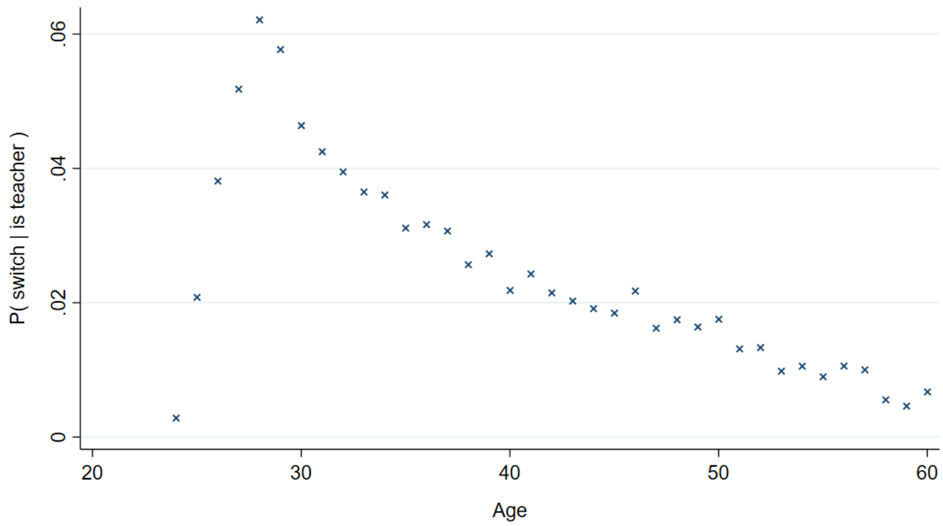
Table A2: First Stage in Teacher Score Sample

Total Share 60	0.066
	(0.043)
Total Share 60 * Post-Quota	–0.224***
	(0.050)
Municipal * Post-Quota FE	X
Year FE	X

Note: Estimates for Equation 7 for restricted sample of municipalities with at least one teacher test score observable in any year. Standard errors clustered at the municipality level. Regression weighted by population. * p<0.1, ** p<0.05, *** p<0.01 (back)

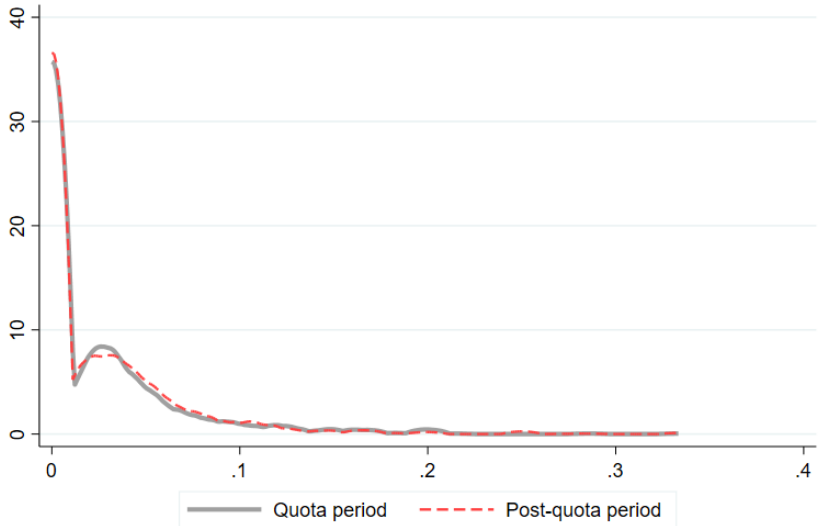
C Teacher Switching and Retirement

Figure A3: Probability of Switching Municipality of Work for Active Teachers by Age



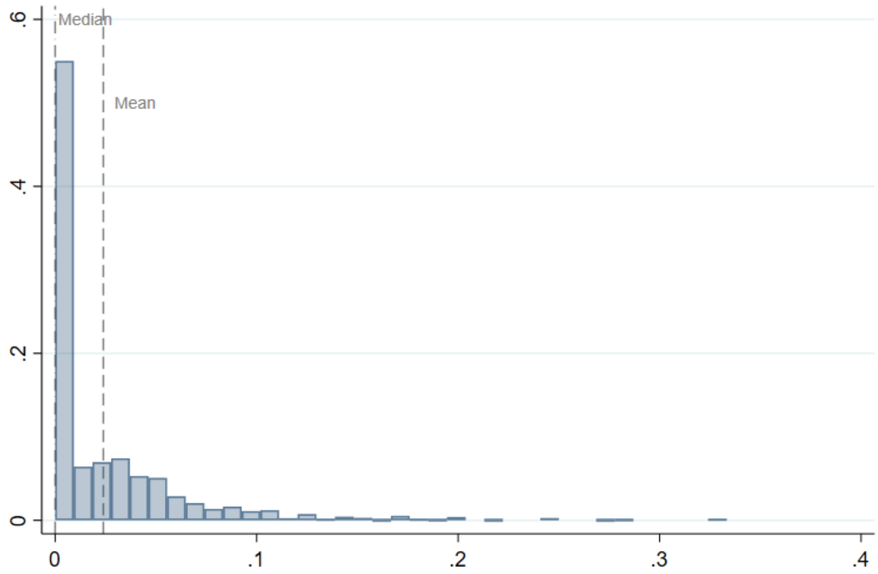
Note: Share of primary school teachers having switched the municipality in which they are working as a primary school teacher at a given age, conditional on having worked as a primary school teacher in the previous year. Data for all active primary school teachers in the years 1990-2000. (back)

Figure A4: Distribution of Share Primary Teachers Turning 60



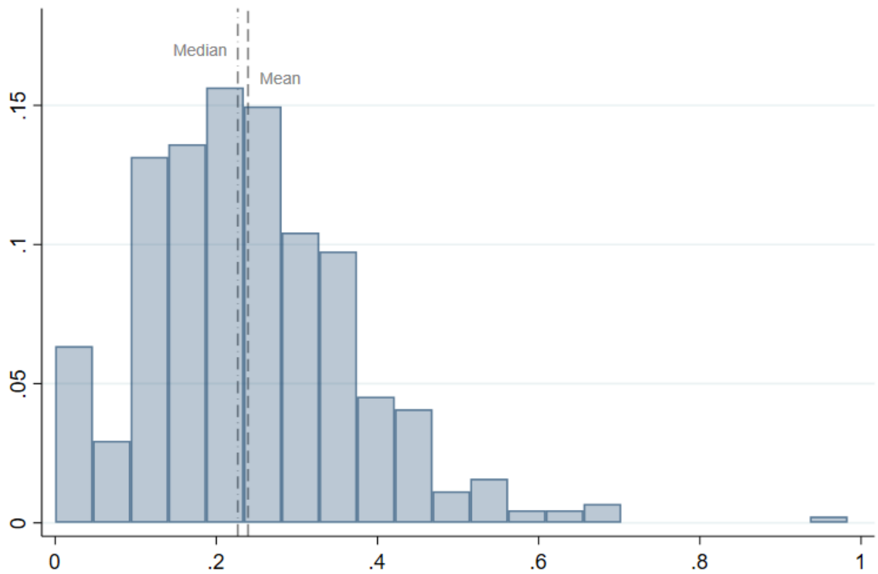
Note: Smoothed density of yearly municipal share of primary school teachers turning 60, separately by years in the quota period (1991-93) and post-quota period (1994-2000). (back)

Figure A5: Distribution of Share Primary Teachers Turning 60 (yearly)



Note: Histogram of yearly municipal share of primary school teachers turning 60. (back)

Figure A6: Distribution of Total Share Primary Teachers Turning 60, 1991-2000



Note: Histogram of cumulative municipal share of primary school teachers turning 60 (adding up all retirements within municipality from 1991-2000). (back)

D Additional IV Estimates

D.1 Sensitivity Main Results

Table A3: IV Estimates: Applications and Enrollment for Post-Compulsory Education, Municipal and Cohort Fixed Effects only

	Apply directly	Apply late	Apply never	Pref. choice	Enrolled at 16	Enrolled ever
Avg share male	0.696***	–0.553***	–0.143*	0.723***	1.071***	0.212**
	(0.234)	(0.193)	(0.078)	(0.276)	(0.384)	(0.087)
Municipal FE	X	X	X	X	X	X
Cohort FE	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X
MOP F^{eff}	17.64	17.64	17.64	17.64	14.40	14.40
Obs	825,094	825,094	825,094	825,094	695,340	695,340
Dep mean	.911	.067	.022	.862	.861	.98
Std effect	.156	-.141	-.061	.133	.193	.094

Note: IV estimates for Equation 9.

Applications: Outcomes in columns 1-3 are mutually exclusive categories of applications to upper secondary education: Pupils apply directly in spring of the year in which they turn 16 (Apply directly), they apply up to four years after they have turned 16 (Apply late), or they apply never or later than five years after having turned 16 (Apply never).

Allocation: (column 4) Pupils obtain one of their first two preferred choices in the application (Pref. choice).

Enrollment: (columns 5-6): Pupils are enrolled in upper secondary education in the fall of the year in which they turn 16 (Enrolled at age 16), and ever enrolled in upper secondary education up to age 25 (Ever enrolled).

Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

D.2 Additional Results on Applications:

Table A4: Obtained Choices in Post-Compulsory Applications

	First	Second	Third	Fourth	Fifth	Switch	No Spot	Apply Never
Avg Share Male	0.334	0.212	−0.137	−0.103	0.021	−0.112	−0.145	−0.071
	(0.287)	(0.161)	(0.096)	(0.065)	(0.040)	(0.091)	(0.098)	(0.073)
Municipal FE	X	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X	X
MOP F^{eff}	15.28	15.28	15.28	15.28	15.28	15.28	15.28	15.28
Obs	825,094	825,094	825,094	825,094	825,094	825,094	825,094	825,094
Dep mean	.777	.085	.035	.015	.007	.019	.04	.022
Std effect	.051	.049	-.048	-.054	.016	-.052	-.047	-.031

Note: IV estimates of Equation 9. Outcomes are mutually exclusive categories for allocation of slots in post-compulsory education application, from left to right: Pupils obtain their First, ..., Fifth choice. Pupils switch from assigned slot to other option (Switch), do not obtain any slot at all (No Spot), and do not put in an application within five years after middle school (Never Apply). Individual level controls are measured at age 7 and include language (SE/FI), foreign origin, single parent HH, highest degree attained in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A5: By Gender: Aspirations for Post-Compulsory Education

	Apply	Choose:		Obtain:		
	never	any Voc	only Acad	no spot	Voc	Acad
Boys * Avg	–0.073	0.811**	–0.744**	–0.207**	0.263	0.017
Share Male	(0.076)	(0.381)	(0.375)	(0.104)	(0.324)	(0.350)
Girls * Avg	–0.070	–0.367	0.431	–0.073	–0.820**	0.962***
Share Male	(0.072)	(0.402)	(0.395)	(0.099)	(0.348)	(0.371)
Municipal FE	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X
MOP F^{eff}	15.28	15.28	15.28	15.28	15.28	15.28
Obs	825,094	825,094	825,094	825,094	825,094	825,094
Boys: Dep mean	.025	.629	.346	.041	.501	.433
Girls: Dep mean	.02	.443	.537	.039	.329	.612
Boys: Std effect	-.03	.107	-.1	-.067	.034	.002
Girls: Std effect	-.032	-.047	.055	-.024	-.111	.126

Note: IV estimates for Equation 9.

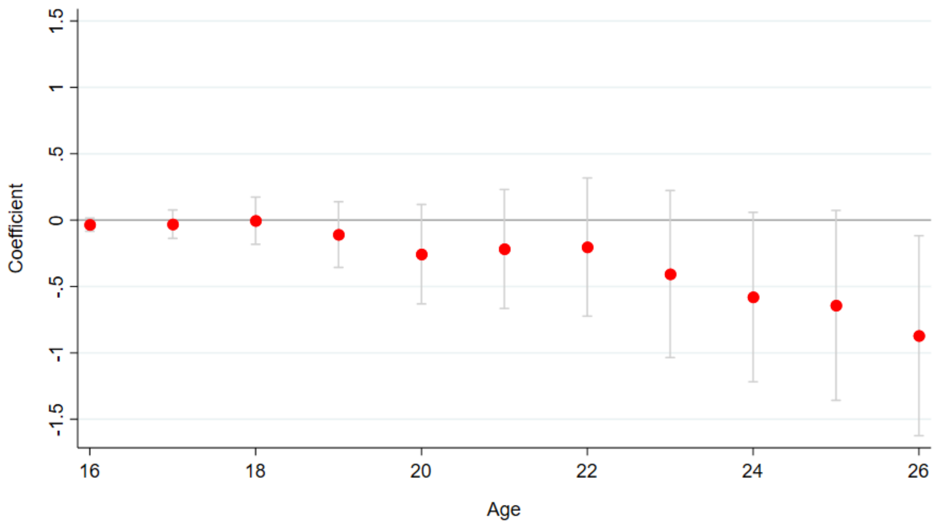
Choices: Outcomes are mutually exclusive categories for columns 1-3: Pupils 'Apply never', pupils put in a vocational degree in any of five available choices (Choose any Voc), or pupils put in only academic track choices (Choose only Acad) (We don't report an estimate for the separate category of 287 pupils who never put in a choice, but obtain a study slot nevertheless).

Slots obtained: Columns 1 and 4-6 are also mutually exclusive categories: Pupils 'Apply never', pupils get allocated 'no spot', a spot in a vocational track (Voc), or a spot in the academic track (Acad).

Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

D.3 Fertility

Figure A7: Female Pupils: Probability of First Birth Having Occurred by Age



Note: IV estimates of Equation 9. Outcome: First birth having occurred by age. Individual level controls are measured at age 7 and include language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Table A6: Female Pupils: Probability of First Birth Having Occurred by Age

	16	17	18	19	20	21	22	23	24	25	26
Avg Share Male	−0.034	−0.031	−0.004	−0.109	−0.257	−0.217	−0.203	−0.406	−0.579*	−0.642*	−0.871**
	(0.026)	(0.055)	(0.091)	(0.126)	(0.191)	(0.229)	(0.266)	(0.321)	(0.326)	(0.365)	(0.384)
Municipal FE	X	X	X	X	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X	X	X	X	X
MOP F^{eff}	15.75	15.75	15.75	15.75	15.75	15.75	15.75	15.75	15.75	15.75	15.75
Obs	396,108	396,108	396,108	396,108	396,108	396,108	396,108	396,108	396,108	396,108	396,108
Dep mean	.001	.005	.013	.03	.056	.087	.119	.152	.188	.229	.273
Std effect	-.059	-.028	-.002	-.041	-.071	-.049	-.04	-.072	-.094	-.097	-.124

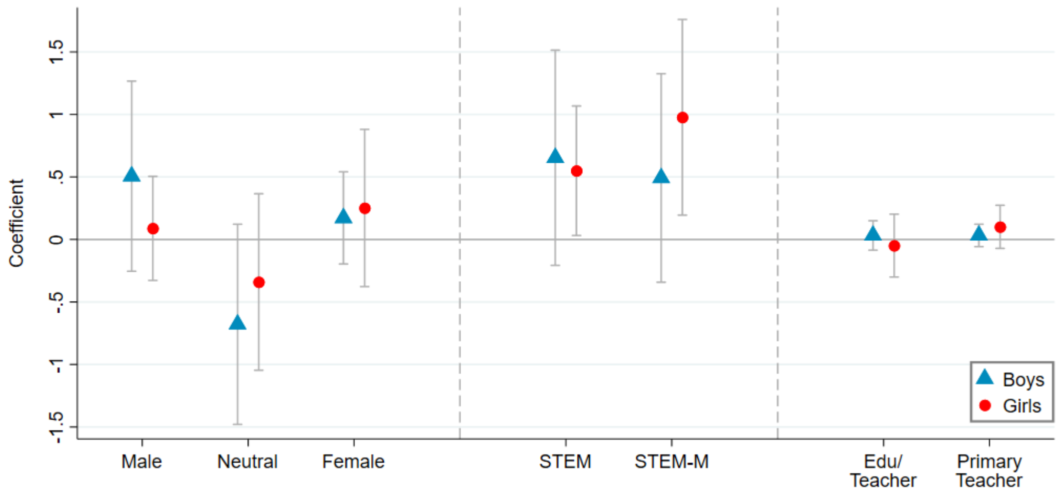
Note: IV estimates for Equation 9. Outcome is the likelihood of having had the first birth by age, from 16 to 26. Individual level controls are measured at age 7 and include language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

D.4 Field of Study

Table A7: IV Estimates: Field of Education at Age 25

	Male	Neutral	Female	STEM	STEM-M	Education/ Teacher	Primary Teacher
Avg Share Male	0.302	–0.500*	0.197	0.595**	0.707**	–0.013	0.063
	(0.229)	(0.286)	(0.191)	(0.273)	(0.323)	(0.073)	(0.049)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.37	15.37	15.37	15.37	15.37	15.37	15.37
Obs	811,392	811,392	811,392	811,392	811,392	811,392	811,392
Dep mean	.303	.433	.264	.264	.379	.023	.011
Std effect	.042	-.064	.028	.086	.093	-.006	.039

Figure A8: IV Estimates by Gender: Field of Education at Age 25



Note: IV estimates of Equation 9. Outcomes from left to right:
Field is 'Male' dominated ($\geq$ 60% male), (gender) 'Neutral' or 'Female' dominated ($\geq$ 60% female).
Field is STEM or STEM + Medicine (STEM-M).
Field is Education Science or Teacher. Field is Primary School Teacher.
Lower panel reports regressions separate by pupil gender. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Table A8: Boys: Field of Education at Age 25

	Male	Neutral	Female	STEM	STEM-M	Education/ Teacher	Primary Teacher
Boys: Avg Share Male	0.506	–0.679*	0.172	0.653	0.492	0.032	0.032
	(0.388)	(0.408)	(0.188)	(0.439)	(0.425)	(0.060)	(0.046)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	14.91	14.91	14.91	14.91	14.91	14.91	14.91
Obs	415,571	415,571	415,571	415,571	415,571	415,571	415,571
Dep mean	.526	.39	.084	.412	.446	.009	.005
Std effect	.065	-.089	.04	.085	.063	.022	.031

Table A9: Girls: Field of Education at Age 25

	Male	Neutral	Female	STEM	STEM-M	Education/ Teacher	Primary Teacher
Girls: Avg Share Male	0.089	–0.340	0.252	0.550**	0.977**	–0.049	0.101
	(0.212)	(0.360)	(0.321)	(0.264)	(0.399)	(0.128)	(0.088)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.75	15.75	15.75	15.75	15.75	15.75	15.75
Obs	395,821	395,821	395,821	395,821	395,821	395,821	395,821
Dep mean	.069	.478	.453	.108	.309	.038	.017
Std effect	.022	-.043	.032	.113	.134	-.016	.05

Note: IV estimates for Equation 9, separate regressions by gender. Outcomes from left to right: Field is 'Male' dominated ($\geq$ 60% male), (gender) 'Neutral' or 'Female' dominated ($\geq$ 60% female), based on previous generation. Field is STEM or STEM + Medicine (STEM-M). Field is Education Science or Teacher. Field is Primary School Teacher. Individual level controls are measured at age 7 and include language (SE/FI), foreign origin, single parent HH, highest degree attained in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

E IV Estimates by Pupil Gender

E.1 IV Estimates by Pupil Gender

Table A10: By Gender: Applications and Enrollment for Post-Compulsory Education

	Apply directly	Apply late	Apply never	Pref. choice	Enrolled at 16	Enrolled ever
Boys * Avg	0.478**	–0.405**	–0.073	0.585**	0.720**	0.110
Share Male	(0.205)	(0.184)	(0.076)	(0.254)	(0.323)	(0.079)
Girls * Avg	0.364*	–0.294*	–0.070	0.498**	0.474	0.141*
Share Male	(0.193)	(0.176)	(0.072)	(0.247)	(0.301)	(0.074)
Municipal FE	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X
MOP F^{eff}	15.28	15.28	15.28	15.28	13.00	13.00
Obs	825,094	825,094	825,094	825,094	695,340	695,340
Boys: Dep mean	.889	.086	.025	.857	.845	.977
Girls: Dep mean	.933	.047	.02	.867	.876	.982
Boys: Std effect	.097	-.092	-.03	.106	.124	.046
Girls: Std effect	.093	-.088	-.032	.093	.09	.066
P-value	.096	.067	.922	.478	.057	.465

Note: IV estimates for Equation 9. Outcomes in columns 1-3 are mutually exclusive categories of applications to upper secondary education: Pupils apply directly in spring of the year in which they turn 16 (Apply directly), they apply up to four years after they have turned 16 (Apply late), or they apply never or later than five years after having turned 16 (Apply never). "Allocation" (column 4): Pupils obtain one of their first two preferred choices in the application (Pref. choice). "Enrollment" (columns 5-6): Pupils are enrolled in upper secondary education in the fall of the year in which they turn 16 (Enrolled at age 16), and ever enrolled in upper secondary education up to age 25 (Ever enrolled). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A11: By Gender: Highest Degree Achieved at Age 25

	Compulsory schooling	Vocational Sec	Vocational Sec Plus	Vocational Tert	Academic Sec	Academic Tert: BA	Academic Tert: MA
Boys * Avg	0.242	–0.194	0.458**	–0.262	–0.432*	0.317**	–0.128
Share Male	(0.176)	(0.277)	(0.216)	(0.217)	(0.232)	(0.143)	(0.091)
Girls * Avg	–0.632***	0.092	0.391*	0.133	–0.446*	0.465***	–0.003
Share Male	(0.185)	(0.274)	(0.208)	(0.219)	(0.236)	(0.156)	(0.101)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.36	15.36	15.36	15.36	15.36	15.36	15.36
Obs	810,065	810,065	810,065	810,065	810,065	810,065	810,065
Boys: Dep mean	.152	.378	.081	.094	.231	.042	.022
Girls: Dep mean	.101	.251	.136	.201	.19	.067	.054
Boys: Std effect	.043	-.025	.107	-.057	-.065	.101	-.056
Girls: Std effect	-.133	.013	.073	.021	-.072	.118	-.001
P-value	.000	.11	.428	.000	.9	.019	.011

Note: IV estimates for Equation 9. Outcomes are mutually exclusive categories of pupils' highest degree achieved at age 25: Having only Compulsory education. For the Vocational track: Having a basic three year secondary degree (Sec), having additional qualifications or high school coursework beyond a basic degree (Sec Plus), having a tertiary degree from a polytechnic (Tert). For the Academic track: Having a three year high school degree (Sec), having a three year university BA degree (Tert: BA), having a two year university MA degree (Tert: MA) or higher. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A12: By Gender: Labor Market Outcomes at Age 25

	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Boys * Avg	0.685***	–0.007	–0.145*	–0.526***
Share Male	(0.254)	(0.161)	(0.078)	(0.143)
Girls * Avg	0.320	–0.079	–0.102	–0.099
Share Male	(0.243)	(0.156)	(0.077)	(0.139)
Municipal FE	X	X	X	X
Region*Cohort FE	X	X	X	X
Ind. controls	X	X	X	X
MOP F^{eff}	15.37	15.37	15.37	15.37
Obs	811,392	811,392	811,392	811,392
Boys: Dep mean	.84	.102	.019	.037
Girls: Dep mean	.845	.07	.015	.07
Boys: Std effect	.119	-.001	-.067	-.179
Girls: Std effect	.056	-.02	-.053	-.025
P-value	.002	.4	.159	.000

Note: IV estimates for Equation 9. Outcomes are mutually exclusive categories of pupils' labor market status measured at age 25: Being in employment or a student, unemployed, on disability insurance (DI) or receiving pension payments, or being out of the labor force for other reasons. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

E.2 IV Estimates by Pupil Gender: Split Sample Estimates

Table A13: Applications and Labor Market Attachment

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Boys * Avg	0.341	–0.351	0.010	0.271	–0.045	–0.079	–0.123
Share Male	(0.260)	(0.243)	(0.104)	(0.315)	(0.237)	(0.099)	(0.132)
Girls * Avg	0.511**	–0.353*	–0.158*	0.731**	–0.012	–0.171*	–0.528**
Share Male	(0.218)	(0.182)	(0.087)	(0.290)	(0.172)	(0.097)	(0.218)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.28	15.28	15.28	15.37	15.37	15.37	15.37
Obs	825,094	825,094	825,094	811,392	811,392	811,392	811,392
Boys: Dep mean	.889	.086	.025	.84	.102	.019	.037
Girls: Dep mean	.933	.047	.02	.845	.07	.015	.07
P-value	.533	.993	.18	.219	.908	.464	.086

Table A14: Highest Degree Achieved at Age 25

	Compulsory	Vocational			Academic		
	schooling	Sec	Sec Plus	Tert	Sec	Tert: BA	Tert: MA
Boys * Avg	0.185	0.052	0.210	–0.088	–0.819**	0.534***	–0.074
Share Male	(0.234)	(0.385)	(0.223)	(0.231)	(0.344)	(0.193)	(0.092)
Girls * Avg	–0.511**	–0.192	0.639**	–0.083	–0.026	0.241	–0.068
Share Male	(0.233)	(0.329)	(0.300)	(0.304)	(0.238)	(0.173)	(0.154)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	15.36	15.36	15.36	15.36	15.36	15.36	15.36
Obs	810,065	810,065	810,065	810,065	810,065	810,065	810,065
Boys: Dep mean	.152	.378	.081	.094	.231	.042	.022
Girls: Dep mean	.101	.251	.136	.201	.19	.067	.054
P-value	.046	.62	.194	.989	.035	.188	.971

Note: IV estimates for Equation 9, split sample estimates by gender. Outcomes are pupils' applications in the last year of middle school (see Table 4) and labor market status (see Table 6). Individual level controls are measured at age 7 and include language (SE/FI), foreign origin, single parent HH, highest degree attained in HH. Standard errors clustered at the municipality level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$ (back)

F Reduced Form Estimates: Main Results

F.1 Reduced Form:

Table A15: Reduced Form: Applications and Enrollment for Post-Compulsory Education

	Apply directly	Apply late	Apply never	Pref. choice	Enrolled at 16	Enrolled ever
Total Share 60	0.018	–0.021	0.003	–0.010	0.026	0.005
	(0.021)	(0.019)	(0.009)	(0.027)	(0.028)	(0.008)
Total Share 60 *	–0.071***	0.059**	0.012	–0.092***	–0.089**	–0.018*
Post-Quota	(0.027)	(0.025)	(0.012)	(0.035)	(0.035)	(0.010)
Municipal FE	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X
Obs	825,094	825,094	825,094	825,094	695,340	695,340
Dep mean	.911	.067	.022	.862	.861	.98

Note: Reduced Form estimates as in Equation 10. Outcomes in columns 1-3 are mutually exclusive categories of applications to upper secondary education: Pupils apply directly in spring of the year in which they turn 16 (Apply directly), they apply up to four years after they have turned 16 (Apply late), or they apply never or later than five years after having turned 16 (Apply never). “Allocation” (column 4): Pupils obtain one of their first two preferred choices in the application (Pref. choice). “Enrollment” (columns 5-6): Pupils are enrolled in upper secondary education in the fall of the year in which they turn 16 (Enrolled at age 16), and ever enrolled in upper secondary education up to age 25 (Ever enrolled). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A16: Reduced Form: Highest Degree Achieved at Age 25

	Compulsory	Vocational			Academic		
	schooling	Sec	Sec Plus	Tert	Sec	Tert: BA	Tert: MA
Total Share 60	–0.007	–0.005	0.029	–0.005	–0.034	0.027	–0.005
	(0.019)	(0.036)	(0.023)	(0.028)	(0.027)	(0.016)	(0.012)
Total Share 60 *	0.029	0.009	–0.072**	0.013	0.074**	–0.065***	0.012
Post-Quota	(0.024)	(0.044)	(0.029)	(0.036)	(0.034)	(0.022)	(0.015)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
Obs	810,065	810,065	810,065	810,065	810,065	810,065	810,065
Dep mean	.127	.316	.108	.146	.211	.054	.038

Note: Reduced Form estimates as in Equation 10. Outcomes are mutually exclusive categories of pupils' highest degree achieved at age 25: Having only Compulsory education. For the Vocational track: Having a basic three year secondary degree (Sec), having additional qualifications or high school coursework beyond a basic degree (Sec Plus), having a tertiary degree from a polytechnic (Tert). For the Academic track: Having a three year high school degree (Sec), having a three year university BA degree (Tert: BA), having a two year university MA degree (Tert: MA) or higher. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

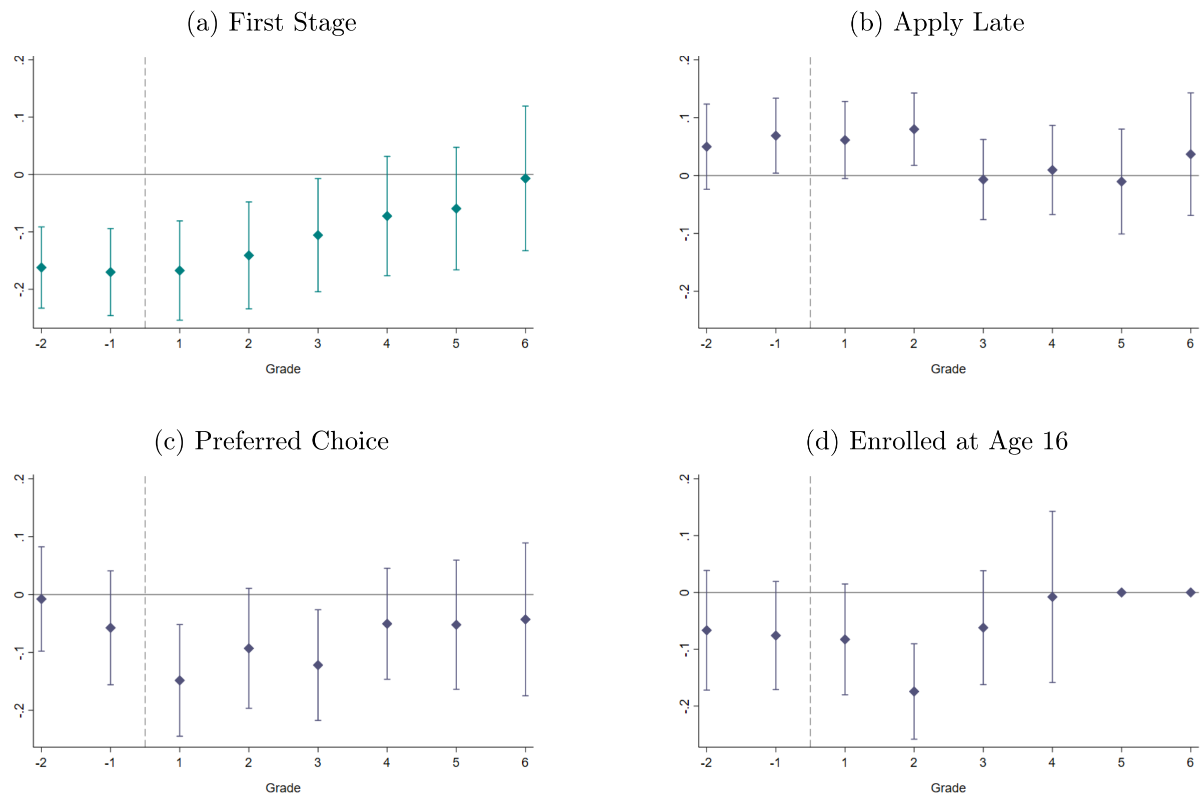
Table A17: Reduced Form: Labor Market Outcomes at Age 25

	Employed/ Student	Un- employed	DI/ Pension	Other out of LF
Total Share 60	0.042	–0.001	–0.005	–0.034**
	(0.029)	(0.022)	(0.009)	(0.015)
Total Share 60 *	–0.086***	0.006	0.021*	0.055***
Post-Quota	(0.033)	(0.026)	(0.011)	(0.018)
Municipal FE	X	X	X	X
Region*Cohort FE	X	X	X	X
Ind. controls	X	X	X	X
Obs	811,392	811,392	811,392	811,392
Dep mean	.842	.086	.017	.053

Note: Reduced Form estimates as in Equation 10. Outcomes are mutually exclusive categories of pupils' labor market status measured at age 25: Being in employment or a student, unemployed, on disability insurance (DI) or receiving pension payments, or being out of the labor force for other reasons. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

F.2 Reduced Form: Grade-Level for Selected Outcomes

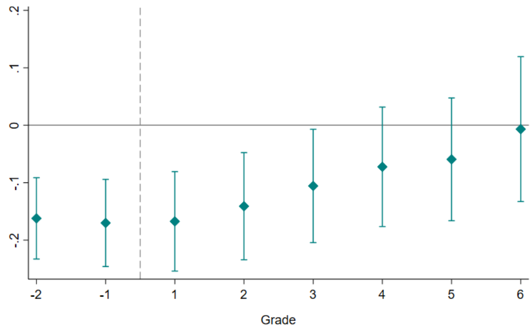
Figure A9: Intermediate Outcomes: Grade Level Estimation



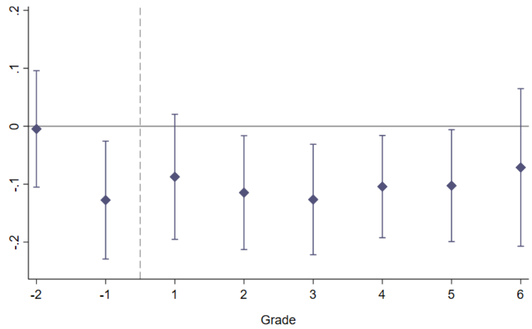
Note: Grade level estimation of pupil level first stage and reduced form (see Equation 10). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Figure A10: Long-Term Outcomes: Grade Level Estimation

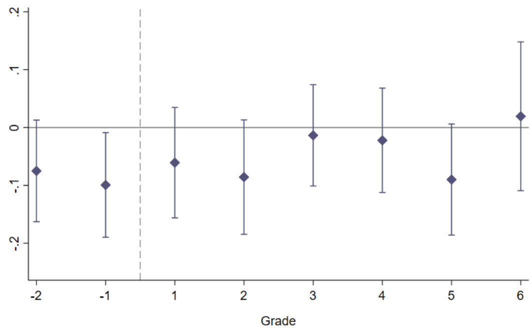
(a) First Stage



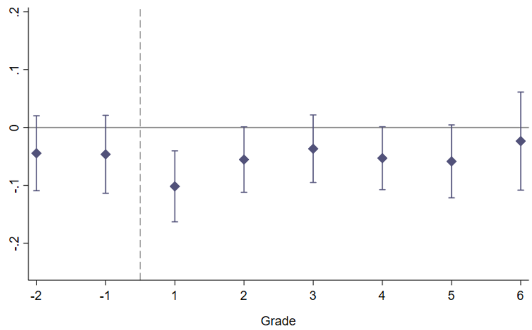
(b) Employed/Student



(c) Vocational Plus Degree



(d) BA Degree



Note: Grade level estimation of pupil level reduced form (see Equation 10). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

G Mechanism

G.1 Complementarities

Table A18: Reduced Form: Complementarities in Production

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Low: Total share 60	0.013	−0.013	0.000	0.054	−0.010	−0.009	−0.030
	(0.033)	(0.028)	(0.014)	(0.043)	(0.033)	(0.014)	(0.022)
High: Total share 60	0.037	−0.034	−0.003	0.026	0.008	−0.007	−0.030
	(0.027)	(0.024)	(0.012)	(0.040)	(0.031)	(0.012)	(0.022)
Low: Total share 60 * post	−0.090**	0.065*	0.026	−0.086*	0.019	0.023	0.036
	(0.038)	(0.033)	(0.017)	(0.049)	(0.038)	(0.017)	(0.028)
High: Total share 60 * post	−0.063*	0.056*	0.006	−0.081*	−0.001	0.025	0.057**
	(0.036)	(0.033)	(0.017)	(0.045)	(0.037)	(0.016)	(0.024)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
Obs	819,831	819,831	819,831	806,226	806,226	806,226	806,226
Low: Dep mean	.904	.07	.026	.846	.082	.017	.053
High: Dep mean	.928	.058	.014	.832	.097	.018	.052
P-value	.594	.858	.409	.943	.695	.938	.579

Note: Reduced form estimates for Equation 10 by median initial share male teachers in a municipality in 1990. P-values reported for test of equality of coefficients in the post-quota period. Outcomes are pupils' application choices and labor market status (see Tables 4 and 6). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A19: Complementarities: Low Share Male (IV estimates)

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Low: Avg Share Male	0.775 (0.589)	–0.547 (0.464)	–0.228 (0.195)	0.752 (0.599)	–0.149 (0.342)	–0.199 (0.197)	–0.344 (0.294)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	3.69	3.69	3.69	3.78	3.78	3.78	3.78
Obs	590,156	590,156	590,156	579,101	579,101	579,101	579,101
Dep mean	.904	.07	.026	.846	.082	.017	.053
Std effect	.126	-.102	-.069	.099	-.026	-.073	-.073

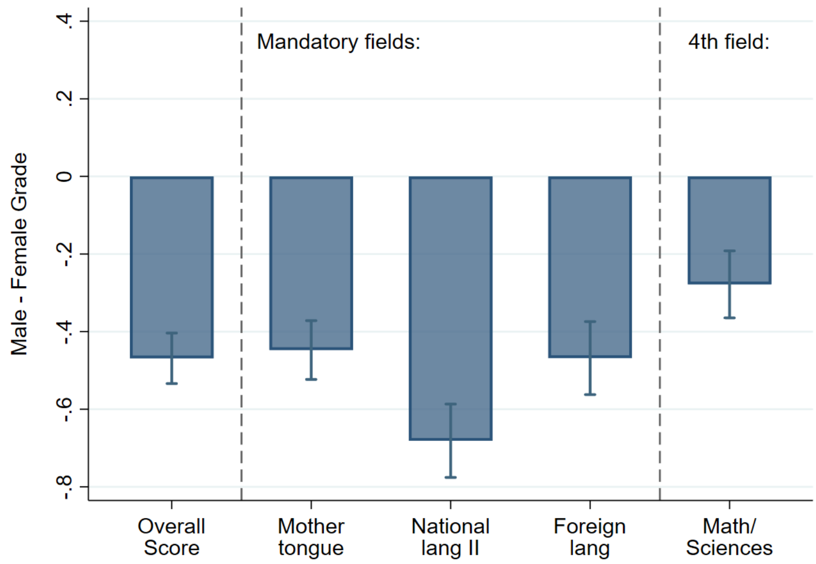
Table A20: Complementarities: High Share Male (IV estimates)

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
High: Avg Share Male	0.547 (0.387)	–0.495 (0.360)	–0.052 (0.138)	0.663 (0.488)	0.002 (0.301)	–0.208 (0.161)	–0.453 (0.280)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
MOP F^{eff}	5.04	5.04	5.04	5.00	5.00	5.00	5.00
Obs	229,342	229,342	229,342	226,794	226,794	226,794	226,794
Dep mean	.928	.058	.014	.832	.097	.018	.052
Std effect	.117	-.118	-.024	.098	0	-.088	-.113

Note: IV estimates for Equation 9, split sample by initial share male teachers in a municipality in 1990. Outcomes are pupils' application choices and labor market status (see Tables 4 and 6). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

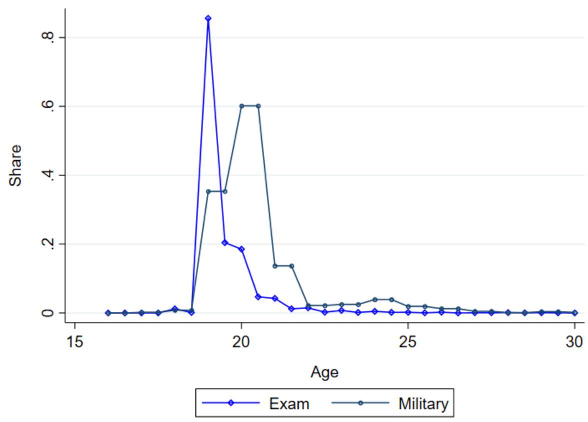
G.2 Performance Differences in Matriculation Exam

Figure A11: Performance Differences by Gender on Matriculation Exam by Field, Teachers



Note: Performance differences in the matriculation exam by gender for the full population of exam takers. Figure displays the gender difference in exam score ("Overall Score") and in each of the exam fields in the matriculation exam that count towards the overall score. Each displayed field receives equal weight of 25% in the calculation of overall score. The score for Math/Sciences is based on the best grade received if an exam taker chose both fields. Sample based on the full population of primary teachers in the main analysis sample who are first time matriculation exam takers in the high school track (lukion opiskelija) under age 22 for the years 1983 - 1985, to ensure comparability with Figure 9. (back)

Figure A12: Timing of Matriculation Exam and Military Service, Male Teachers



Note: Share of male teachers taking the matriculation exam and being listed as a military recruit by age and season of the matriculation exam (spring/fall). Data on matriculation exams takes are bi-yearly, while military service is available once per year. (back)

G.3 Teacher team performance and exam scores

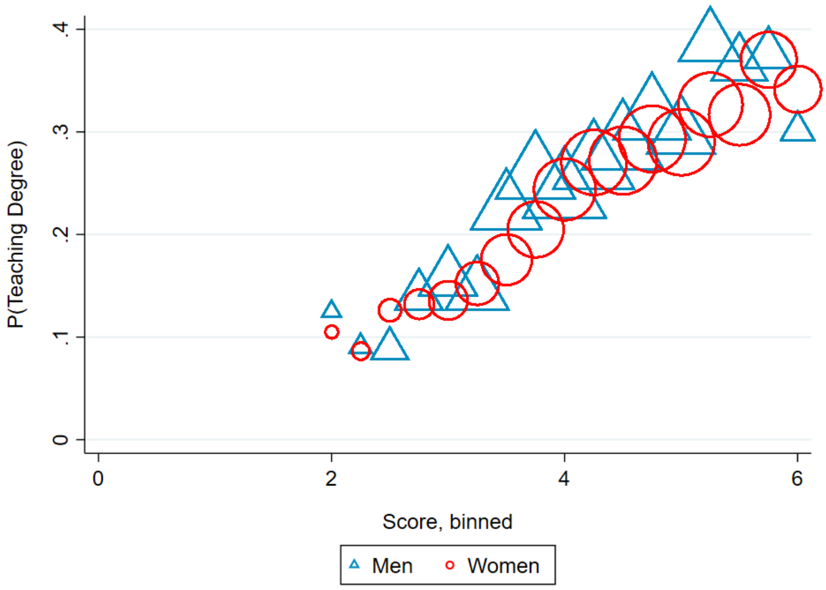
Table A21: Teacher Team Performance and Teacher Exam Scores

	Apply				Employed/Student			
Overall Exam Score	–0.0039				0.0011			
	(0.0062)				(0.0082)			
Language		–0.0074	–0.0085			–0.0055	–0.0058	
		(0.0064)	(0.0068)			(0.0080)	(0.0084)	
Math or Science		0.0045				0.0083		
		(0.0053)				(0.0060)		
Math			0.0026				0.0004	
			(0.0033)				(0.0035)	
Science			0.0029				0.0066	
			(0.0052)				(0.0057)	
Math Background				0.0015				0.0247
				(0.0153)				(0.0157)
Municipal FE	X	X	X	X	X	X	X	X
Cohort FE	X	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X	X
Adj. R^2	0.070	0.070	0.070	0.070	0.025	0.025	0.025	0.025
Obs	825,032	825,032	825,032	825,032	811,331	811,331	811,331	811,331
Dep mean	.911	.911	.911	.911	.842	.842	.842	.842

Note: Estimates for Equation 9 of pupil outcomes on teachers' grades in the matriculation exam. The score for Language is comprised of the grade for Mother Tongue, Second National Language and Foreign Language. Math or Science is based on the best grade received if an exam taker chose both fields. Math background measures the share of teachers who have taken mathematics in their matriculation exam. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

G.4 Compensating Differentials

Figure A13: Likelihood of Obtaining Primary Teaching Degree by Exam Score and Gender



Note: Share of primary school teaching degree holders among all applicants to primary teaching by binned score in the matriculation exam. Weighted by number of applicants per bin. Sample contains all applicants to primary teacher education in the years 2000 - 2005, with data for 2003 missing and considers all degrees obtained by the year 2020. The exam score is based on a candidate's best performance if they had multiple takes. Exam score is assigned on a scale between 2-6 with 6 corresponding to the highest grade. (back)

Table A22: Likelihood of Teaching Degree by Score and Gender, Among Applicants to Primary Teaching

	(1)	(2)	(3)	(4)
Score	0.073***	0.073***	0.068***	0.069***
	(0.003)	(0.004)	(0.004)	(0.004)
Male		–0.003		–0.010
		(0.036)		(0.043)
Male * Score		0.007		0.013
		(0.008)		(0.010)
Adj. R^2	0.100	0.100	0.088	0.089
Obs	20,563	20,563	18,243	18,243
Dep mean	.266	.266	.3	.3

Note: Outcome is an indicator variable for whether an applicant holds a primary teaching degree in the year 2020. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers all degrees obtained by the year 2020. In columns 3 and 4, sample is restricted to applicants who ever obtain any university degree. Controls for age, experience (since degree), and application year. If an applicant applies multiple years in a row during this time period, only the last application is considered. Exam score based on best performance if applicant had multiple takes. Score is assigned on a scale between 2-6 with 6 corresponding to the highest grade. Higher observation count relative to Table A23 due to considering degrees for each year up to 2020, whereas wages are measured in 2020 only (and thus some applicants do not reside in the country in that year).
* p<0.1, ** p<0.05, *** p<0.01 (back)

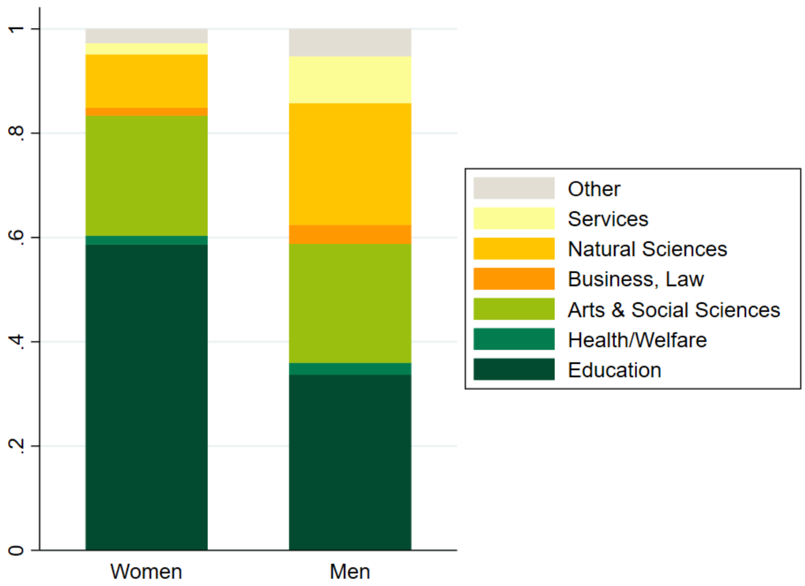
G.5 Wage Premia

Table A23: Wage Premia for Primary Teaching Degree Holders by Gender

	(1)	(2)	(3)	(4)
Teaching Degree	3,398***	3,386***	2,729***	2,714***
	(304)	(305)	(300)	(300)
Male * Teaching Degree	–4,978***	–4,911***	–6,637***	–6,547***
	(646)	(647)	(651)	(656)
Male	11,216***	12,127***	12,855***	14,444***
	(336)	(1,422)	(372)	(1,559)
Score	1,362***	1,406***	978***	1,047***
	(134)	(150)	(141)	(155)
Male * Score		–218		–374
		(330)		(356)
Adj. R^2	0.098	0.098	0.100	0.100
Obs	20,433	20,433	18,163	18,163
Dep mean	40229	40229	41363	41363

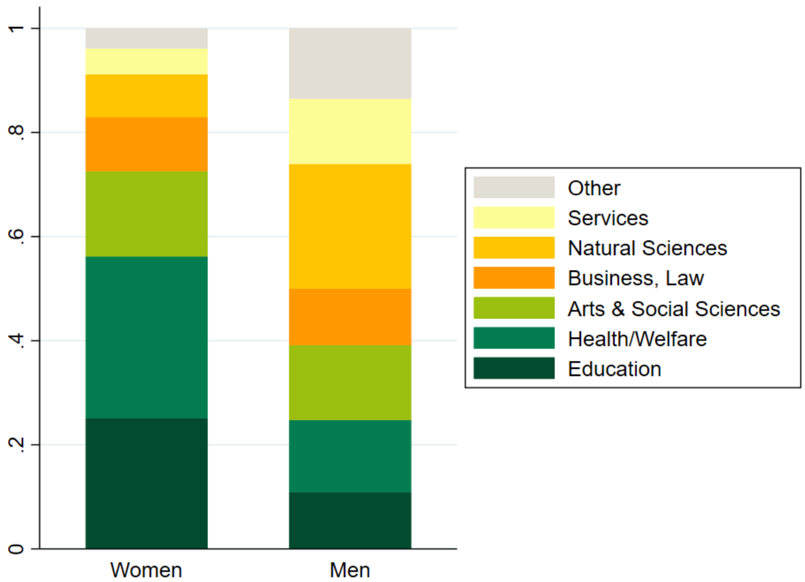
Note: Outcome is total wage earnings in the year 2020. In columns 3 and 4, sample is restricted to applicants who ever obtain any university degree. Controls for age, experience (since degree), and application year. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers all degrees obtained by the year 2020. If an applicant applies multiple years in a row during this time period, only the last application is considered. Exam score based on best performance if applicant had multiple takes. Exam score is assigned on a scale between 2-6 with 6 corresponding to the highest grade. * $p<0.1$, ** $p<0.05$, *** $p<0.01$ (back)

Figure A14: Share of Fields Applied to Outside Primary Teacher



Note: Share of applications to fields other than primary school teaching among all applicants with at least one application to primary teaching (with 60% of women's and 65% of men's total applications going to primary teaching not displayed). Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing). (back)

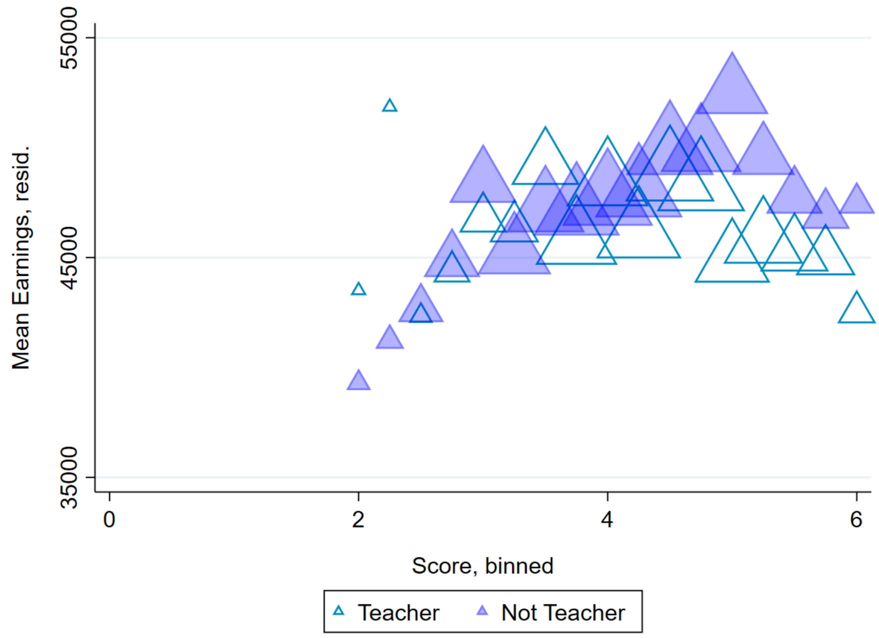
Figure A15: Actual Fields Obtained among Applicants Outside Primary Teaching



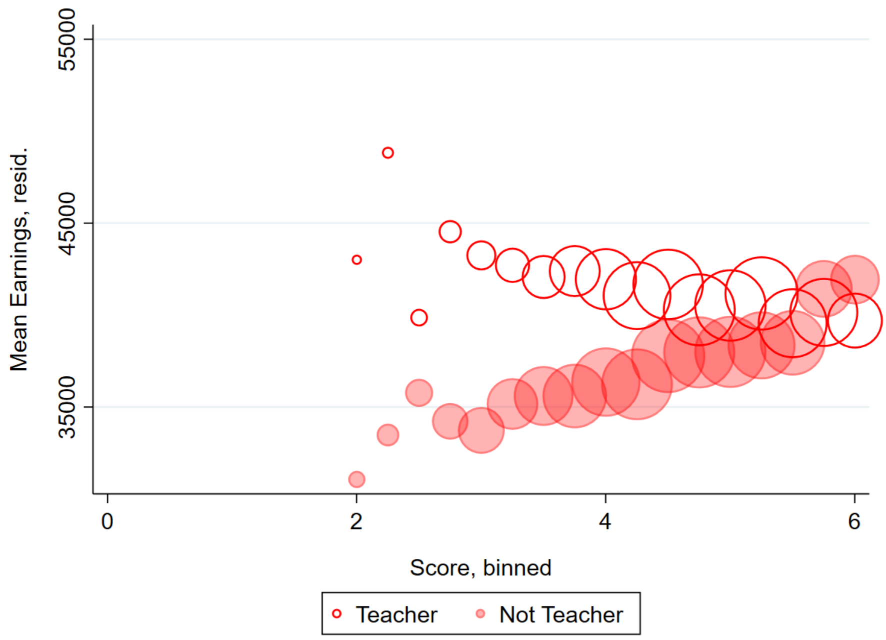
Note: Share of actual fields obtained among those who do not obtain a primary teaching degree by 2020. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers most recent degree obtained in 2020. (back)

Figure A16: Outside Options of Teacher Applicants: Residualized Earnings

(a) Men: Residualized Earnings



(b) Women: Residualized Earnings



Note: Residualized wage earnings in the year 2020 among all applicants to primary teaching by whether candidate holds a primary teaching degree against binned score in the matriculation exam. Weighted by number of applicants per bin. Controls for age, experience (since degree), and application year. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers all degrees obtained by the year 2020. Exam score is based on best performance if multiple takes. Exam score is assigned on a scale between 2-6 with 6 corresponding to the highest grade. (back)

H Robustness

H.1 Maternity/Paternity of teachers

Table A24: Teachers Becoming Parents

	Birth total	Birth fem	Birth male	Maternity leave
Share 60	–0.127***	–0.109***	–0.018	–0.003
	(0.040)	(0.029)	(0.025)	(0.030)
Share 60 * Post-Quota	0.075	0.036	0.039	–0.040
	(0.050)	(0.036)	(0.029)	(0.032)
Year FE	X	X	X	X
Adj. R^2	0.007	0.006	0.012	0.005
Obs	4448	4448	4448	4448
Dep mean	.06	.03	.02	.02

Note: Estimates for Equation 8. Outcomes from left to right: Share of teachers with the birth of a child, share of teachers who are female and have a birth (column 2), and who are male and have a birth (column 3), share of teachers who are female and on leave after birth (defined as not being an active teacher in the year subsequent to having given birth). Standard errors clustered at the municipality level. Regressions weighted by population, means unweighted. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A25: Effect on Main Outcomes of Female Teachers Having a Newborn Child

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Female Teachers Having a Child	–0.009	0.006	0.003	–0.003	–0.007	0.006*	0.003
	(0.008)	(0.008)	(0.003)	(0.010)	(0.008)	(0.003)	(0.005)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
Obs	825,094	825,094	825,094	811,392	811,392	811,392	811,392
Dep mean	.911	.067	.022	.842	.086	.017	.053

Note: Specification equivalent to Equation 10, but estimating the effect of total exposure to female teachers having a newborn child while pupils are in primary school. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

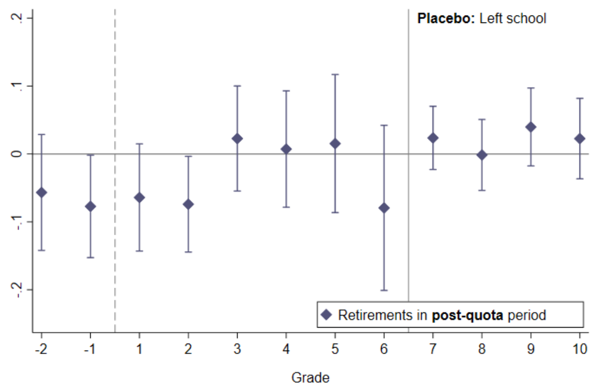
Table A26: Effect on Main Outcomes of Male Teachers Having a Newborn Child

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Male Teachers Having a Child	–0.008 (0.010)	0.006 (0.009)	0.002 (0.005)	–0.021 (0.013)	0.009 (0.010)	–0.007* (0.004)	0.018** (0.007)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
Obs	825,094	825,094	825,094	811,392	811,392	811,392	811,392
Dep mean	.911	.067	.022	.842	.086	.017	.053

Note: Specification equivalent to Equation 10, but estimating the effect of total exposure to male teachers having a newborn child while pupils are in primary school. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

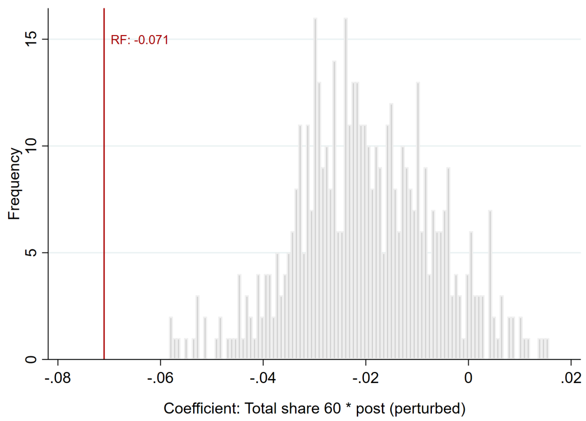
H.2 Placebo Check and Randomization Inference

Figure A17: Placebo Reduced Form by Grade: Applications to Post-Compulsory Education



Note: Grade level estimation of pupil level reduced form (see Equation 10). Outcome is binary indicator for pupils applying to post-compulsory education directly after middle school, regressed on the share of teachers turning 60 just before a pupil enters the respective grade, including "fake grades" in which the pupil has already left primary school. Figure depicts coefficients for the post-quota period *relative* to the quota period. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Figure A18: Reduced Form: Randomization Inference



Note: Perturbed reduced form. 500 iterations of randomly assigning the share of teachers turning 60 in any post-quota year across municipalities, keeping fixed quota-period retirements, and estimating Equation 10. Outcome is binary indicator for pupils applying to post-compulsory education directly after middle school. Line indicates the reduced form coefficient in Table A15. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

H.3 Sample Attrition and Further Robustness

Table A27: Test for Selective Sample Attrition

	RF		IV	
	Left 16	Left 25	Left 16	Left 25
Avg Share Male			0.026	–0.019
			(0.029)	(0.051)
Total Share 60	0.006	0.006	0.005	0.007
	(0.005)	(0.007)	(0.004)	(0.006)
Total Share 60 *	–0.004	0.003		
Post-Quota	(0.005)	(0.009)		
Municipal FE	X	X	X	X
Cohort FE	X	X	X	X
Region*Cohort FE	X	X	X	X
Ind. controls	X	X	X	X
MOP F^{eff}			15.28	15.28
Adj. R^2	0.042	0.049		
Obs	826,180	826,180	826,180	826,180
Dep mean	.005	.018	.005	.018

Note: Reduced form, and IV estimates for Equation 9. Outcomes are a binary indicator for pupils having left the sample at age 16 or age 25, excluding registered deaths. Pupils are defined as having left the sample if they do not appear in the register data at the respective age. Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A28: Further Robustness

	Apply				Employed/Student			
	Main	No capital	No cities	Parent UB	Main	No capital	No cities	Parent UB
Avg Share Male	0.424**	0.437**	0.458**	0.411**	0.512**	0.491**	0.503**	0.487**
	(0.197)	(0.198)	(0.214)	(0.195)	(0.243)	(0.237)	(0.255)	(0.238)
Municipal FE	X	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X	X
MOP F^{eff}	15.28	15.59	13.86	15.28	15.37	15.69	13.96	15.37
Obs	825,094	758,379	648,930	825,094	811,392	746,392	639,043	811,392
Dep mean	.911	.911	.911	.911	.842	.842	.842	.842

Note: IV estimates for Equation 9. Outcomes are pupils' applications in the last year of middle school (see Table 4) and labor market status (see Table 6), in turn examining the main specification for comparison (column 1 and 5), dropping Helsinki (column 2 and 6), dropping the five most populous municipalities based on place of living at age 7 (column 3 and 7), and controlling for parental unemployment status at age 7 (column 4 and 8). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

Table A29: Reduced form estimates by size of municipality

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Small: Total	–0.098**	0.082**	0.016	–0.112	–0.020	0.030*	0.095***
share 60 * post	(0.045)	(0.038)	(0.019)	(0.068)	(0.051)	(0.016)	(0.029)
Large: Total	–0.063**	0.052*	0.011	–0.079**	0.014	0.018	0.044**
share 60 * post	(0.028)	(0.027)	(0.013)	(0.033)	(0.027)	(0.012)	(0.020)
Municipal FE	X	X	X	X	X	X	X
Region*Cohort FE	X	X	X	X	X	X	X
Ind. controls	X	X	X	X	X	X	X
Obs	825,427	825,427	825,427	811,723	811,723	811,723	811,723
Low: Dep mean	.91	.067	.023	.842	.086	.017	.053
High: Dep mean	.923	.056	.021	.849	.084	.016	.049
p-value	.448	.455	.809	.637	.524	.472	.097

Note: Reduced form estimates (Equation 10), estimating separate coefficients for small (12 teachers or fewer) and large municipalities (> 12 teachers). P-values for test of equality of coefficients in the post-quota period. Outcomes are pupils' applications in the last year of middle school (see Table 4) and labor market status (see Table 6). Individual level controls are measured at age 7 and include gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. * p<0.1, ** p<0.05, *** p<0.01 (back)

H.4 Brief Discussion of Macro-Economic Shocks

During the period of our study, two major macro-economic shocks warrant a brief mention: The depression in Finland during the early 1990s, as well as the financial crisis in 2008/09, initiating the global great recession. We study the cohorts born between 1981 - 1993, starting school between 1988 - 2000.

Finland experienced a 14% contraction of GDP from 1990-93, accompanied by a more permanent rise in unemployment. Region-by-cohort fixed effects in all of our specification allow for differential impacts of this shock across different parts of the country. In addition, we run our main specification controlling for parental unemployment status at age 7 in columns 4 and 8 of Appendix Table A28, with the main effects quantitatively unchanged.

Regarding the great recession, it is worth noting that our treatment assignment is based on the place where pupils live when they are age 7. Our study cohorts turn 25 years old in the years 2006-2018. It is thus the earlier and middle cohorts that are both more exposed to male quota teachers and turn 25 during the time of the financial crisis and subsequent recession.

H.5 TWFE Robustness

This section explores potential bias in $\hat{\beta}_{fe}$ from negative weights in TWFE estimation due to heterogeneous treatment effects in our setting. The main concern - outlined by the relevant literature - is that previously treated units exhibiting dynamic treatment effects over time are used as a control group for newly treated units. When treatment effects are e.g. increasing over time, the fixed effects difference out a change in the control group consisting of previously treated units that is "too large", leading to potential sign reversal in the estimator.

De Chaisemartin and d'Haultfoeuille (2020) decompose $\hat{\beta}_{fe}$ into a weighted sum of average treatment effects (ATE) for treated units, with weights proportional to the residual from a regression of the treatment variable on fixed effects. If treatment effects are heterogeneous, problems with sign reversal arise when treated observations receive a negative weight due to their *residualized* treatment value in a particular period being negative (intuitively, these negative weights arise because that particular observation serves as a control in that period). Negative weights by themselves are mechanically the product of any TWFE specification – it is in combination with heterogeneous treatment effects that problems with sign reversal may arise. While the literature to date has not offered diagnostic tools for our particular case where treatment is continuous and infinite, we can use the intuition developed in De Chaisemartin and d'Haultfoeuille (2020) (also highlighted by Jakiela, 2021) to probe for such issues in our setting.

First, the highlighted concern arises only when treatment effects are heterogeneous. The way in which treatment effect heterogeneity matters in our IV set-up is through the first stage relationship by using residualized treatment assignment to generate predicted values for the endogenous variable in the second stage. The first stage in our setting estimates a mechanical relationship between local retirements and teacher gender composition, with a clear ex-ante prior on sign and magnitude. While there is no direct test of assessing treatment effect heterogeneity, reporting sensitivity to particular groups or time periods may at least be partly illuminating about whether the first stage coefficients are driven by any particular group of observations. To this extent, leave-one-out estimation in the following section (Appendix H.6) reports coefficients that show no worrisome patterns.

A further probing for treatment effect heterogeneity consists in examining the relationship between the residualized outcome and residualized treatment variable. Under homogeneous treatment effects, this relationship should be linear and not differ by treatment assignment status. In the first stage of our setting, pupil cohorts that experience relatively more retirements in the post-quota period are 'treated' by being exposed to fewer male quota teachers, whereas pupils with relatively more retirements in the quota period serve as the 'control' group. Appendix Figures A19a and A19b plot the residualized share male against the residualized treatment variable both for the municipal and the pupil level first stage.[64] A test for differences in slopes between treatment and control observations shows that these are small and not significant.

[64]I.e. for the first stage equation 7 at the municipal level, residualized treatment corresponds to the residuals of the following specification:

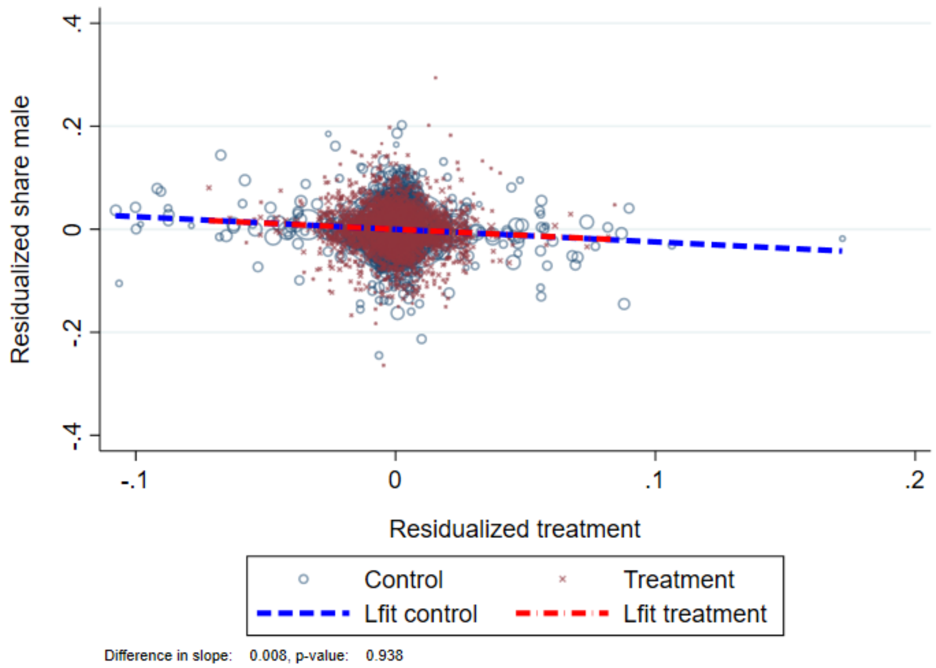
$$\textit{total share 60 post}_{mt} = \beta_0 + \beta_1 \textit{total share 60}_{mt} + \gamma_{mp} + \eta_t + \epsilon_{mt} \tag{13}$$

With $\textit{total share 60 post}_{mt}$ the share of teachers turning 60 interacted with an indicator for the post period.

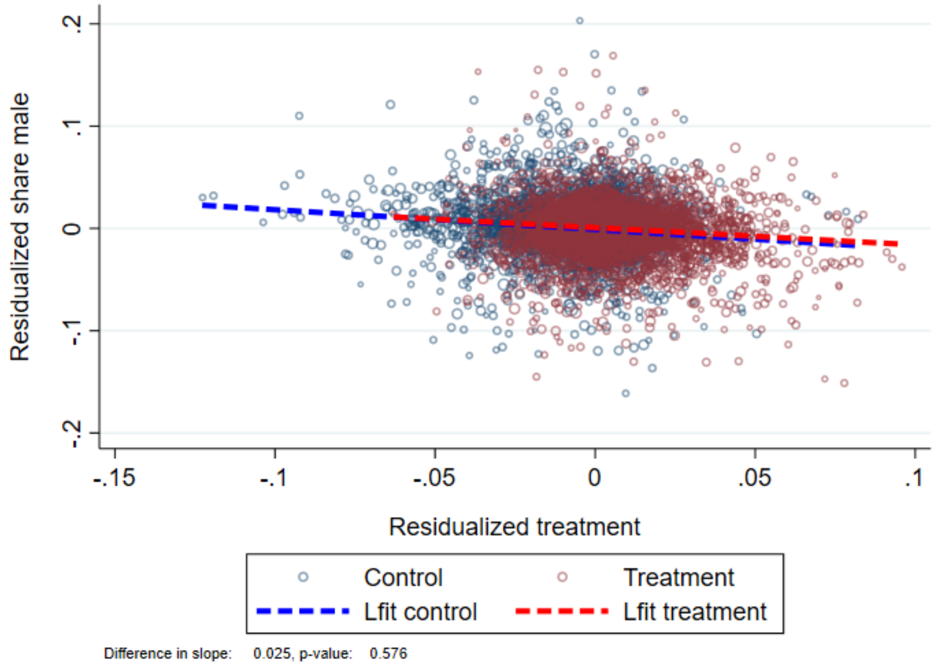
Second, following De Chaisemartin and d'Haultfoeuille (2020), we can examine the weights that observations receive. The focus here is to understand how treatment assignment based on actual treatment status maps into treatment assignment based on the *residualized* treatment variable. In our setting, municipality-by-year or municipality-by-cohort observations with more retirements in the quota relative to the post-quota period should serve as a control group based on actual treatment assignment, and thus receive a negative weight (i.e. exhibit a negative *residualized* treatment assignment). Appendix Figure A20 plots residualized treatment assignment against actual treatment assignment both for the municipal and the pupil panel separately. Reassuringly, the mapping between residualized and actual treatment assignment follows a clear pattern: observations with higher retirement in the quota period are those that exhibit on average a negative residualized treatment value (i.e. receive a negative weight).

Figure A19: Treatment Effect Heterogeneity

(a) Municipality Level: Residuals of First Stage



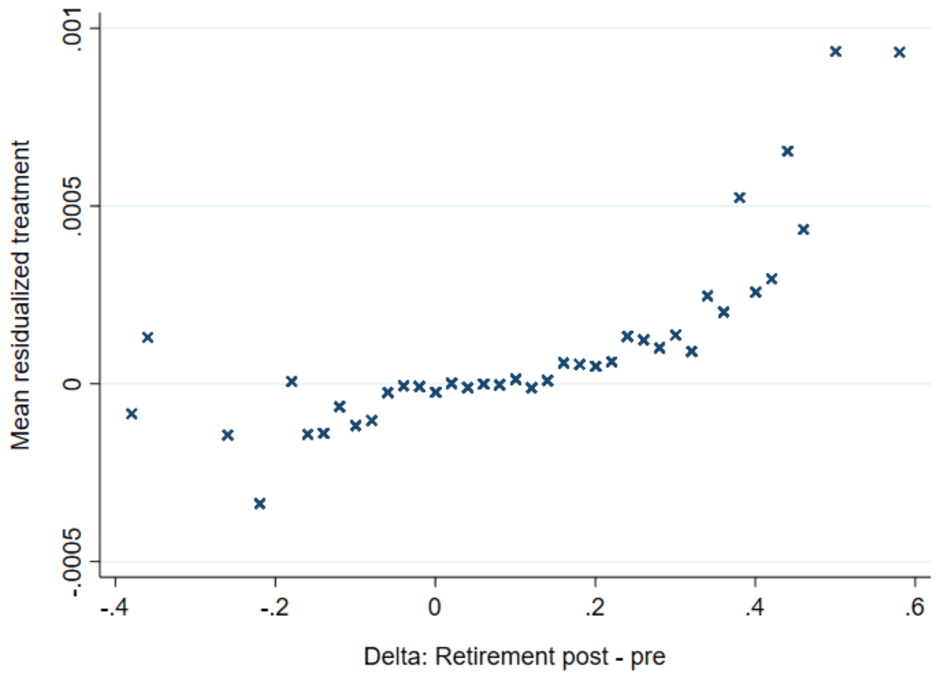
(b) Pupil Level: Residuals of First Stage



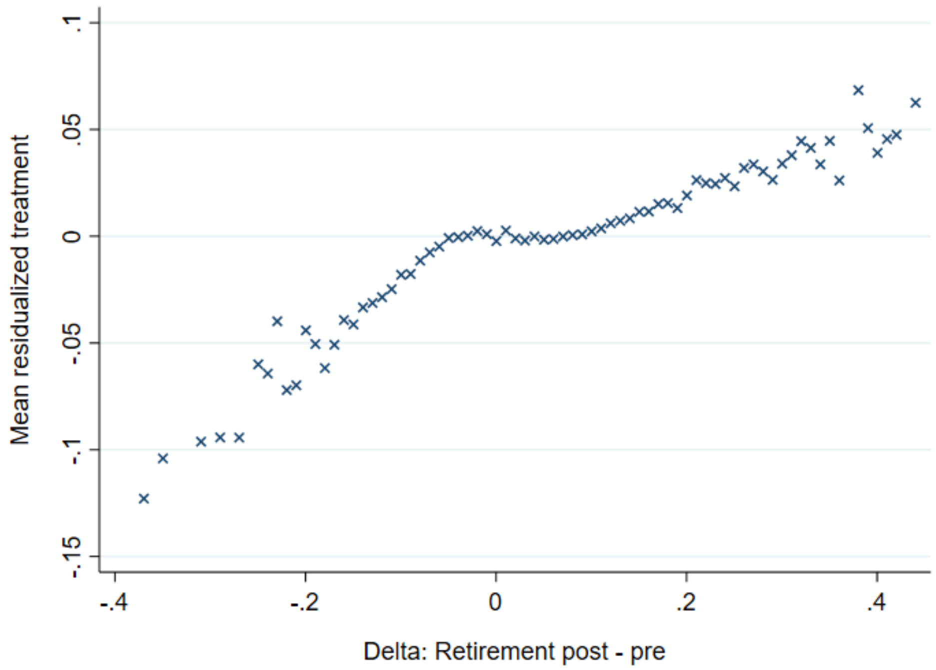
Note: Residualized share male against residualized treatment (see see Equation 13), at the municipal level (upper panel) and pupil level (lower panel). Weighted by number of observations. (back)

Figure A20: Residualized Treatment

(a) Municipality Level: Residualized Treatment ("weights")



(b) Pupil Level: Residualized Treatment ("weights")

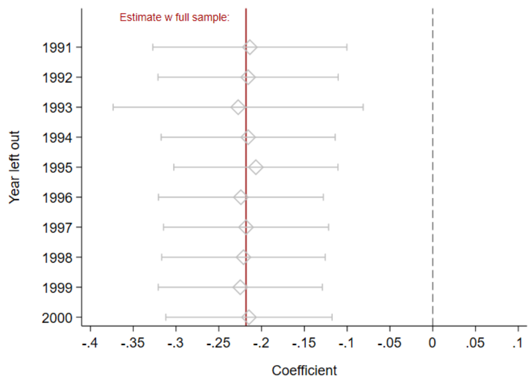


Note: Mean residualized treatment (see Equation 13) against actual treatment assignment (binned) at the municipal level (upper panel) and pupil level (lower panel). (back)

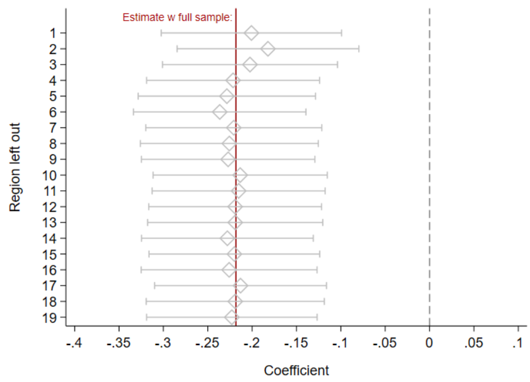
H.6 Leave-One-Out Estimation

Figure A21: First Stage: Leave-One-Out

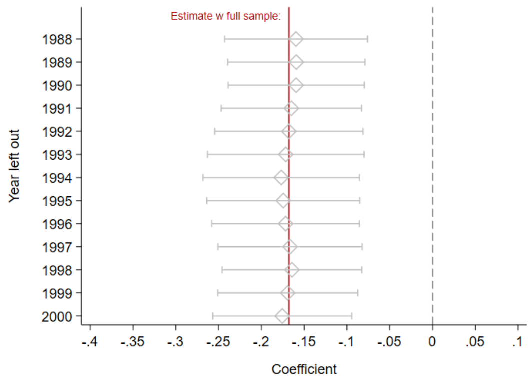
(a) Municipal Level: Leave-One-Out Years



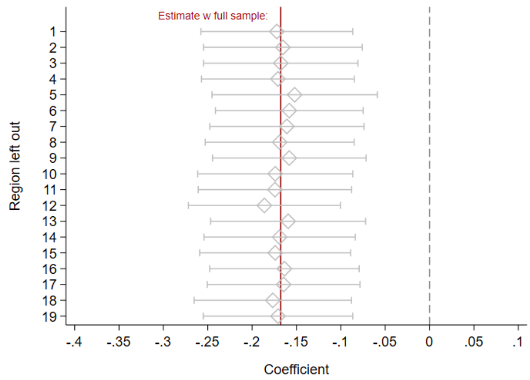
(b) Municipal Level: Leave-One-Out Region



(c) Pupil Level: Leave-One-Out Cohorts



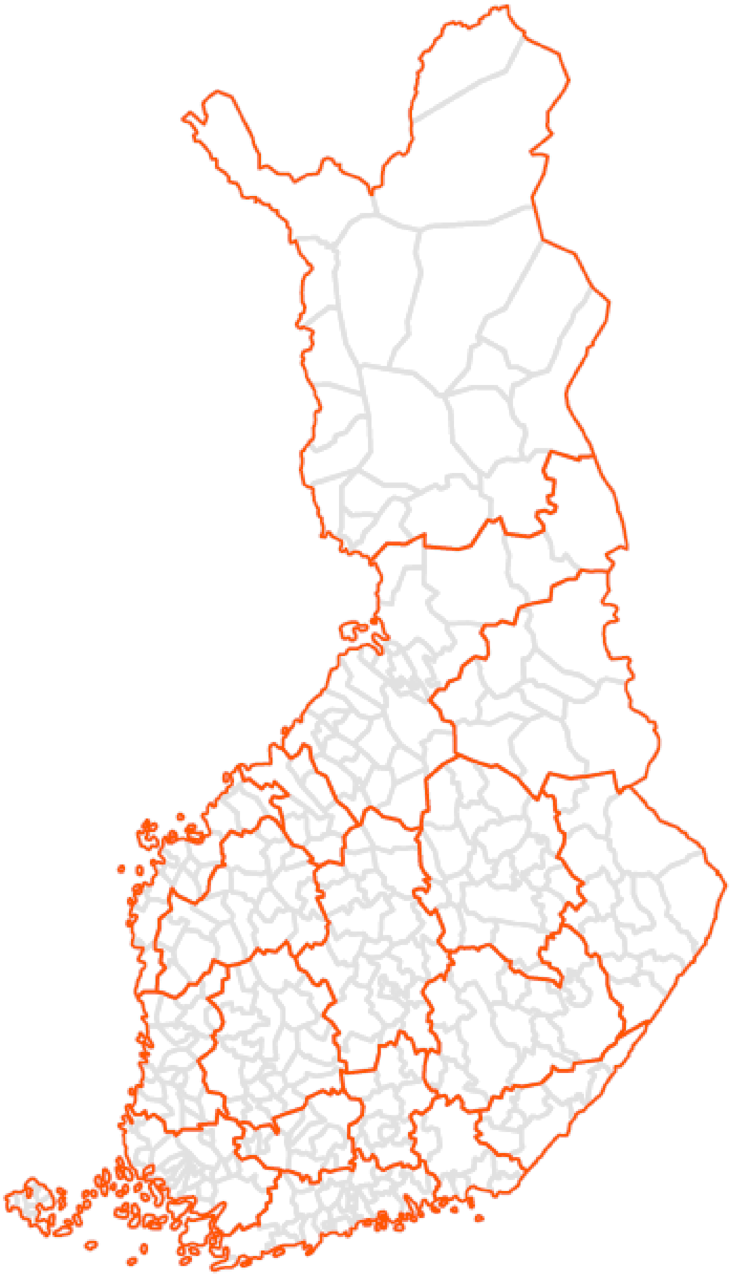
(d) Pupil Level: Leave-One-Out Region



Note: Leave-one-out estimation of treatment coefficient in municipal level first stage Equation 7 (Panel a and b) and pupil level first stage Equation 10 (Panel c and d), with respect to regions and years/cohorts. Indicated years/cohorts and regions on the y-axis are the respective observations dropped in the estimation of the coefficient. (back)

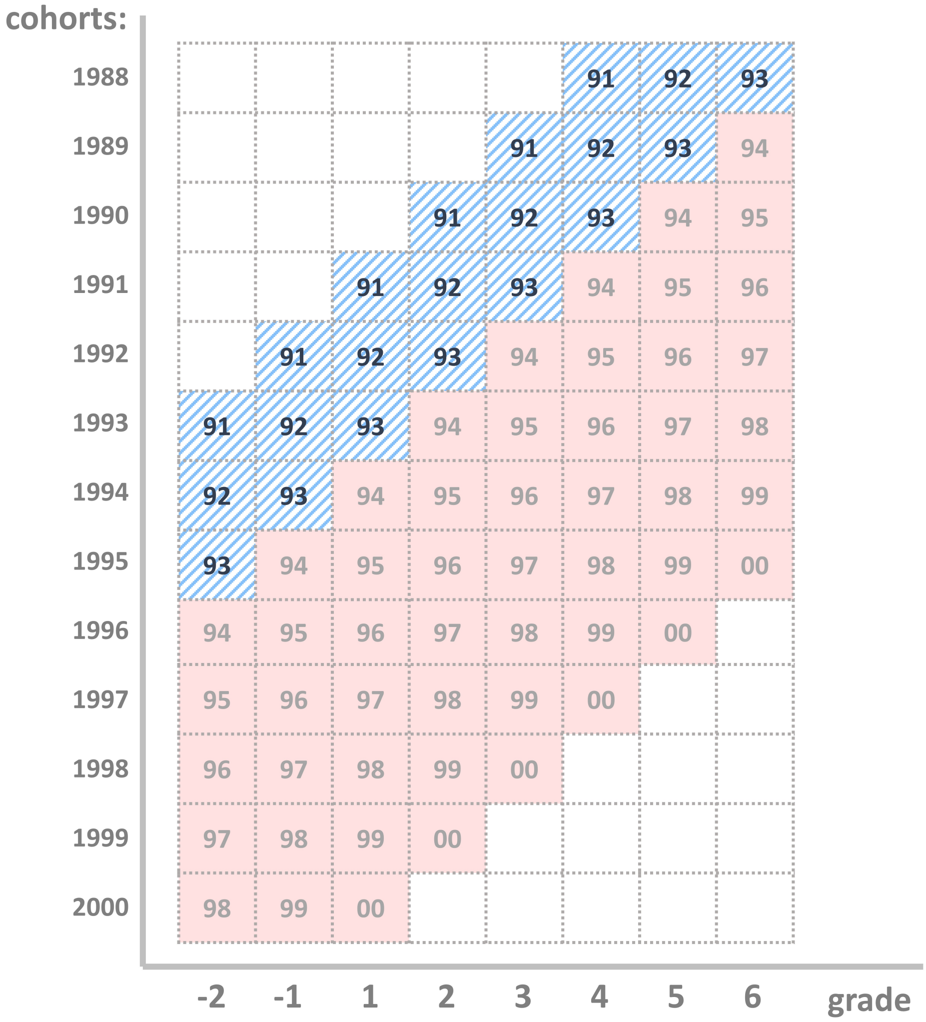
I Context

Figure A22: Region and Municipality Borders, Finland



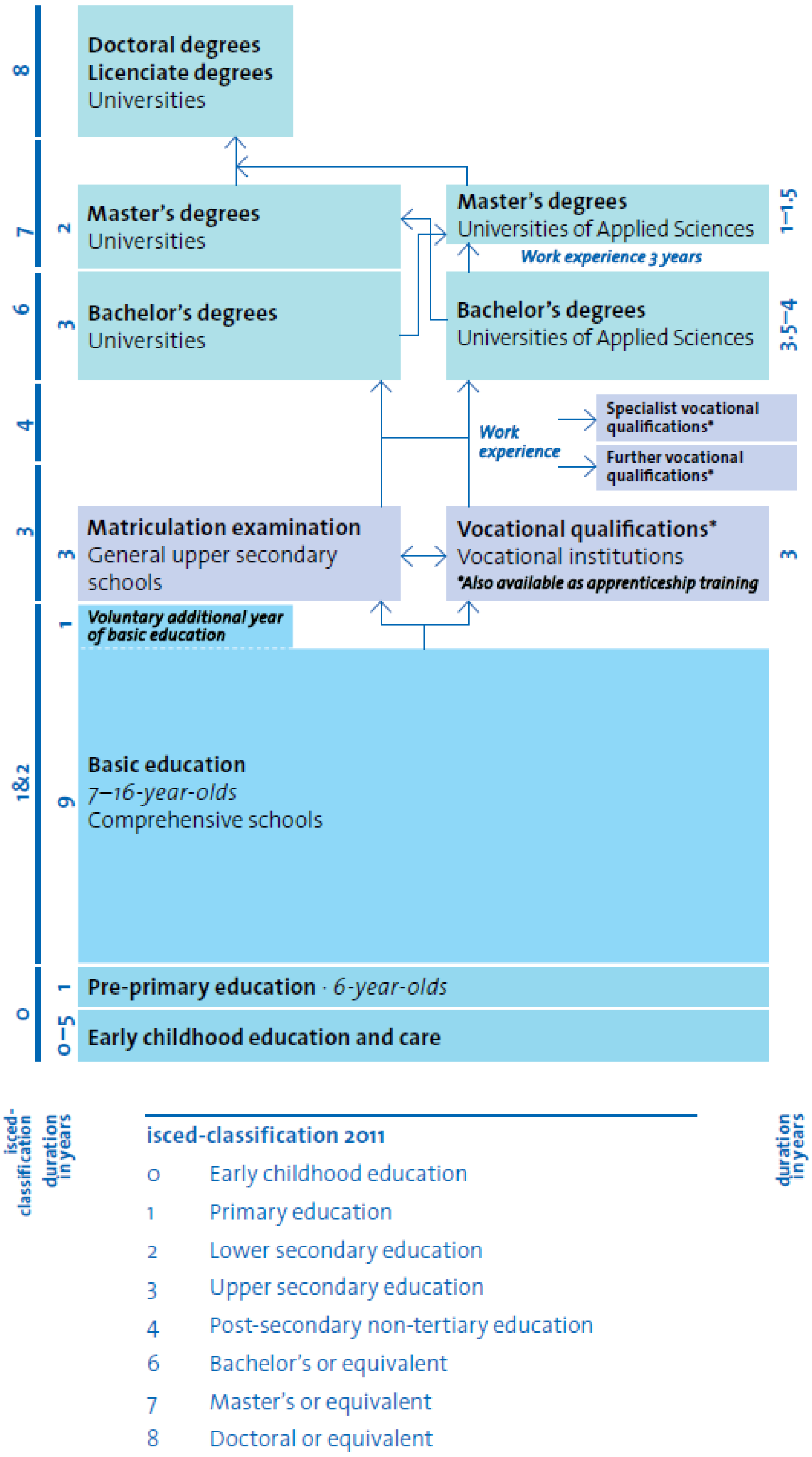
Note: Borders for 2019, shapefiles provided by Statistics Finland.

Figure A23: Pupil Cohorts and Exposure to Quota Period



Note: Figure shows cohorts by year in which they turn seven years old, and exposure to the quota by the grades which they spend in primary school. Years in which the quota was still in place colored in blue (with stripes), years in which the quota was abolished in red. (back)

Figure A24: Finnish Education System



Note: Source: Ministry of Education, Finland. (back)