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

Ursina Schaede¶ Ville Mankki‖

June 2024

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-bargained 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. Evidence suggests that the quota improved the allocation of talent by mending imperfections in the unconstrained 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, Muriel Niederle, 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.

¶Tufts University, Department of Economics: ursina.schaede@tufts.edu (corresponding author)
‖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 (Barnes, 2023; Green, 2022; Dhume, 2019). However, we still have a relatively limited understanding of the equity-efficiency trade-offs associated with such 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' 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 benefitting academically lower-scoring male candidates. We document that these "male quota teachers" perform better on the job 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. We then show that these productivity gains materialize because the unconstrained selection criteria disadvantage the underrepresented group and are not related to output. Instead, male teachers in this context exhibit several attributes, namely career attachment and intrinsic motivation, that may explain their pupils' perseverance regarding their own professional paths.

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-bargained 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 male quota teachers, 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 teachers due to the lifting of the quota.

We start by outlining a general conceptual framework of university admissions as a signal extraction problem. In the presence of group-specific bias or mean ability, an unconstrained admissions office sets optimal threshold scores that differ by group. We show that requiring admissions to set equal scores across groups — i.e. abolishing affirmative action, or the quota in the Finnish case — can be detrimental both for candidate quality and equal representation when selection criteria exhibit bias against the underrepresented group.

We then examine the efficiency effects of the male teacher 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 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 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

[1] 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 admitted otherwise. Throughout the paper, we refer to teachers turning 60 as "retirement."

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.

Turning to long-term impacts up to early adulthood, we examine pupils' educational attainment, labor market attachment and field of study by age 25. For pupils who were exposed to a higher share 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 standard deviation increase in the share of male quota teachers, which corresponds to a 4% increase over the mean. Using data from degree registers, we further document 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.

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 selection 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 admission criteria disadvantaged the underrepresented group. The matriculation exam score attaches a weight of 75% to language fields, in which men perform relatively worse, and 25% to math and science fields, in which men perform relatively better. Second, we document that this selection criterion is not directly related to productivity on the job. 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 document that the underrepresented group exhibits several attributes that may enhance productivity in this setting. In particular, we show that male teachers display a high degree of dedication to their career as a teacher, which may explain why their pupils are more career-oriented themselves. Male teachers are less likely to leave their career as a primary teacher and have higher earnings — which due to the deterministic nature of the salary scale in this setting — is indicative of taking over additional responsibilities beyond regular teaching hours. These differences between male and female teachers are not fully accounted for by family formation and exist even for teachers without children. We further show evidence that highlights male teachers' intrinsic motivation when pursuing a stereotypically female occupation. Analyzing compensating differentials, we document that 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 are indicative of men's intrinsic motivation in the teaching profession, as they forgo higher earnings relative to their outside option. Both of these patterns are consistent with male teachers setting an effective role model for their pupils in terms of career attachment, thus inspiring them to persevere in their own pursuits. This impact is not gender specific as both boys and girls benefit to a similar extent from exposure to male quota teachers, and it does not operate via improving pupils' grades.

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. While providing a role model in terms of career orientation may explain the underrepresented group's productivity in the context of this study, the particular group-specific characteristics that matter for output will certainly be setting-specific.

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 (Bleemer, 2022; Otero et al., 2021; Black et al., 2023; Bleemer, 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).[2]

Second, 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; Dobbie et al., 2022). 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, 2023), to the detriment of equal representation and aggregate productivity.

Third, our paper relates to studies on teacher value-added and gender-based role models. A

[2]Peck (2017) and Cortés et al. (2023) 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).

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 economic outcomes for pupils (Rivkin et al., 2005; Chetty et al., 2014), with recent work emphasizing dimensions of teacher value-added that are unrelated to pupils' test scores but matter for long-term outcomes (Jackson, 2018; Beuermann et al., 2023, Petek and Pope, 2023; Rose et al., 2022). 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, 2020; Carrell et al., 2010), but others fail to find positive impacts (see de Gendre et al. (2022) for a meta analysis).

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 exceeds 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.[3] 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 does not map into 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. We turn to mechanisms in Section 7 and robustness checks in Section 8; the final section concludes.

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

[3] In the presence of factors such as stigma against the underrepresented group, one would *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).

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.

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 of 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, p.27).

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 those reforms, including that "the Ministry of Education has maintained the sex quota system for the training of classroom teachers" (Sysiharju, 1987, p.33): In a first step, applicants were ranked in a centralized system according to 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). 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.[4] 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. While the quota was in place, men on average scored about .45 grade points, or a bit more than half a standard deviation 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. Appendix Table A1 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. Regarding educational trajectories in Panel B, there is no difference in having obtained a high school degree and being a certified teacher. In Panel C, statistics on the matriculation exam show similar pass rates, 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

[4] 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 A1) or the composition of applicants would have drastically changed with its lifting. Figure 2 shows no discontinuity in the matriculation exam scores of those admitted post-quota.

basic level exam.

3 Conceptual Framework

This section develops a brief conceptual framework that serves to illustrate the potential trade-offs between candidate ability and group representation. In our setting, an unbiased admissions office would like to maximize future teacher ability, taking the role of a social planner. However, true teacher ability is unobserved so that the admissions office has to select candidates based on scores. 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 allows for group-specific admission thresholds and takes the form of a binding representation target in the Finnish context.[5] 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 underrepresented groups' lower performance on evaluation 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.[6] 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 $h_g(s)$. The admissions office is required to set score thresholds above which every candidate within a specified group is admitted, such that randomization is not permitted. Scores are an imperfect measure of ability a_g due to group-specific bias, b_g, and overall noise e. In particular:

$$s_g = a_g + e - b_g \tag{1}$$

$$where \;\; a_g \sim \mathcal{N}(\mu_{a_g}, \sigma_a^2) \;\; and \;\; e \sim \mathcal{N}(0, \sigma_e^2)$$

e independent of a_g; and b_g a constant.[7] The expected ability of a candidate from group g given

[5] 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 7.2.3.2, we document that post-quota admission probabilities conditional on exam score do not differ by gender in the Finnish context. 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.

[6] The admissions office's problem is similar to Chan and Eyster (2003), who show that forbidding universities that have an exogenous taste parameter for diversity may result in lower quality of admits as the optimal admissions rule under race-blind admissions is generically non-monotone (Ray and Sethi, 2010). In contrast, we do not assume a taste-parameter for diversity and add an explicit distinction between observable scores and unobserved ability. Note also that the teacher admissions office in our setting is unlikely to practice taste-based discrimination against men by choosing a test that favors women (Becker, 1957), as the matriculation exam in Finland is commonly used for university admissions in most fields (Kupiainen et al., 2018) and teacher education departments themselves campaigned for a reinstatement of the quota after it was abolished (Tasa-Arvovaltuutetu, 1992).

[7] The distributional assumptions we make here are equivalent to Autor and Scarborough (2008) who document that a more precise additional employee screening test raises productivity without decreasing minority repre-

their score s is then given by

$$E[a|s,g] = \frac{\sigma_e^2}{\sigma_e^2+\sigma_a^2}\,\mu_{a_g} + \frac{\sigma_a^2}{\sigma_e^2+\sigma_a^2}\,(s+b_g) \tag{2}$$

with $\frac{\sigma_a^2}{\sigma_e^2+\sigma_a^2}$ equal to the squared correlation coefficient between ability and scores, $\rho^2 \equiv \rho_{a,s}^2$. An unconstrained admissions office sets optimal threshold scores s_g^* such that total expected ability of admits is maximized, subject to a capacity constraint:

$$\max_{s_M, s_F} \int_{s_M} E(a_M \mid s)\, h_M(s)\, ds + \int_{s_F} E(a_F \mid s)\, h_F(s)\, ds \tag{3}$$

$$s.t. \;\; N_M + N_F = c$$

with N_g the mass of candidates admitted from group g. This results in group-specific cut-off scores given by:

$$s_M^* = s_F^* - \frac{\sigma_e^2}{\sigma_a^2}\,(\mu_{a_M} - \mu_{a_F}) - (b_M - b_F) \equiv s_F^* - \kappa \tag{4}$$

The mass of candidates of each group admitted is then given by (see Appendix B):

$$N_g = 1 - \Phi(z_g^*) \tag{5}$$

with $\Phi(\cdot)$ the CDF of the standard normal distribution, and $z_g^* = \frac{\rho}{\sigma_a}(s_g^* - \mu_{a_g} + b_g)$. The mean ability among admits of each group is:

$$E[a_g|s_g > s_g^*] = \mu_{a_g} + \sigma_a\, \rho\, \lambda(z_g^*) \tag{6}$$

with $\lambda(\cdot)$ the inverse Mills Ratio.

Given this set-up, adding a quota constraint for the share of male admits in the admissions office's optimization problem (Equation 3) is analogous to mandating the implied distance between group-specific threshold scores. If binding, this constraint adds a wedge to optimal threshold scores: $s_M^* = s_F^* - \kappa - \delta$, with δ the Lagrange multiplier of the representation constraint. 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.

3.2 Representation and Efficiency

In the following, we discuss how restricting an admissions office to admit candidates based on the same score threshold impacts the quality of admitted candidates, using Equation 4. It is

sentation since the relative bias between tests remains unchanged. This is in contrast to models of statistical discrimination (Aigner & Cain, 1977; Phelps, 1972) that rely on the assumption of group-specific signal precision. In our setting, similar to Autor and Scarborough (2008), the score distributions between groups exhibit a mean shift, but similar variance (see also Section 7.2.1). This mirrors group differences in test taking and on-the-job performance 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 and is equally predictive of job performance (Hartigan and Wigdor, 1989).

straightforward to see that when scores fully reflect teaching ability, i.e. the test is not biased against one group and there are no group-specific differences in mean ability, optimal admissions will set equal threshold scores across groups. However, whenever $\kappa \neq 0$, optimal scores are group-specific, and forcing an admissions office to use equal score thresholds (i.e. abolishing affirmative action) poses a binding constraint to its optimization problem. When the admissions office is restricted to admit both groups based on the same cut-off score in the presence of group differences, there will be fewer students admitted from the group g for whom the group-specific cutoff is lower than the single threshold, $(s^* > s_g^*)$, which raises mean ability among the remaining admits of this group (Equation 6).

1) Group differences in bias: The test is biased against one group if $b_M - b_F \neq 0$, such that there are differences in test performance that do not reflect ability differences. Optimal admissions will adjust for this bias by setting group-specific score thresholds that are lower for the group that the test disadvantages. All else equal, when scores are biased against men $(b_M > b_F = 0)$, total expected ability of admits is maximized when men are admitted with a lower threshold score that exactly offsets the bias.

2) Group differences in mean ability: Similarly, optimal admissions will set group-specific scores if there are group differences in mean ability, $\mu_{a_M} - \mu_{a_F} \neq 0$. Intuitively, this is because scores are measured with noise e, such that unconstrained admissions places weight not just on scores, but also on group-specific mean ability when assessing a candidate's expected ability (Equation 2). If, for example, male applicants exhibit characteristics that make them better teachers on average $(\mu_{a_M} - \mu_{a_F} > 0)$, an unconstrained admissions office will take this into account by setting a lower threshold score for men, and more so the less precise the signal.

3) General signal precision: By assumption, the noise term e is not group-specific and thus matters for group-specific admission thresholds only insofar as gaps in mean ability are present. As the variance of noise, σ_e^2, increases, scores become an increasingly unreliable signal of ability (ρ decreases) and the admissions office thus emphasizes group mean ability rather than scores when assessing a candidate's expected ability (Equation 4). However, noise also weakens the relationship between observable scores and ability, which we discuss below.

Taking stock: Group-specific biases and overall signal precision may each contribute to make scores an imperfect signal of a candidate's true ability. An unconstrained admissions office will take these factors into account when setting optimal admissions thresholds. In the Finnish setting, once the quota for men is abolished, admissions are required to apply the same score threshold across both male and female candidates. In the presence of group differences in scores that disadvantage men ($\kappa > 0$), this will result in fewer male candidates admitted, and overall teacher quality will be lower (as long as the quota did not overcompensate for this disadvantage). This decrease will be stronger the less precise the test signal in the presence of higher mean ability among male applicants. In general, when 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.

3.3 Affirmative Action and the Bias-Precision Trade-Off

The conceptual framework also highlights a trade-off with respect to bias and precision of the signal when an admissions office has to choose between two imperfect measures of ability (e.g. GPA and SAT scores) and group-specific thresholds are no longer allowed. Consider two different tests that deliver scores, s_1 and s_2. Let b_{g_j} denote the bias of Test j against group g, and ρ_j the correlation between test score j and ability. Suppose Test 2 is more biased against the underrepresented group, but provides a more precise signal of ability for both groups:

$$b_{M_2} > b_{M_1} > b_F \;\; and \;\; \rho_2 > \rho_1$$

For ease of exposition, we assume $\mu_a = \mu_{a_F} = \mu_{a_M} = 0$, $b_F = 0$ in the following and we assume that selection of candidates from each group is from the right tail. To illustrate how representation and efficiency change when admitting candidates based on s_2 rather than s_1, let

$$\Upsilon_j = \left\{1 - \Phi(z^*_{M_j})\right\} \sigma_a \rho_j \lambda(z^*_{M_j}) + \left\{1 - \Phi(z^*_{F_j})\right\} \sigma_a \rho_j \lambda(z^*_{F_j}) \tag{7}$$

denote the total expected ability of admits for Test j. Appendix Section B.2 contains the proofs for this section.

Unconstrained Admission: An unconstrained admissions office will always prefer to admit based on the test with higher precision (i.e. Test 2) as it selects higher expected ability candidates from both groups. Intuitively, this is because optimal threshold scores are group-specific and offset any bias (Equation 4). There are no changes in representation when admitting based on Test 2 instead of Test 1, but because Test 2 sends a clearer signal of a candidate's true ability, higher expected ability candidates are admitted from both groups ($\Upsilon_2 > \Upsilon_1$).

Abolishing Affirmative Action: When an admissions office is not allowed to set group-specific thresholds, admissions based on Test 2 will admit an excess of female candidates and under-admit men relative to Test 1. Both higher bias and higher precision contribute to admitting fewer candidates from the underrepresented group as increasing precision under-admits the group against which the test is biased. However, the net effect on total expected ability under Test 2 is ambiguous: While it delivers gains in average male talent both due to a better signal of ability and by admitting fewer men (thus moving to the right in the ability distribution), the impact on average female talent is ambiguous. Gains from precision may be offset by losses from admitting an excess of female candidates. The net effect on total expected ability of admits when moving from Test 1 to Test 2 thus depends on how the gains from increased precision compare to the loss from under-admitting candidates from the disadvantaged group.

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 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, degree and income]. 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 7.2.3.2.

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, and 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 link to pupils' matriculation exam grades [YTL moduuli].

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 both within and across schools in the same municipality (see Section 2.1). 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 grade levels in 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).[8]

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-bargained 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. 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, using 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 A19 outlines the timeline of the 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.

[8] 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.

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:

$$share\ male_{mt} = \pi_0 + \pi_1\ total\ share\ 60_{mt} + \pi_2\ \mathbb{1}_{t=post}\ total\ share\ 60_{mt} + X_{mt}\delta + \eta_{rt} + \gamma_{mp} + \zeta_{mt} \quad (8)$$

with $share\ male_{mt}$ the share of male teachers in municipality m in a given year t, and $total\ share\ 60_{mt}$ the cumulative teacher retirements in a municipality up to that point in time in the sample.[9] 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 cumulative retirements once the post-period starts to separately estimate how retirements affect the local share of male teachers post-quota.[10]

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

Structural equation: We would like to estimate how increasing the share of male teachers via the quota affects pupils' outcomes. As such, we are not interested in the impact of male teachers *per se*, but in the local average treatment effect of swapping quota men for marginal women. This is estimated in a two-stage least-squares framework using the timing of local retirement shocks as an instrument for the average share of male teachers in the following structural equation:

$$y_{im,t+x} = \beta_0 + \beta_1\ \overline{share\ male}_{mt} + \beta_2\ \overline{total\ share\ 60}_{mt} + X_i\delta + \gamma_m + \eta_{rt} + u_{imt} \quad (10)$$

with $y_{im,t+x}$ the outcome of interest at time $t+x$ for pupil i who at age seven lived in municipality m, $\overline{share\ male}_{mt}$ the average of the share of male teachers across the years we observe pupils in primary school and X_i individual level controls.[11] We add municipal fixed effects γ_m, as well as

[9] The fixed effects specification of equation 8 uses the stock of the dependent variable (the share of male teachers) and the independent variable (the cumulative share of teachers retiring 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 of teachers retiring in each year, dropping the municipal fixed effects:

$$\Delta share\ male_{mt} = \pi_0 + \pi_1\ share\ 60_{mt} + \pi_2\ \mathbb{1}_{t=post}\ share\ 60_{mt} + X_{mt}\delta + \eta_{rt} + \zeta_{mt} \quad (9)$$

We report first stage results for both equations, and use equation 9 when thinking in flows is more intuitive for robustness checks on hiring patterns.

[10] The municipality-by-period fixed effects are necessary because of the cumulative nature of *total share 60* and ensure that we are using variation that arises from deviations from the municipality*treatment-period mean instead of the municipal mean only.

[11] The controls we include are indicator variables for pupil gender, language (Swedish, Finnish, other), foreign origin, single parent household, and highest level of education in the household (Compulsory, Secondary, Tertiary, n/a) measured at age seven. Our pupil panel spans 13 cohorts that are starting school in the years 1988-2000

region-by-cohort fixed effects η_{rt}.

Our empirical strategy isolates variation in the share of male quota teachers from compositional changes in the inflow of recently graduated teachers that is caused by retirements. As rookie teachers may differ from older teachers along various dimensions, we account for pupils' overall exposure to rookie teachers via retirement by controlling for the exposure-weighted share of teachers retiring during a pupils' time in primary school, $\overline{\textit{total share 60}}_{mt}$. We discuss this measure in more detail below.

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 8. 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:

$$\overline{\textit{share male}}_{mt} = \pi_0 + \pi_1\,\overline{\textit{total share 60}}_{mt} + \pi_2\,\mathbb{1}_{t=post}\,\overline{\textit{total share 60}}_{mt} + X_i\delta + \gamma_m + \eta_{rt} + \epsilon_{imt} \quad (11)$$

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 two-stage least-squares set-up, $\hat{\beta}_1$ thus measures how increasing the average share of male teachers via the quota affects pupils' outcomes.

Exposure to retirement: We construct a measure of exposure to teacher retirements that reflects that retirements that occur early in (or just prior to) a given pupil's school career will affect them over more time than those that occur later. We average cumulative retirements during the six grades g in which a pupil attends primary school:

$$\overline{\textit{total share 60}}_{mt} = \frac{1}{6}\sum_{g=1}^{6} R_{mtg} \quad (12)$$

with $R_g = \textit{share 60}_g + R_{g-1}$ and $R_1 = \textit{share 60}_{-2} + \textit{share 60}_{-1} + \textit{share 60}_1$.[12]

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 share retirements that occur just before a pupil enters the first grade and the two years before that. R_2 adds the share of teachers retiring just before a pupil enters second grade to the sum of share retirements experienced up to that point: $R_2 = \textit{share 60}_2 + R_1$, and so forth up to grade 6. In the empirical analysis, we report grade level results for the first stage that directly motivate the construction of this measure.

(see Appendix Figure A20).

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

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 requires 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. We test for such patterns in Section 8 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 plots the exit probability by age for all primary school teachers in our panel. 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-bargained retirement age of 60. In our estimation, we are only using variation from teacher exits that is due to teachers turning 60 years old.[13]

We start by illustrating the intuition of the first stage using raw data. Figure 4 displays the

[13] Appendix A 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 A2. 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.

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.

Figure 5 formalizes this intuition by running the first stage Equation 8 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 1 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.

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 relative to the quota period. 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.

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 Appendix Table A3, 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 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 when admitting more women and fewer men post-quota (see Figure 2) and corresponds to about a .3 SD increase in local teacher teams' exam grades. 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. 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 post-quota have a large negative 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 2, the first three columns report results for the pupil level first stage. The post-quota interaction with retirements measures by how much the average share of male (quota) teachers changes relative to the quota period 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).

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. 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).[14] In the centrally-organized application process, each pupil can submit

[14] Virtually everyone (99.7% of a cohort) successfully graduates from compulsory education (Virtanen, 2016). 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 higher ranked 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.

up to five choices for institution (and field), and a student-proposed deferred acceptance (DA) algorithm allocates available study slots. 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 pupils' 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 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.

Considering pupils' exposure to male (quota) teachers and retirements over their entire time in primary school, Table 2 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. 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 results in higher likelihood of pupils applying to post-compulsory education. The coefficients report the effect size associated with an increase 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 2.7 percentage points higher likelihood of applying, corresponding to a 3% increase over the mean of the dependent variable.

Table 3 reports IV results on the full set of outcomes regarding pupils' application timing and choices after compulsory schooling.[15] Having more male quota teachers makes pupils more likely

[15]We provide multiple sensitivity checks: Appendix Table A5 reports IV estimates for the main application outcomes in this section with cohort instead of region-by-cohort fixed effects, documenting that results are not sensitive to this choice. Appendix F.1 reports the full set of corresponding reduced form results and Appendix F.2 displays grade-level reduced form estimates. Results for the full set of mutually exclusive categories regarding

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. These patterns translate into higher enrollment rates in upper secondary education. Why are pupils who are exposed to more male quota teachers more successful in obtaining their top choices? In Appendix Table A8, we document that having more male quota teachers makes pupils apply more in line with attainable options. 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, Educational Attainment and Field of Study

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 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.

Table 4 presents IV results for educational attainment by examining the highest degree achieved by age 25 using mutually exclusive education categories.[16] We observe a shift towards higher attainment both in vocationally-oriented as well as in academic education paths.

which slot pupils obtain are reported in Appendix Table A7.

[16]Appendix Table A6 shows the first stage, reduced form and IV results for gradually adding in controls for the main outcome (Employed/Student) of this section. Reduced form results for long-run outcomes are reported in Appendix F. Appendix Figure A7 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.

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. 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 5 reports effects for mutually exclusive labor market status categories. Being exposed to more male quota teachers 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 3 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 for reasons other than disability.

Field choice: 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, male teachers could also more broadly motivate pupils to pursue different education fields.

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 70% within a field and degree level cell, we define the field as male or female leaning, and gender neutral otherwise. 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 general and primary school teacher in particular.[17]

Table 6 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 Table A12, the STEM

[17] 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. 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.

shift is similarly pronounced for both pupil genders.[18] Regarding teaching fields overall and primary teacher education specifically, we fail to reject a null effect. 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.

7 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 admission criteria disadvantage the underrepresented group as the composite score of the matriculation exam attaches considerably less weight to mathematics and natural science fields, in which men perform relatively better. We then show that scores are not directly related to productivity on the job. Teacher team performance on the exam is not predictive of pupil outcomes, suggesting that the score criterion inefficiently disadvantages men.

Third, we shed light on the particular attributes that may explain the disadvantaged groups' productivity in the context of this setting: Male teachers may serve as a role model for pupils in terms of their own career attachment and intrinsic motivation for their job – two factors which may inspire pupils to be more ambitious in their own pursuits, i.e. by continuing education, being more attached to the labor market and choosing fields with higher earnings. We show that these role model effects are not gender-specific as both girls and boys benefit similarly from exposure to male quota teachers. We further rule out that our main effects are driven by male quota teachers merely increasing pupils' grades. While providing a role model in terms of career attachment may explain higher performance among marginal candidates of the underrepresented group in the context of this study, the particular characteristics that determine output will be specific across different contexts for which policy makers consider more equal representation.

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

7.1 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%). Appendix Table A18 shows the reduced form for the main outcomes. 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.[19]

7.2 Imperfect Selection of Individual Talent

7.2.1 Admission Criteria Disadvantage the Underrepresented Group

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 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 language fields, and a weight of 25% on 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 8, 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. 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 (.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 A8. Among primary teachers, men perform worse on all fields compared to women, but the relative patterns are similar to the overall population: The gender score gap in mathematics and sciences

[19] This interpretation presumes that the gender composition in schools within the same municipality corresponds to the average municipal gender composition. We examine this assumption in Section 8. 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, but similar conclusions (Appendix Table A19).

is .25 grade points lower relative to the gap in languages.

The weighting of fields in the matriculation exam score thus discounted fields in which men performed relatively better when ranking primary teacher applicants. This score served as a main criterion for primary teacher admissions, and we show in Section 7.2.3.2 that the second round interview did not overcome this misalignment. 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.

7.2.2 Impact of Admission Criteria on Productivity

In Table 7 we examine the relationship between teachers' matriculation exam scores and "teacher team value-added" at the cohort-by-municipality level, exploiting variation in teacher team test performance that different pupil cohorts within the same municipality are exposed to. We estimate a precise zero 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 grades in language and math or sciences in column 2 and 3 show similarly negligible impacts for any field. 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 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.

7.2.3 Why Are Male (Quota) Teachers More Productive?

7.2.3.1 Career Attachment

We examine two proxies of career attachment for teachers, using our main panel of active primary school teachers from 1990-2000: Gender differences in career exit and compensation for additional responsibilities beyond regular teaching hours.

Table 8 documents gender differences in exit and leave taking among active teachers, restricting the sample to teachers below age 55 to not conflate career exits with retirement. Male teachers are less likely to exit the teaching career, defined as not returning to teaching from a leave that occurs in the first five years of the panel until the end of the ten year period (column 1). This differential drop-out is not explained merely via channels of family formation, as only a bit more than half of the gender difference in exits from the teaching career is accounted for by exits that are initiated following the birth of a child (column 2). Overall, male teachers are about 20% less likely to more permanently leave teaching as a career compared to women. Columns 3 and 4 show that male teachers are also less likely to go on leave in any given year, defined as switching from being an active teacher to not working as a teacher in the following year. However, this gender difference is almost entirely explained by female teachers being absent following a birth. We show in robustness section 8.2 that female teachers giving birth *per se* does not lead to worse student outcomes, pointing to team teaching structures being able to absorb such shocks. Instead, these leave and exit patterns are indicative of differential attachment to the profession with male teachers having a more sustained presence in their pupils' schools. We do not observe any gender differences in taking sick leave (column 5).

We next turn to examine residual earnings as a second proxy for career attachment. The salary scale of teachers is strictly deterministic, and overtime and additional responsibilities are compensated with supplemental payments above the regular payroll according to a fixed key (Statistics Finland, 1995; OAJ, 2023). We can thus examine gender differences in residual wage earnings that likely reflect differential engagement at schools in activities beyond regular teaching hours. We start by plotting the evolution of (raw) teacher earnings by age and gender in Appendix Figure A9, separately depicting women who never have children. Initially, the earnings profile of men and women who never have children evolve similarly, while women with children have a somewhat lower level (likely due to these being childbearing years). However, by age 40, the earnings of women who never have children fully converge with those of women with children, while male teachers' compensation levels are higher. Appendix Table A20 confirms these patterns, highlighting that male teachers' earnings are higher even when accounting for differential impacts of children or leave taking: Controlling for year and municipality fixed effects, male teachers on average earn about 3,000 EUR (approximately 3,000 USD) more (column 1), but there is no gender-specific return to experience defined as time since degree (column 2). This earnings gap persists even when restricting the sample to teachers who do not receive any government transfers (sick pay or parental leave pay) in a given year (column 3). Comparing only teachers who never have children in column 4, the gender gap in salary becomes smaller, but does not disappear and corresponds to men without children being compensated an additional 4.5% over their female counterparts. Men are more likely to become principal (column 5), but promotions may not be free of potential gender bias. Importantly, though, the salary gap is not explained by men being more likely to be promoted (column 6). Taken together, these patterns in teacher salary and teacher exit are suggestive of male teachers picking up additional responsibilities at school and being more attached to their profession, thus providing a role model for pupils in terms of their own career attachment and commitment to their profession.

7.2.3.2 Intrinsic Motivation

We next assess a proxy for male and female teachers' intrinsic motivation for this profession 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 by 2020.[20]

We start by documenting the relationship between matriculation exam score and the likelihood of obtaining a primary teaching degree by gender in Appendix Figure A10. For both male and female applicants, the likelihood of obtaining a primary teaching degree is increasing in score. Appendix Table A21 formalizes these patterns and confirms that there is not differential score premium by gender after the quota is abolished: 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 differentially reverse the impact of exam scores for men.[21]

We measure compensating differentials by comparing labor market earnings among similarly qualified applicants within gender:

$$Wage_i = \beta_0 + \beta_1\, Teacher_i + \beta_2 Male_i * Teacher_i + \beta_3\, Male_i + X_i + u_i \tag{13}$$

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 experience. Appendix Table 9 documents wage premia, 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.[22] These wage penalties are more pronounced when restricting the sample to applicants who ever obtain a university degree in columns 3 and 4. In Appendix Figures A11 and A12, 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, 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. As men are not accepted into primary teaching, they thus obtain fields with higher earnings as their next best alternative.

Taken together, these patterns suggest that among similarly scoring applicants, men possess

[20] University study slots in Finland are allocated in a centralized university application system by deferred acceptance. The year 2000 is the first time that data on application fields in the national application system is available. 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 use information on whether applicants ever obtain a primary teaching degree by 2020. Completion rates for primary teaching are at 90% (Nissinen and Välijärvi, 2011). Data on field choice for 2003 is missing and we exclude applicants in this year from the analysis.

[21] 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.

[22] In Appendix Figure A13a, we show that these wage penalties are present throughout the exam score distribution.

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 become primary teachers are willing to accept a wage penalty, highlighting their intrinsic motivation when choosing this stereotypically female profession.

7.2.3.3 Role Model Impacts by Pupil Gender

Do male quota teachers serve as a role model primarily for boys? While the main effects clearly demonstrate that the overall impact of male quota teachers was positive, this could mask heterogeneous effects by pupil gender. Figure 9 and Appendix Table A9 report heterogeneity by pupil gender for the main application outcomes at age 16. We run our main specification (Equation 10) with separate treatment effects for boys and girls while estimating controls and fixed effects jointly.[23] 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 A9: 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.[24]

7.2.3.4 Do Male Quota Teachers Increase Pupils' Grades?

We finally document limited evidence that male quota teachers increase their pupils' grades. This further strengthens the interpretation that effects operate through male (quota) teachers providing a more general role model of career attachment rather than opening up additional opportunities to pupils by increasing scores. We use two sources of pupil grades: Final grades obtained in middle school that determine which slot applicants obtain in their applications to post-compulsory education, and grades obtained in the nationally standardized matriculation exam. We observe GPA for 97% of the full sample of pupils and restrict the analysis of grades to those who apply to post-compulsory education. Field specific grades for mathematics and languages (the average of mother tongue and second national language) are available for all pupil cohorts except the last two (i.e. starting school before 1999). Grades for the matriculation

[23] When estimating these fixed effects jointly, the underlying assumption is that time-varying region-specific economic shocks affect e.g. the choice of whether to apply for boys and girls to a similar extent.

[24] 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. In line with this notion are studies that do not find evidence for same-gender match effects across a range of OECD countries (de Gendre et al., 2022; Cho, 2012; Holmlund and Sund, 2008).

exam are available for all cohorts, but we can only observe those 49% of the full sample who sat this exam.

Appendix Table A22 documents male quota teachers' impact on pupil grades, with coefficients consistently indistinguishable from zero, albeit imprecisely estimated. For middle school grades, we observe a noisy and negative impact on GPA across all subjects (column 1), with estimates for languages fields negative and estimates for mathematics closer to zero. Column 5 shows that exposure to male quota teachers increases the likelihood of observing middle school grades. For matriculation exam grades, the observed patterns are quite similar, with impacts on overall GPA noisy and negative and impacts on the likelihood of observing a pupil's exam positive. Across both exams these patterns suggest that — if anything — male quota teachers pull in pupils at the margin of the grade distribution to apply to post-compulsory education or sit the matriculation exam, with no discernible improvements in pupils' grades. Appendix Table A23 further documents precisely estimated null effects of teacher's own matriculation exam performance on their pupils grades in both middle school and the matriculation exam, consistent with results in Section 7.2.2.

8 Robustness

8.1 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 10 reports municipal level regressions, with all specifications assessing changes in flows for consistency (see Equation 9). Our goal is to understand whether teacher retirements in the post-quota period differentially impact municipalities' teacher hiring or exit strategies.

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, i.e. municipalities do not react to the policy reform by trying to recruit more teachers laterally. 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.

8.2 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 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 leave taking patterns due to maternity from more female teachers post-quota are unlikely to drive our results.

8.3 Placebo Test and Randomization Inference

To further assess robustness, we perform a placebo test in Appendix Figure A14, 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 A15 plots the distribution of reduced form treatment coefficients for applications (Equation 11), 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 and renders a p-value of 0.002.

8.4 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 (See, among others: De Chaisemartin and d'Haultfoeuille, 2020; Arkhangelsky et al., 2021; Sun and Abraham, 2020; Imai and Kim, 2021; Goodman-Bacon, 2021). 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. We conclude that treatment effect heterogeneity leading to sign reversal in $\hat{\beta}_{fe}$ is not a major concern in our setting.

8.5 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. We further assess whether the impacts of male (quota) teachers are plausibly linearly additive (Appendix Table A29). In small municipalities, the municipal and school level share of male teachers coincide and we document that treatment impacts are the same. We also discuss the main macro-economic shocks in Finland during our study period.

9 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 the study slots on top of their list, 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, have higher educational attainment as measured by their highest degree achieved, and are more likely to pursue education fields in STEM subjects.

Multiple pieces of evidence suggest that the quota in our setting reduced selection imperfections, even as it constrained admissions decisions. While selection criteria in absence of the quota were group-neutral 'on paper', in practice they discounted productivity-relevant attributes 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 whether a binding representation target is an optimal policy. However, the main trade-offs highlighted in this paper generalize to any setting in which more equitable representation is a policy goal. When a key criterion for choosing candidates discounts the abilities of an underrepresented group, conscious re-weighting of selection 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.

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, *30*(2), 175–187.

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.

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*.

Autor, D. H., & Scarborough, D. (2008). Does Job Testing Harm Minority Workers? Evidence from Retail Establishments. *The Quarterly Journal of Economics*, *123*(1), 219–277.

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.

Barnes, R. (2023). Supreme Court Rejects Race-based Affirmative Action in College Admissions [2023-06-29]. *Washington Post*.

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. (2023). Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes. *American Economic Journal: Applied Economics*, *15*(1), 26–67.

Bleemer, Z. (2021). *Top Percent Policies and the Return to Postsecondary Selectivity* (Research & Occasional Paper Series: CSHE. 1.2021). Center for Studies in Higher Education.

Bleemer, Z. (2022). Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209. *The Quarterly Journal of Economics*, *137*(1), 115–160.

Bohren, J. A., Hull, P., & Imas, A. (2022). *Systemic Discrimination: Theory and Measurement* (Working Paper No. 29820). 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. (2023). Labor Market Nationalization Policies and Exporting Firm Outcomes: Evidence from Saudi Arabia. *Economic Development and Cultural Change*, *71*(4), 1397–1426.

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.

de Gendre, A., Feld, J., Salamanca, N., & Zölitz, U. (2022). Do Same-Sex Teachers Affect Test Scores and Job Preferences?

Dhume, S. (2019). India's Ethnic Quotas Are a Cautionary Tale [2019-01-24]. *The Wall Street Journal*.

Dobbie, W., Hull, P., & Arnold, D. (2022). Measuring Racial Discrimination in Bail Decisions. *American Economic Review*, *112*(9).

Eckbo, B. E., Nygaard, K., & Thorburn, K. S. (2021). Valuation Effects of Norway's Board Gender-Quota Law Revisited. *Management Science*.

Etelä Suomen Sanomat. (1988). Koulutuspaikkojen Lisäzksellä Eroon Opettajapulasta [Getting Rid of Teacher Shortage by Increasing Training Places] [1988-08-25]. *Etelä Suomen Sanomat*.

Eurostat. (2021). Education and Training in the EU — Facts and Figures [Online; accessed 21-Sep-2021: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Education_and_training_in_the_EU_-_facts_and_figures].

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. *American Economic Journal: Economic Policy.*

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.

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.

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].

Li, D., Raymond, L. R., & Bergman, P. (2020). *Hiring as Exploration* (Working Paper No. 27736). National Bureau of Economic Research.

Liimatainen, S. (2002). Opettajankoulutuksen Valintayhteistyö [Selection for Teacher Education]. 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 [Male Quotas are Considered for Education and Social Work] [2012-10-06]. *Helsingin Sanomat*.

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 Markets. *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. (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]. *Tutkimusselosteita Koulutuksen Tutkimuslaitos*, (43).

OAJ. (2023). Työaika ja palkka – joka jäsenen sopimusopas [Working Time and Salary – Member's Contract Guide].

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. (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.

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. (2023). 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 [It Will Never Change! On the Difficulty of Reform of Student Selections for Primary Teacher Education]. *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* (Working Paper No. 30274). 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 [2019-08-30]. *Washington Post.*

Sethi, R., & Somanathan, R. (2023). Meritocracy and Representation. *Journal of Economic Literature*, *61*(3), 941–957.

Statistics Finland. (1995). Hinta- ja Palkkatiedote 1/95 [Price and Wage Bulletin 1/95].

Stock, J. H., & Yogo, M. (2002). *Testing for Weak Instruments in Linear IV Regression* (Working Paper No. 0284). National Bureau of Economic Research.

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].

Tasa-Arvovaltuutetu. (1992). Tasa-Arvovaltuutetun Toimisto 27/59/1992 [Office of the High Commissioner of Equality (Petäjäniemi, Tuulikki), Opinion on Quota based on request by Abo Akademi 27/59/1992, 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.

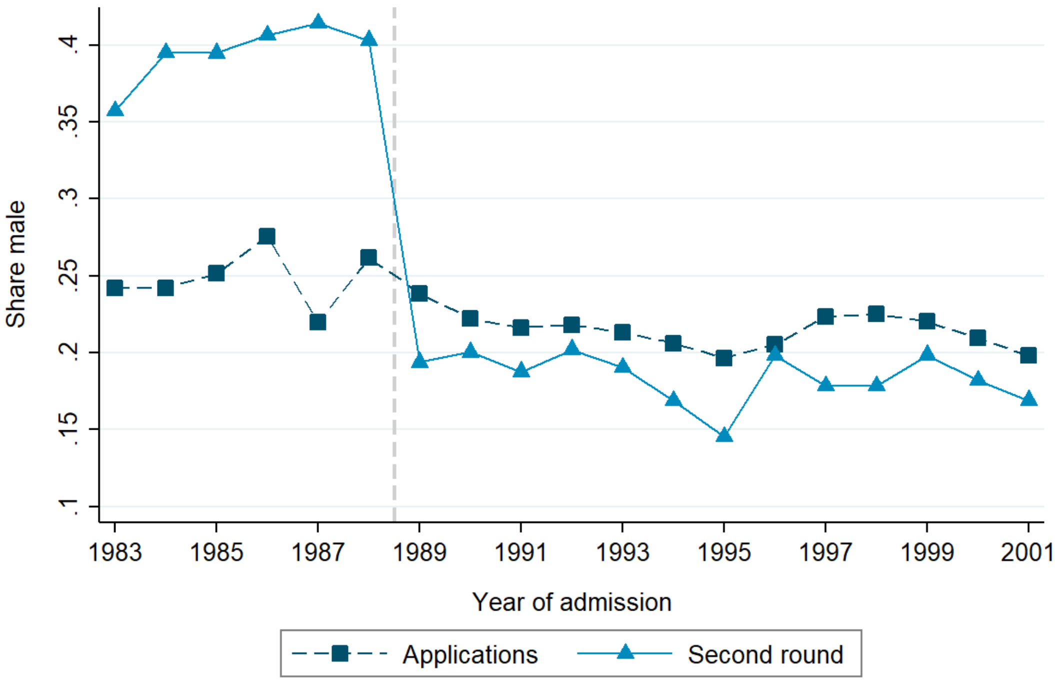
Valtiokonttori. (1988). Valintaoikeus: Valitsijan Opas [Selection Guide (State Pension System)].

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

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

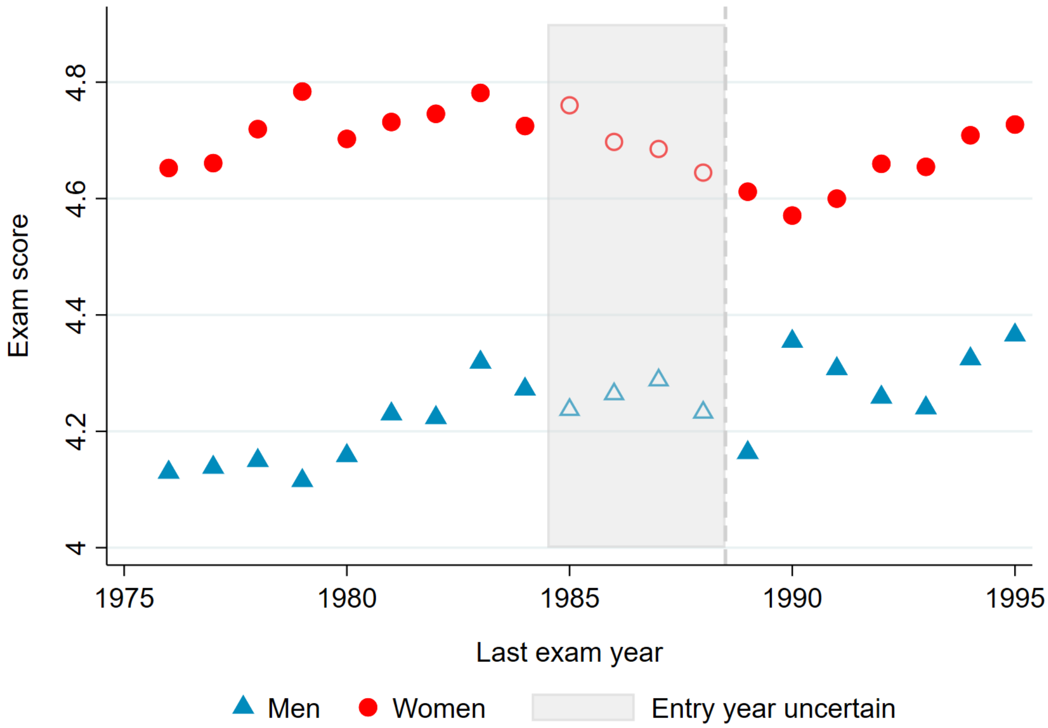
Tables and Figures

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)

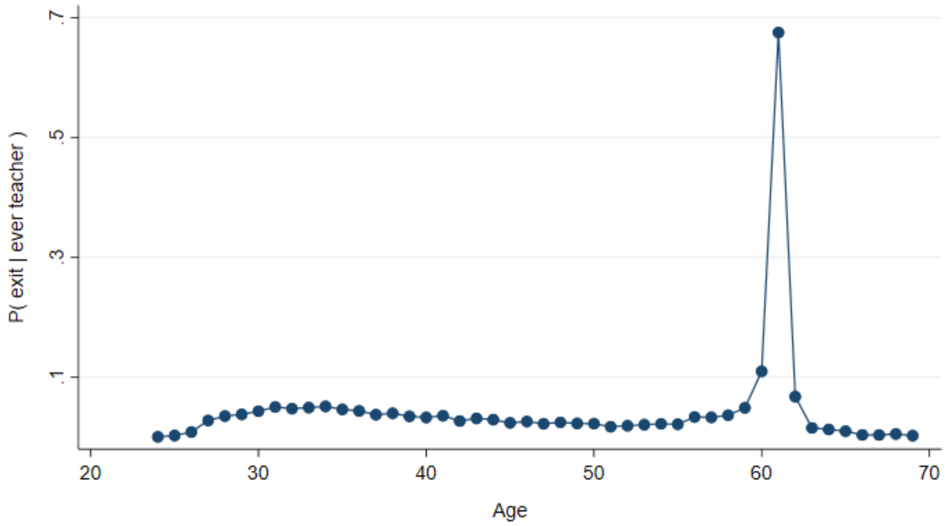
Figure 2: Matriculation Exam Score among Primary School Teachers



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 quota was abolished, but there is uncertainty with respect to the start date of exam takers in the years prior to 1989.

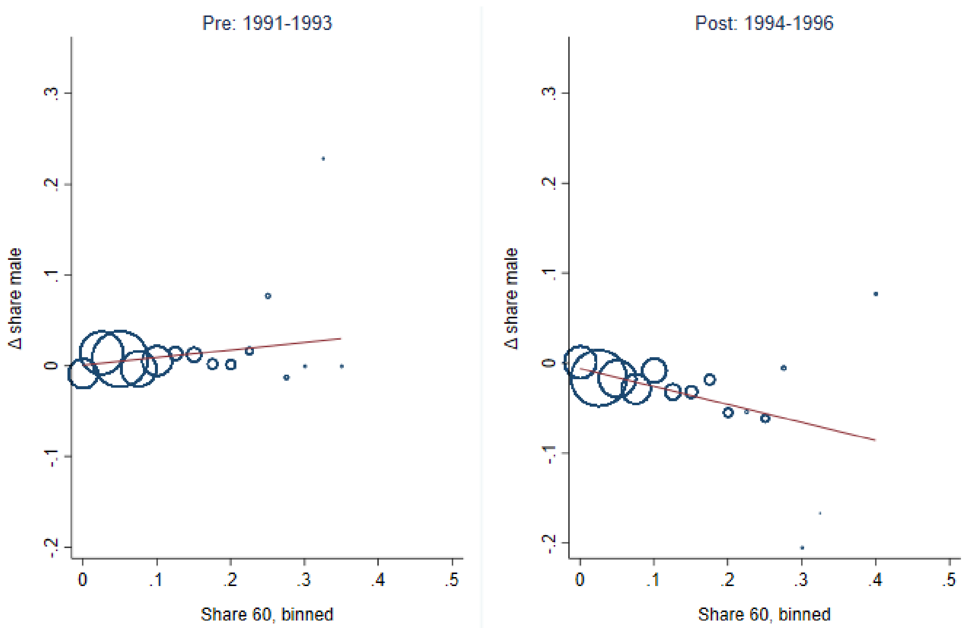
This figure plots the final score obtained in the *first* attempt at the exam in order to get at a measure of cognitive ability that is not influenced by repeated test taking. We plot the exam grade against the date of someone's *last* exam take to most closely approximate the point of entry to university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

Figure 3: Probability of Teacher Exit by Age



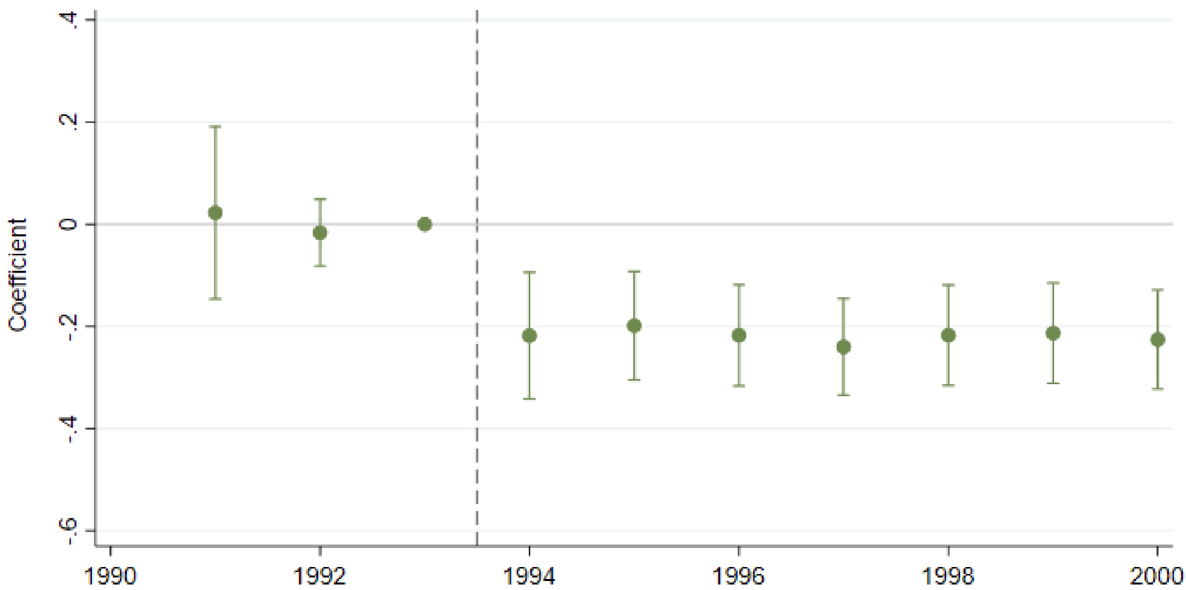
Note: Share of primary school teachers not working as a primary school teacher at a given age, conditional on having worked as a primary school teacher in the previous year. Data for all active primary school teachers in the years 1990-2000. Multiple exits per teacher possible. (back)

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 8, 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)

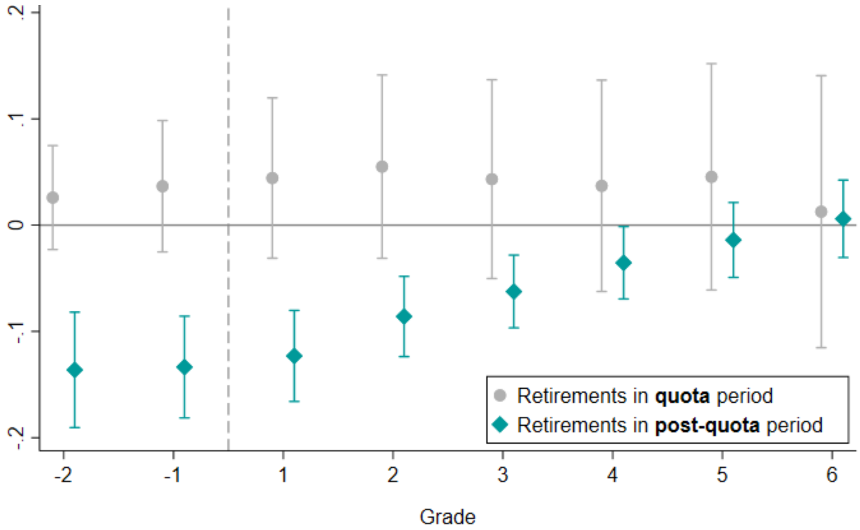
Table 1: First Stage at the Municipal Level

	First Differences				Fixed Effects		
	Δ Share Male				Share Male		
Share 60 * Post-Quota	−0.165	−0.170	−0.175	−0.161			
	(0.044)	(0.044)	(0.046)	(0.044)			
Share 60	0.062	0.062	0.070	0.072			
	(0.038)	(0.039)	(0.041)	(0.039)			
Total Share 60 * Post-Quota					−0.218	−0.243	−0.194
					(0.049)	(0.054)	(0.049)
Total Share 60					0.068	0.099	0.078
					(0.043)	(0.045)	(0.043)
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	0.022	0.869	0.867	0.869
Obs	4448	4448	4448	4448	4443	4443	4443
Dep mean	.0007	.0007	.0007	.0007	.3601	.3601	.3601

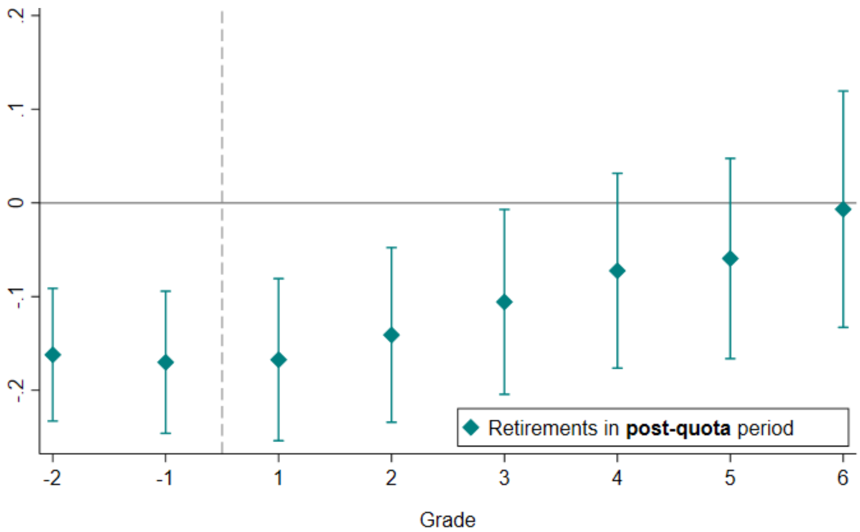
Note: Estimates for Equation 9 (columns 1-4): Year-on-year changes of the share of male teachers (Δ Share Male) on the share of teachers reaching retirement age (Share 60), and the corresponding fixed effects specification in Equation 8 (columns 5-7) of local share of male teachers on cumulative teacher retirement (Total Share 60). Observation counts between specifications change due to municipal consolidation. Dep mean reports mean of the dependent variable in the quota period, i.e. before 1994. 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. (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 (Equation 11). 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)).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

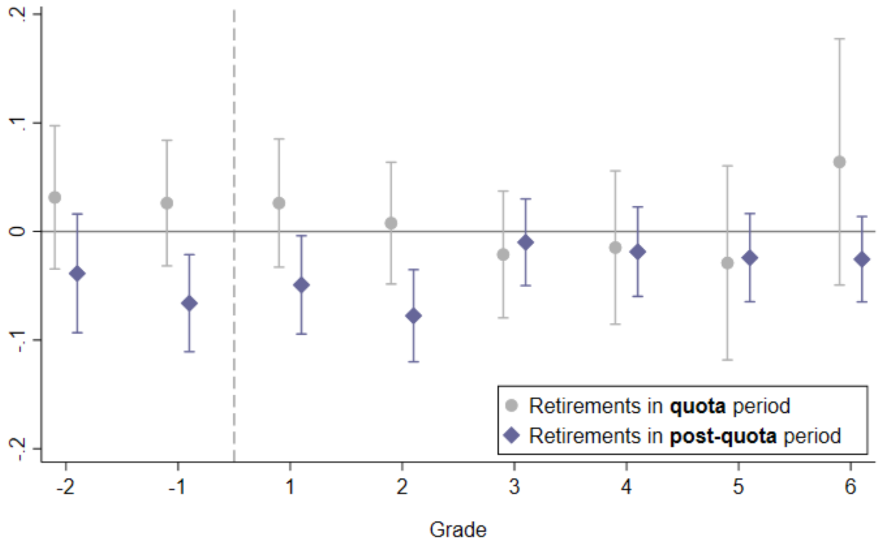
Table 2: First Stage, Reduced Form and IV: Applications for Post-Compulsory Education

	First Stage			RF			IV		
	Avg Share Male			Apply			Apply		
Avg Share Male							0.547	0.699	0.421
							(0.202)	(0.235)	(0.196)
Total Share 60 *	–0.176	–0.176	–0.168	–0.096	–0.096	–0.058			
Post-Quota	(0.042)	(0.042)	(0.043)	(0.026)	(0.026)	(0.027)			
Total Share 60	0.033	0.033	0.043	0.049	0.049	0.019	0.031	0.025	–0.001
	(0.036)	(0.036)	(0.037)	(0.020)	(0.020)	(0.021)	(0.025)	(0.027)	(0.020)
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	0.038	0.038			
Obs	825,095	825,095	825,095	825,095	825,095	825,095	825,095	825,095	825,095
Dep mean	.313	.313	.313	.911	.911	.911	.911	.911	.911

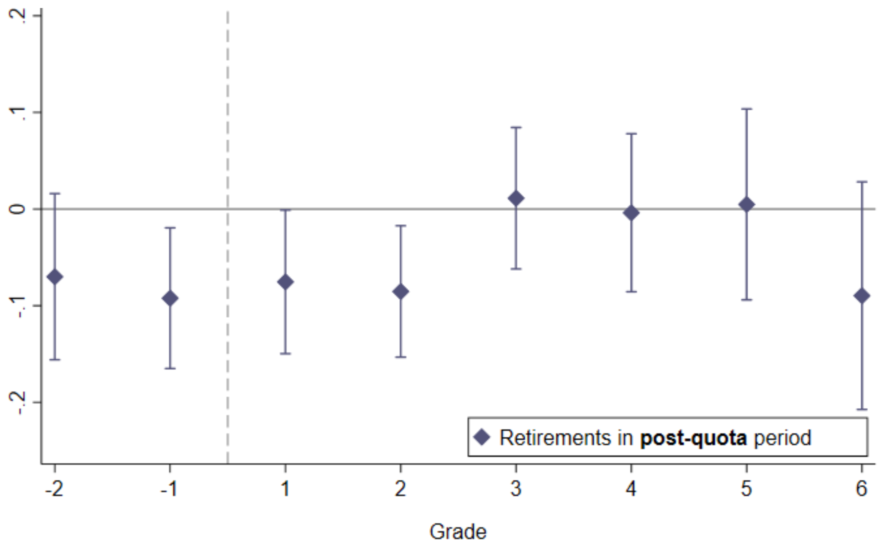
Note: Columns 1-3 show estimates for Equation 11 with the average share male teachers pupils are exposed to during primary school as the outcome. Columns 4-6 show reduced form estimates (corresponding to Equation 11), and Columns 7-9 show IV estimates of Equation 10, with a pupil applying directly in the spring of the year they turn 16 (i.e. the last year of middle school) as the outcome. Individual-level controls measured at age 7: 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. (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 (Equation 11). 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)).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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: IV Estimates: Applications and Enrollment for Post-Compulsory Education

	Apply directly	Apply late	Apply never	Top choice	Enrolled at 16	Enrolled ever
Avg Share Male	0.421	−0.345	−0.076	0.530	0.608	0.124
	(0.196)	(0.176)	(0.073)	(0.246)	(0.309)	(0.074)
MOP F^{eff}	15.28	15.28	15.28	15.28	13.00	13.00
Obs	825,095	825,095	825,095	825,095	695,341	695,341
Dep mean	.911	.066	.023	.858	.861	.98
Std effect	.094	-.088	-.032	.097	.11	.055

Note: IV estimate of Equation 10.

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 choices in the application (Top 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). Data on enrollment available for cohorts starting school in 1990 and after (see Section 4).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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: 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	–0.439	0.386	–0.070
	(0.154)	(0.260)	(0.208)	(0.211)	(0.228)	(0.146)	(0.093)
MOP F^{eff}	15.36	15.36	15.36	15.36	15.36	15.36	15.36
Obs	810,066	810,066	810,066	810,066	810,066	810,066	810,066
Dep mean	.127	.316	.108	.146	.211	.054	.038
Std effect	-.032	-.008	.088	-.014	-.068	.108	-.023

Table 5: 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)
MOP F^{eff}	15.37	15.37	15.37	15.37
Obs	811,393	811,393	811,393	811,393
Dep mean	.842	.086	.017	.053
Std effect	.089	-.009	-.061	-.093

Note: IV estimates of Equation 10.

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. 1327 observations have no degree information.

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.

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level.

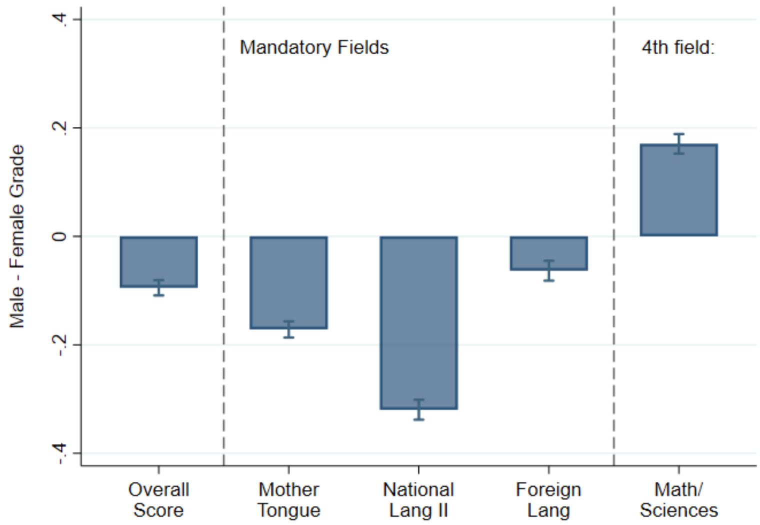
(back)

Table 6: IV Estimates: Field of Education at Age 25

	Male	Neutral	Female	STEM	STEM-M	Education/ Teacher	Primary Teacher
Avg Share Male	0.302	–0.499	0.197	0.595	0.707	–0.013	0.063
	(0.229)	(0.286)	(0.191)	(0.273)	(0.323)	(0.073)	(0.049)
MOP F^{eff}	15.37	15.37	15.37	15.37	15.37	15.37	15.37
Obs	811,393	811,393	811,393	811,393	811,393	811,393	811,393
Dep mean	.303	.433	.264	.264	.379	.023	.011
Std effect	.042	-.064	.028	.086	.093	-.006	.039

Note: IV estimates of Equation 10. Outcomes from left to right:
Field is 'Male' dominated (> 70% male), (gender) 'Neutral' or 'Female' dominated (> 70% female), based on previous generation. Field is STEM or STEM + Medicine (STEM-M). Field is Education Science or Teacher. Field is Primary School Teacher. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 8: 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. Gender difference in score ("Overall Score") and in each of the exam fields 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 the high school track (lukion opiskelija) under age 22 for the years 1983 - 1985. 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 7: Teacher Team Performance and Teacher Exam Scores

	Apply				Employed/Student			
Overall GPA	−0.0040				0.0012			
	(0.0061)				(0.0082)			
Language		−0.0073	−0.0085			−0.0057	−0.0060	
		(0.0063)	(0.0068)			(0.0080)	(0.0084)	
Math or Science		0.0043				0.0081		
		(0.0053)				(0.0060)		
Math			0.0027				0.0006	
			(0.0032)				(0.0035)	
Science			0.0027				0.0064	
			(0.0052)				(0.0057)	
Math Background				0.0016				0.0236
				(0.0152)				(0.0157)
Adj. R^2	0.070	0.070	0.070	0.070	0.025	0.025	0.025	0.025
Obs	825,033	825,033	825,033	825,033	811,332	811,332	811,332	811,332
Dep mean	.911	.911	.911	.911	.842	.842	.842	.842

Note: Estimates of pupil outcomes on teachers' exam scores in the matriculation exam. Outcomes are pupils' application likelihood and employment status (see Tables 3 and 5). 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 8: Career Attachment

	Exit	Exit (Birth)	Leave	Leave (Birth)	Sick Leave
Male	–0.0073	–0.0042	–0.0169	–0.0147	–0.0006
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Adj. R^2	0.013	0.003	0.010	0.008	0.002
Obs	21,482	21,482	186,586	186,586	105,224
Dep mean	.037	.008	.041	.015	.016
Dep mean men	.032	.005	.03	.004	.016
Dep mean women	.04	.01	.047	.019	.016

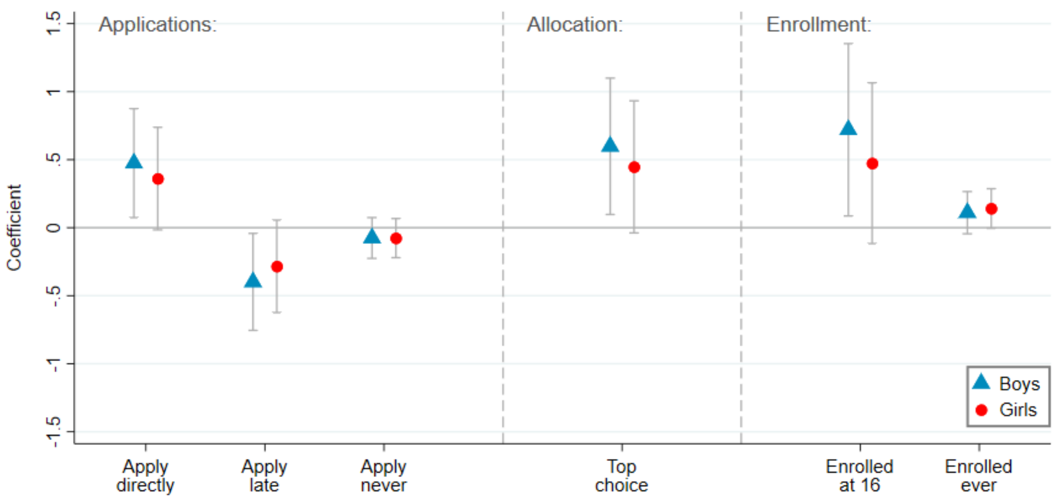
Note: Regressions of outcomes related to leave taking for the panel of active primary teachers (ages 26-55), on an indicator for being Male. Outcomes from left to right: Exit: Defined as switching from being an active teacher to not being a teacher in the next year at any point within the first five years of the panel, and not returning as an active teacher until the end of the panel. Exit (Birth): Exit for which either the year just before or the first year of leave coincides with the birth of a child. Leave: Defined as switching from being an active teacher to not being a teacher in the next year in any given year. Leave (Birth): Leave for which either the year just before or the first year of leave coincides with the birth of a child. Sick Leave: Receiving payments for sick leave in any given year, data available for 1995-2000 only. All specifications include year and municipality fixed effects. (back)

Table 9: Wage Premia for Primary Teaching Degree Holders by Gender

	(1)	(2)	(3)	(4)
Teaching Degree	3,397	3,381	2,739	2,719
	(304)	(305)	(300)	(300)
Male * Teaching Degree	–4,972	–4,887	–6,779	–6,659
	(646)	(654)	(652)	(657)
Male	11,213	12,379	13,010	15,244
	(336)	(1,420)	(376)	(1,573)
Score	1,362	1,418	978	1,074
	(134)	(149)	(141)	(156)
Male * Score		–279		–525
		(330)		(359)
Adj. R^2	0.098	0.098	0.102	0.102
Obs	20,430	20,430	17,857	17,857
Dep mean	40,224	40,224	41,456	41,456

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 (time 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. (back)

Figure 9: IV Estimates: Applications and Enrollment for Post-Compulsory Education by Pupil Gender



Note: IV estimate of Equation 10.

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 choices in the application (Top 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).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 10: Exit and Hiring Patterns in Municipalities

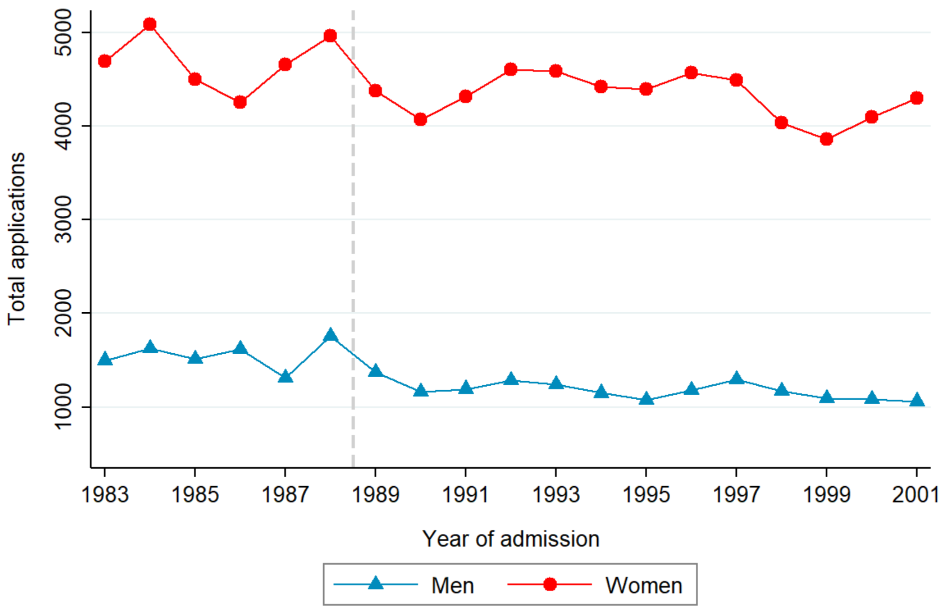
	Leave	Leave ≤ 59	Δ Age	Δ Time since degree	First entrants
Share 60 * Post-Quota	0.041	–0.001	–1.300	–1.243	–0.043
	(0.067)	(0.070)	(1.387)	(1.341)	(0.375)
Share 60	0.873	0.046	–16.088	–15.355	0.335
	(0.061)	(0.062)	(1.215)	(1.172)	(0.315)
Adj. R^2	.176	.008	.222	.206	.038
Obs	4448	4448	4448	4448	3746
Dep mean	.1	.08	-.21	-.21	.35

Note: Estimates for Equation 9. Outcomes from left to right: 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). All specifications include year fixed effects. Standard errors clustered at the municipality level. Regressions weighted by population, means unweighted.
(back)

For Online Publication: Appendix Tables and Figures

A Summary Statistics

Figure A1: Total Applications by Gender



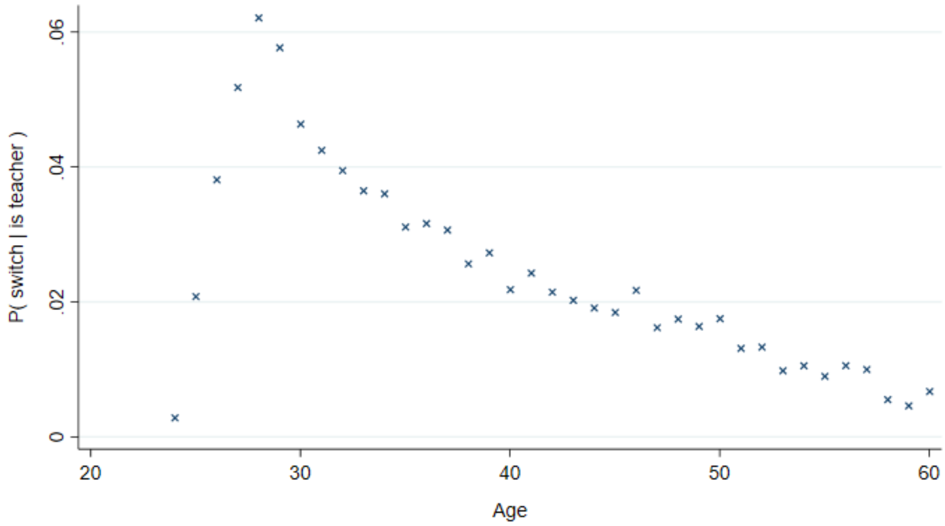
Note: Total number of male and female applicants to primary teacher studies. *Source:* Liimatainen (2002). (back)

Table A1: 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	0.613	0.590	-0.022
			(0.007)			(0.007)
Rural residence at birth	0.371	0.392	0.021	0.359	0.376	0.017
			(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 obs	14,995	7,298		18,074	7,887	
Education path (born after 1952)						
High school degree	0.980	0.979	-0.002	0.978	0.971	-0.007
			(0.003)			(0.003)
Teaching degree	0.897	0.891	-0.006	0.905	0.886	-0.018
			(0.006)			(0.005)
Primary teaching degree	0.845	0.870	0.025	0.833	0.855	0.022
			(0.007)			(0.006)
Age at high school degree	19.23	19.47	0.24	19.19	19.40	0.21
			(0.02)			(0.01)
Age at teaching degree	26.45	27.04	0.59	26.84	27.51	0.67
			(0.12)			(0.09)
Age at primary teacher degree	25.21	26.37	1.16	25.87	26.88	1.01
			(0.08)			(0.07)
Academic performance (born after 1952)						
Matriculation exam	0.986	0.984	-0.003	0.982	0.976	-0.006
			(0.003)			(0.002)
Overall GPA, first take	4.78	4.22	-0.56	4.72	4.22	-0.49
			(0.02)			(0.01)
Overall GPA, 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 obs	7,053	3,273		11,701	4,614	

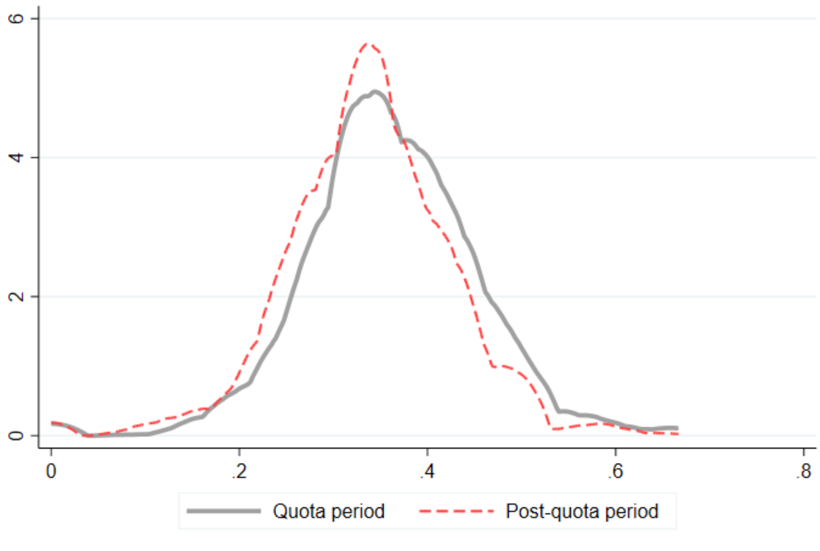
Note: Characteristics of male and female primary school teachers who are active teachers for at least one year in the quota period (1990-93) or in the post-quota period (1994-2000) and who are between 24 and 60 years old. (back)

Figure A2: Probability of Switching Municipality of Work for Active Teachers by Age



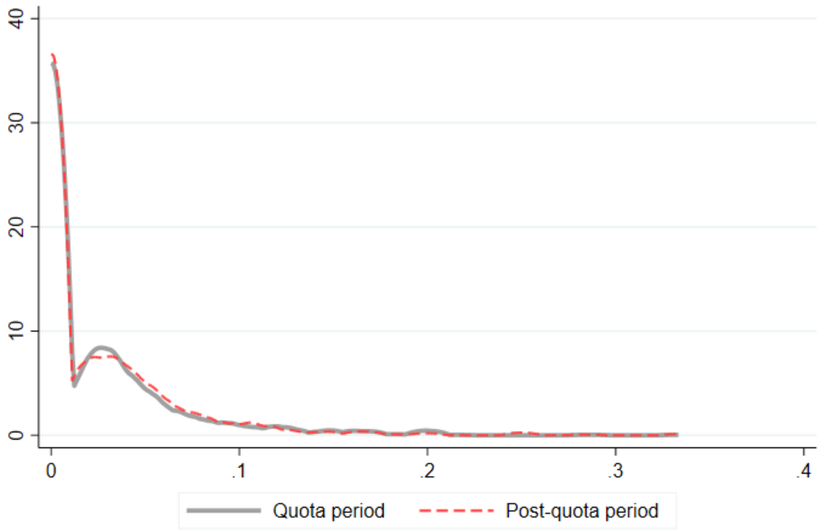
Note: Share of primary school teachers having switched the municipality in which they are working as a primary school teacher at a given age, conditional on having worked as a primary school teacher in the previous year. Data for all active primary school teachers in the years 1990-2000. (back)

Figure A3: Distribution of Share Male Primary Teachers



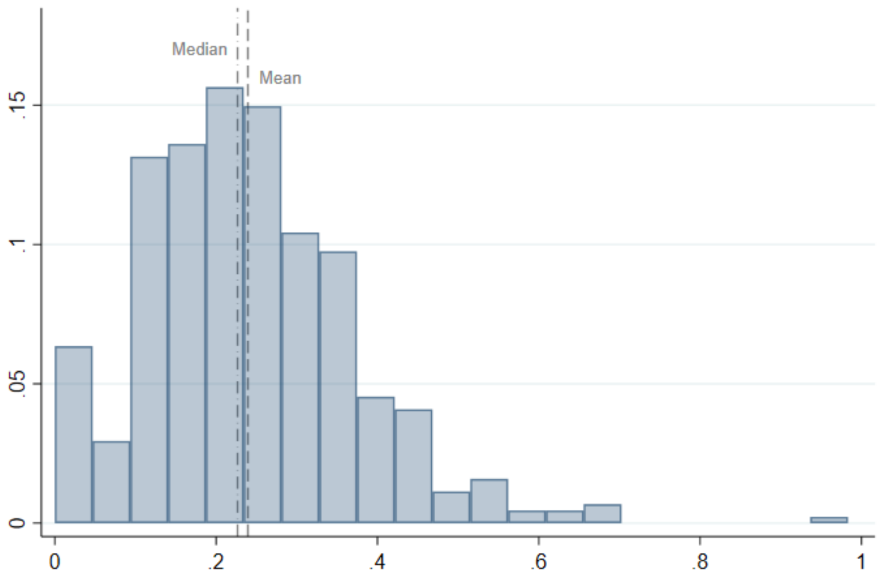
Note: Smoothed density of yearly share of male primary teachers at the municipal level, separately by years in the quota period (1991-93) and post-quota period (1994-2000). (back)

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 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)

Table A2: Summary Statistics for Main Treatment and Outcome Variables

Variable	Mean	SD
Treatment		
Municipality Level		
Share Male	0.344	0.090
Share 60	0.024	0.039
Share Male, 1990	0.359	0.105
Share Male, 2000	0.309	0.082
Pupil Level		
Avg Share Male	0.313	0.064
Total Share 60	0.078	0.056
Outcomes		
Applications and Enrollment		
Apply	0.911	0.285
Apply late	0.066	0.249
Apply never	0.023	0.150
Top Choice	0.858	0.349
Enrolled at 16	0.861	0.346
Enrolled ever	0.980	0.142
Labor Market		
Employed/Student	0.842	0.364
Unemployed	0.086	0.281
DI/Pension	0.017	0.130
Other out of LF	0.053	0.223
Highest Degree		
Compulsory schooling	0.127	0.333
Vocational Sec	0.316	0.465
Vocational Sec Plus	0.108	0.310
Vocational Tert	0.146	0.353
Academic Sec	0.211	0.408
Academic Tert: BA	0.054	0.227
Academic Tert: MA	0.038	0.190
Field of Study		
Male	0.303	0.460
Neutral	0.433	0.495
Female	0.264	0.441
STEM	0.264	0.441
STEM-M	0.379	0.485
Education/Teacher	0.023	0.150
Primary Teacher	0.011	0.102

Note: Summary statistics for main treatment and outcome variables. Municipality Level refers to the municipal-by-year panel, while Pupil Level refers to exposure measures throughout primary school for the pupil panel.

B Conceptual Framework: Appendix

B.1 Proofs and Derivations for Section 3.2

Probability of getting admitted (Equation 5): A candidate of group g is admitted if:

$$
\begin{aligned}
s_g &> s_g^* \\
\frac{s_g - \mu_{s_g}}{\sigma_s} &> \frac{s_g^* - \mu_{s_g}}{\sigma_s} \\
&= \frac{\rho}{\sigma_a}\left(s_g^* - \mu_{a_g} + b_g\right) \\
&\equiv z_g^*
\end{aligned}
$$

The mass of candidates admitted for group g is thus equal to $1 - \Phi(z_g^*)$.

Mean expected ability among admits (Equation 6)
We can rewrite $a_g = \mu_{a_g} + \xi_{a_g}$ with $\xi_{a_g} \sim \mathcal{N}(0, \sigma^2_{\xi_{a_g}})$ such that $E[a_g|s_g > s_g^*] = \mu_{a_g} + E[\xi_{a_g}|s_g > s_g^*]$ and $E[\xi_{a_g}|s_g]$ a linear regression of ξ_{a_g} on s_g. Applying iterated expectations $E(\xi_{a_g}) = E_s[E(\xi_{a_g}|s)]$, we thus get that:

$$
\begin{aligned}
E[a_g|s_g > s_g^*] &= \mu_{a_g} + \sigma_a\, \rho\, E\left[\frac{s_g - \mu_{s_g}}{\sigma_s} \mid \frac{s_g - \mu_{s_g}}{\sigma_s} > \frac{s_g^* - \mu_{s_g}}{\sigma_s}\right] \\
&= \mu_{a_g} + \sigma_a\, \rho\, \lambda(z_g^*)
\end{aligned}
$$

with $z_g^* \equiv \dfrac{s_g^* - \mu_{s_g}}{\sigma_s}$ and $\lambda(\cdot) = \dfrac{\phi(\cdot)}{1 - \Phi(\cdot)}$ the inverse Mills Ratio.

B.2 Proofs and Derivations for Section 3.3

Recall that Test 2 is more biased against the underrepresented group ($b_{M_2} > b_{M_1} > b_F$), but delivers a more precise signal of ability ($\rho_2 > \rho_1$) and that we set $\mu_a = \mu_{a_F} = \mu_{a_M} = 0$ and $b_M > b_F = 0$ in the following. We further assume that selection of candidates from each group is from the right tail of the distribution.

B.2.1 Unconstrained Admissions

When an admissions office is unconstrained, optimal threshold scores take into account any group-specific bias (Equation 4): $s^*_{M_j} = s^*_{F_j} - b_{M_j}$, such that:

$$
z^*_{M_j} = z^*_{F_j} = \frac{\rho_j}{\sigma_a}\, s^*_{F_j}
$$

The capacity constraint and total expected ability of admits are, respectively:

$$2\,(1-\Phi(z^*_{F_j})) = c \qquad\qquad \Upsilon = 2\,(1-\Phi(z^*_{F_j}))\,\rho_j\,\sigma_a\,(z^*_{F_j}) = 2\,\rho_j\,\sigma_a\,\phi(z^*_{F_j})$$

In the following, we drop the test subscript j for better readability.

B.2.1.1 Increasing bias

Note that since neither the capacity constraint nor total expected ability of admits depend on b_M, an increase in bias will not change the mass of candidates admitted from either group and has no impact on total expected ability. This is because an unconstrained admissions office can directly off-set an increase in bias by adjusting the group-specific cut-off scores.

B.2.1.2 Increasing precision

Representation: Taking the derivative of the capacity constraint with respect to ρ, we obtain:

$$-2\,\phi(z^*_F)\,\frac{dz^*_F}{d\rho} = 0$$

Since $\phi(\cdot) > 0$, it must be that $\dfrac{dz^*_F}{d\rho} = 0$, and thus the mass of admitted male and female candidates does not change when increasing the precision of the test.

Total Expected Ability of Admits: Taking the derivative of Υ with respect to ρ, we obtain:

$$2\,\sigma_a\,\phi(z^*_F) + 2\,\rho\,\sigma_a\,\phi'(z^*_F)\,\frac{dz^*_F}{d\rho} > 0$$

since $\phi(\cdot) > 0$ and $\dfrac{dz^*_F}{d\rho}$ is equal zero. Total expected ability for both groups therefore increases when precision of the test increases.

B.2.2 No Affirmative Action

B.2.2.1 Increasing Bias

Representation: Differentiating the capacity constraint $(1-\Phi(z^*_M)) + (1-\Phi(z^*_F)) = c$ with respect to b_M, we obtain:

$$\begin{aligned} -\phi(z^*_M)\,\frac{dz^*_M}{db_M} - \phi(z^*_F)\,\frac{dz^*_F}{db_M} &= 0 \\ -\phi(z^*_M)\,\frac{\rho}{\sigma_a}\left(\frac{ds^*}{db_M} + 1\right) &= \phi(z^*_F)\,\frac{\rho}{\sigma_a}\,\frac{ds^*}{db_M} \end{aligned} \tag{14}$$

We can assess in which the direction the score cut-off moves when bias increases by rearranging the above Equation 14:

$$\frac{ds^*}{db_M} = -\frac{\phi(z_M^*)}{\phi(z_M^*) + \phi(z_F^*)} < 0$$

since $\phi(\cdot) > 0$. Thus, the RHS of Equation 14 is negative and $\frac{dz_M^*}{db_M} > 0$. The mass of admitted men falls when bias increases, with the mass of women increasing by the same amount.

Total Expected Ability of Admits: Taking the derivative of Υ with respect to b_M and using the properties that $\frac{dz_F^*}{db_M} = -\frac{\phi(z_M^*)}{\phi(z_F^*)}\frac{dz_M^*}{db_M}$ (see Equation 14) and $\frac{\phi'(x)}{\phi(x)} = -x$, we obtain:

$$\begin{aligned}
&\rho\,\sigma_a\,\phi'(z_M^*)\,\frac{dz_M^*}{db_M} + \rho\,\sigma_a\,\phi'(z_F^*)\,\frac{dz_F^*}{db_M} \\
&= \rho\,\sigma_a\,\left(\phi'(z_M^*)\,\frac{dz_M^*}{db_M} - \phi'(z_F^*)\frac{dz_M^*}{db_M}\,\frac{\phi(z_M^*)}{\phi(z_F^*)}\right) \\
&= \rho\,\sigma_a\,\left(\frac{\phi'(z_M^*)}{\phi(z_M^*)}\,\frac{dz_M^*}{db_M}\phi(z_M^*) - \frac{\phi'(z_F^*)}{\phi(z_F^*)}\,\frac{dz_M^*}{db_M}\phi(z_M^*)\right) \\
&= -\,\rho\,\sigma_a\,\frac{dz_M^*}{db_M}\phi(z_M^*)\,(z_M^* - z_F^*) \\
&= -\,b_M\,\rho^2\,\frac{dz_M^*}{db_M}\phi(z_M^*) < 0
\end{aligned}$$

Total expected ability of admits thus falls when bias b_M increases.

B.2.2.2 Increasing Precision

Representation Differentiating the capacity constraint $(1 - \Phi(z_M^*)) + (1 - \Phi(z_F^*)) = c$ with respect to ρ, we obtain:

$$-\phi(z_M^*)\left(\frac{z_M^*}{\rho} + \frac{ds^*}{d\rho} * \frac{\rho}{\sigma_a}\right) = \phi(z_F^*)\left(\frac{z_F^*}{\rho} + \frac{ds^*}{d\rho} * \frac{\rho}{\sigma_a}\right) \tag{15}$$

We can assess in which direction the score cut-off moves when precision increases by rearranging the above Equation 15:

$$\frac{ds^*}{d\rho} = -\frac{\sigma_a\,(\phi(z_M^*)\,z_M^* + \phi(z_F^*)\,z_F^*)}{\rho^2\,(\phi(z_M^*) + \phi(z_F^*))} < 0$$

since admissions are from the right tail of the distribution. The score cut-off thus decreases as precision increases. Intuitively, a test with higher precision will decrease the variance of the score distribution, such that the tails are thinner. Admitting the same mass of candidates therefore implies a lower s^*.

We can show by contradiction that the mass of men (LHS of Equation 15) falls as precision increases. Note first that it is not possible for the change in the mass of admits of both groups to be zero ($\Delta N_M = -\Delta N_F = 0$), since $\phi(\cdot) > 0$ and $z^*_M > z^*_F$.
Suppose the LHS of Equation 15 were > 0, then it must be that

$$z^*_M < -\frac{ds^*}{d\rho} * \frac{\rho^2}{\sigma_a} \qquad \text{and} \qquad z^*_F > -\frac{ds^*}{d\rho} * \frac{\rho^2}{\sigma_a}$$

But this is not possible since for any test, $z^*_M > z^*_F$. It therefore follows that the LHS of Equation 15 is < 0 (and thus $\frac{dz^*_M}{d\rho} > 0$), i.e. the mass of men **falls** as the precision of the test increases.

Total Expected Ability of Admits: Taking the derivative of Υ with respect to ρ, we obtain:

$$\begin{aligned}
&\sigma_a\, \phi(z^*_M) + \sigma_a\, \phi(z^*_F) + \rho\, \sigma_a\, \phi'(z^*_M)\, \frac{dz^*_M}{d\rho} + \rho\, \sigma_a\, \phi'(z^*_F)\, \frac{dz^*_F}{d\rho} \\
&= \sigma_a \left(\phi(z^*_M) + \phi(z^*_F)\right) - b_M\, \rho^2\, \frac{dz^*_M}{d\rho}\, \phi(z^*_M)
\end{aligned}$$

The first term in the above is positive while the second term is negative, such that the net impact of increasing precision is ambiguous. Intuitively, increasing precision means getting a better signal of ability among those admitted from both groups. However, because the test exhibits a bias against men, increasing precision also incurs a cost by under-admitting candidates from the male group and over-admitting female candidates.

C First Stage: Teachers' Matriculation Exam Scores

Table A3: First Stage: Teachers' Matriculation Exam Scores

	Overall Score	Language	Best of Math or Sciences	Math	Sciences
Total Share 60 * Post-Quota	0.08	0.06	0.15	0.11	0.11
	(0.25)	(0.27)	(0.29)	(0.51)	(0.33)
Total Share 60	–0.08	–0.06	–0.13	–0.31	–0.02
	(0.22)	(0.24)	(0.26)	(0.45)	(0.30)
Adj. R^2	0.84	0.86	0.82	0.81	0.82
Obs	4358	4358	4358	4347	4348
Dep mean	4.42	4.3	4.78	4.01	4.57

Note: Estimates for Equation 8 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, see Section 7.2.1. 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. All specifications include year and municipality-by-post-quota fixed effects. Standard errors clustered at the municipality level. Regressions weighted by population. (back)

Table A4: First Stage in Teacher Score Sample

Total Share 60 * Post-Quota	–0.222
	(0.050)
Total Share 60	0.067
	(0.043)
Adj. R^2	0.870
Obs	4358

Note: Estimates for Equation 8 for restricted sample of municipalities with at least one teacher test score observable in any year (A3). Specification includes year and municipality-by-post-quota fixed effects. Standard errors clustered at the municipality level. Regression weighted by population. (back)

D Additional IV Estimates

D.1 Sensitivity Main Results

Table A5: IV Estimates: Applications and Enrollment for Post-Compulsory Education
Municipal and Cohort Fixed Effects only

	Apply directly	Apply late	Apply never	Top choice	Enrolled at 16	Enrolled ever
Avg Share Male	0.699	–0.540	–0.159	0.714	1.071	0.213
	(0.235)	(0.188)	(0.083)	(0.280)	(0.384)	(0.087)
MOP F^{eff}	17.64	17.64	17.64	17.64	14.40	14.40
Obs	825,095	825,095	825,095	825,095	695,341	695,341
Dep mean	.911	.066	.023	.858	.861	.98
Std effect	.156	-.138	-.068	.13	.193	.094

Note: IV estimate of Equation 10.

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 choices in the application (Top 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).

All specifications include year and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A6: First Stage, Reduced Form and IV: Employed/Student at Age 25

	First Stage			RF			IV		
	Avg Share Male			Employed/Student			Employed/Student		
Avg Share Male							0.297	0.403	0.512
							(0.215)	(0.216)	(0.243)
Total Share 60 *	–0.177	–0.177	–0.168	–0.053	–0.053	–0.076			
Post-Quota	(0.042)	(0.042)	(0.043)	(0.033)	(0.033)	(0.033)			
Total Share 60	0.034	0.034	0.044	0.031	0.031	0.040	0.021	0.019	0.020
	(0.036)	(0.036)	(0.037)	(0.028)	(0.028)	(0.030)	(0.026)	(0.027)	(0.027)
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	0.008	0.008			
Obs	811,393	811,393	811,393	811,393	811,393	811,393	811,393	811,393	811,393
Dep mean	.313	.313	.313	.842	.842	.842	.842	.842	.842

Note: Columns 1-3 show estimates for Equation 11 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 11), and Columns 7-9 show IV estimates of Equation 10 with being either employed or a student at age 25 as the outcome. Individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

D.2 Additional Results on Applications:

Table A7: Obtained Choices in Post-Compulsory Applications

	First	Second	Third	Fourth	Fifth	Switch	No Spot	Apply Never
Avg Share Male	0.321	0.209	–0.136	–0.105	0.011	–0.113	–0.110	–0.076
	(0.289)	(0.162)	(0.097)	(0.066)	(0.039)	(0.091)	(0.103)	(0.073)
MOP F^{eff}	15.28	15.28	15.28	15.28	15.28	15.28	15.28	15.28
Obs	825,095	825,095	825,095	825,095	825,095	825,095	825,095	825,095
Dep mean	.773	.085	.035	.015	.007	.019	.043	.023
Std effect	.049	.048	-.047	-.055	.008	-.052	-.035	-.032

Note: IV estimates of Equation 10. 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).
All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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: By Gender: Aspirations for Post-Compulsory Education

	Apply	Choose:		Obtain:		
	never	any Voc	only Acad	no spot	Voc	Acad
Boys * Avg Share Male	–0.076	0.771	–0.701	–0.202	0.240	0.038
	(0.077)	(0.378)	(0.372)	(0.110)	(0.324)	(0.350)
Girls * Avg Share Male	–0.077	–0.419	0.488	–0.004	–0.900	0.981
	(0.073)	(0.401)	(0.394)	(0.104)	(0.350)	(0.370)
MOP F^{eff}	7.66	7.66	7.66	7.66	7.66	7.66
Obs	825,095	825,095	825,095	825,095	825,095	825,095
Boys: Dep mean	.025	.628	.346	.044	.498	.432
Girls: Dep mean	.02	.442	.537	.041	.327	.611
Boys: Std effect	-.031	.102	-.094	-.063	.031	.005
Girls: Std effect	-.034	-.054	.062	-.001	-.122	.128
P-value	.979	0	0	.002	0	0

Note: IV estimate of Equation 10.
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 do not 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).
All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

E IV Estimates by Pupil Gender: Main Outcomes

E.1 IV Estimates by Pupil Gender

Table A9: By Gender: Applications and Enrollment for Post-Compulsory Education

	Apply directly	Apply late	Apply never	Top choice	Enrolled at 16	Enrolled ever
Boys * Avg	0.474	–0.399	–0.076	0.598	0.720	0.110
Share Male	(0.204)	(0.182)	(0.077)	(0.256)	(0.323)	(0.079)
Girls * Avg	0.361	–0.284	–0.077	0.447	0.474	0.141
Share Male	(0.192)	(0.173)	(0.073)	(0.248)	(0.301)	(0.074)
MOP F^{eff}	7.66	7.66	7.66	7.66	15.64	15.64
Obs	825,095	825,095	825,095	825,095	695,341	695,341
Boys: Dep mean	.889	.085	.025	.852	.845	.977
Girls: Dep mean	.933	.047	.02	.863	.876	.982
Boys: Std effect	.096	-.091	-.031	.107	.124	.046
Girls: Std effect	.092	-.086	-.034	.083	.09	.066
P-value	.097	.06	.979	.215	.057	.471

Note: IV estimate of Equation 10.

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 choices in the application (Top 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).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A10: 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 Share	0.242	–0.194	0.458	–0.262	–0.432	0.317	–0.128
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)
MOP F^{eff}	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Obs	810,066	810,066	810,066	810,066	810,066	810,066	810,066
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	0	.11	.428	0	.89	.019	.011

Note: IV estimate of 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A11: By Gender: Labor Market Outcomes at Age 25

	Employed/ Student	Un- employed	DI/ Pension	Other out of LF
Boys * Avg	0.685	–0.007	–0.144	–0.526
Share Male	(0.254)	(0.161)	(0.078)	(0.143)
Girls * Avg	0.320	–0.079	–0.101	–0.099
Share Male	(0.243)	(0.156)	(0.077)	(0.139)
MOP F^{eff}	7.70	7.70	7.70	7.70
Obs	811,393	811,393	811,393	811,393
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	.396	.159	0

Note: IV estimate of 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A12: By gender: Field of Education at Age 25

	Male	Neutral	Female	STEM	STEM-M	Education/ Teacher	Primary Teacher
Boys * Avg	0.506	–0.679	0.172	0.653	0.492	0.032	0.032
Share Male	(0.388)	(0.408)	(0.188)	(0.439)	(0.425)	(0.060)	(0.046)
Girls * Avg	0.089	–0.340	0.252	0.550	0.977	–0.049	0.101
Share Male	(0.212)	(0.360)	(0.321)	(0.264)	(0.399)	(0.128)	(0.088)
Boys: MOP F^{eff}	14.91	14.91	14.91	14.91	14.91	14.91	14.91
Girls: MOP F^{eff}	15.75	15.75	15.75	15.75	15.75	15.75	15.75
Boys: Obs	415,571	415,571	415,571	415,571	415,571	415,571	415,571
Girls: Obs	395,822	395,822	395,822	395,822	395,822	395,822	395,822
Boys: Dep mean	.526	.39	.084	.412	.446	.009	.005
Girls: Dep mean	.069	.478	.453	.108	.309	.038	.017
Boys: Std effect	.065	-.089	.04	.085	.063	.022	.031
Girls: Std effect	.022	-.043	.032	.113	.134	-.016	.05
P-value	.337	.508	.825	.83	.338	.549	.491

Note: IV estimates of Equation 10. Outcomes from left to right:

Field is 'Male' dominated (> 70% male), (gender) 'Neutral' or 'Female' dominated (> 70% female), based on previous generation.

Field is STEM or STEM + Medicine (STEM-M).

Field is Education Science or Teacher. Field is Primary School Teacher.

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (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	0.346	–0.352	0.006	0.276	–0.053	–0.078	–0.121
Share Male	(0.261)	(0.244)	(0.104)	(0.317)	(0.239)	(0.099)	(0.132)
Girls * Avg	0.498	–0.334	–0.164	0.725	–0.003	–0.171	–0.530
Share Male	(0.217)	(0.178)	(0.089)	(0.290)	(0.172)	(0.097)	(0.219)
Boys: MOP F^{eff}	14.83	14.83	14.83	14.91	14.91	14.91	14.91
Girls: MOP F^{eff}	15.66	15.66	15.66	15.75	15.75	15.75	15.75
Boys: Obs	421,151	421,151	421,151	415,571	415,571	415,571	415,571
Girls: Obs	403,944	403,944	403,944	395,822	395,822	395,822	395,822
Boys: Dep mean	.889	.085	.025	.84	.102	.019	.037
Girls: Dep mean	.933	.047	.02	.845	.07	.015	.07
P-value	.58	.941	.18	.233	.86	.461	.084

Table A14: Highest Degree Achieved at Age 25

	Compulsory		Vocational			Academic	
	schooling	Sec	Sec Plus	Tert	Sec	Tert: BA	Tert: MA
Boys * Avg	0.181	0.061	0.209	–0.101	–0.817	0.541	–0.074
Share Male	(0.234)	(0.386)	(0.223)	(0.231)	(0.345)	(0.195)	(0.092)
Girls * Avg	–0.508	–0.200	0.640	–0.070	–0.029	0.234	–0.068
Share Male	(0.233)	(0.331)	(0.301)	(0.304)	(0.239)	(0.173)	(0.155)
Boys: MOP F^{eff}	14.90	14.90	14.90	14.90	14.90	14.90	14.90
Girls: MOP F^{eff}	15.74	15.74	15.74	15.74	15.74	15.74	15.74
Boys: Obs	414,896	414,896	414,896	414,896	414,896	414,896	414,896
Girls: Obs	395,170	395,170	395,170	395,170	395,170	395,170	395,170
Boys: Dep mean	.152	.378	.081	.094	.231	.042	.022
Girls: Dep mean	.101	.251	.136	.201	.19	.067	.054
P-value	.052	.599	.195	.924	.038	.173	.973

Note: IV estimate of Equation 10, split sample estimates by gender (i.e. fully interacted model). Outcomes are pupils' applications in the last year of middle school (see Table 3) and labor market status (see Table 5). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

F Reduced Form Estimates

F.1 Reduced Form: Main Results

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

	Apply directly	Apply late	Apply never	Top choice	Enrolled at 16	Enrolled ever
Total Share 60 *	–0.071	0.058	0.013	–0.089	–0.089	–0.018
Post-Quota	(0.027)	(0.024)	(0.012)	(0.036)	(0.035)	(0.010)
Total Share 60	0.017	–0.021	0.003	–0.016	0.026	0.005
	(0.021)	(0.019)	(0.009)	(0.027)	(0.028)	(0.008)
Adj. R^2	0.070	0.030	0.138	0.063	0.049	0.024
Obs	825,095	825,095	825,095	825,095	695,341	695,341
Dep mean	.911	.066	.023	.858	.861	.98

Note: Reduced Form estimates as in Equation 11.

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 choices in the application (Top 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).

All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A16: Reduced Form: Highest Degree Achieved at Age 25

	Compulsory schooling	Vocational			Academic		
		Sec	Sec Plus	Tert	Sec	Tert: BA	Tert: MA
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)
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)
Adj. R^2	0.056	0.086	0.012	0.035	0.065	0.042	0.028
Obs	810,066	810,066	810,066	810,066	810,066	810,066	810,066
Dep mean	.127	.316	.108	.146	.211	.054	.038

Note: Reduced Form estimates as in Equation 11. 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A17: Reduced Form: Labor Market Outcomes at Age 25

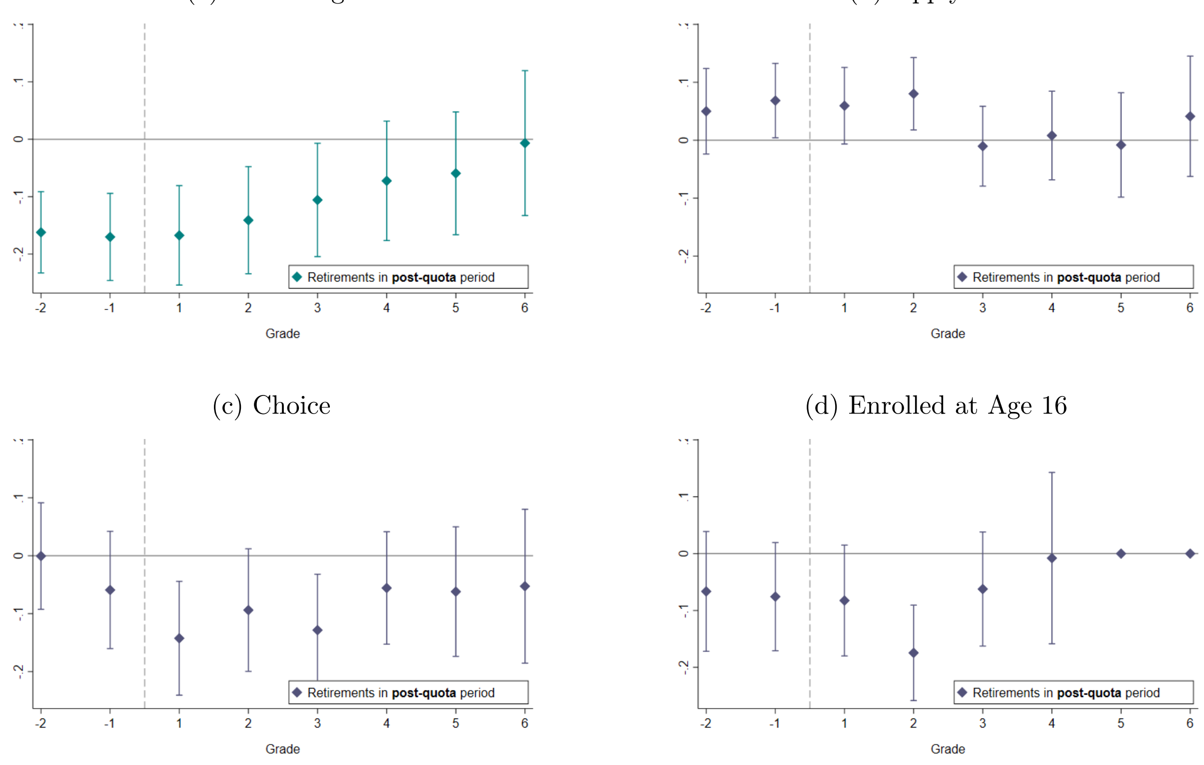
	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Total Share 60 *	–0.086	0.006	0.021	0.055
Post-Quota	(0.033)	(0.026)	(0.011)	(0.018)
Total Share 60	0.042	–0.001	–0.005	–0.034
	(0.029)	(0.022)	(0.009)	(0.015)
Adj. R^2	0.025	0.023	0.003	0.013
Obs	811,393	811,393	811,393	811,393
Dep mean	.842	.086	.017	.053

Note: Reduced Form estimates as in Equation 11. 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

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

Figure A6: Intermediate Outcomes: Grade Level Estimation

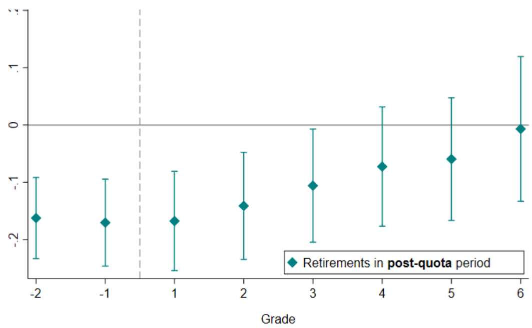
(a) First Stage (b) Apply Late (c) Choice (d) Enrolled at Age 16



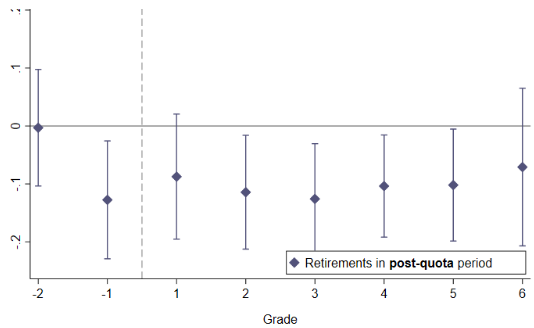
Note: Grade level estimation of pupil level first stage and reduced form (see Equation 11). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

Figure A7: Long-Term Outcomes: Grade Level Estimation

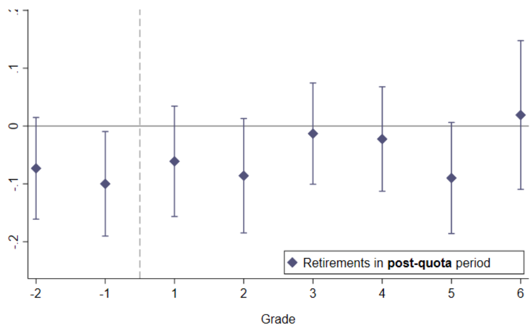
(a) First Stage



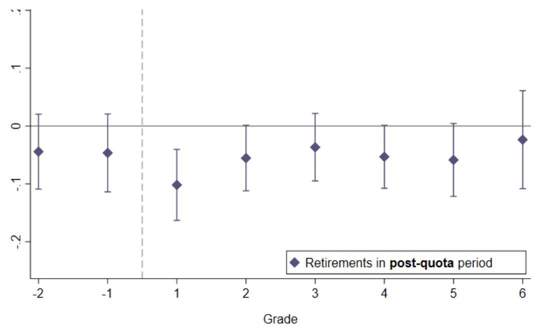
(b) Employed/Student



(c) Vocational Plus Degree



(d) BA Degree



Note: Grade level estimation of pupil level reduced form (see Equation 11). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 * Post	–0.086	0.060	0.025	–0.085	0.017	0.023	0.039
	(0.038)	(0.034)	(0.016)	(0.049)	(0.038)	(0.017)	(0.028)
High: Total Share 60 * Post	–0.066	0.058	0.009	–0.080	–0.000	0.025	0.055
	(0.036)	(0.033)	(0.017)	(0.045)	(0.036)	(0.016)	(0.024)
Low: Total Share 60	0.008	–0.010	0.002	0.053	–0.007	–0.008	–0.033
	(0.033)	(0.028)	(0.014)	(0.043)	(0.033)	(0.014)	(0.022)
High: Total Share 60	0.039	–0.033	–0.006	0.028	0.009	–0.007	–0.031
	(0.027)	(0.024)	(0.012)	(0.040)	(0.031)	(0.012)	(0.021)
Obs	819,498	819,498	819,498	805,895	805,895	805,895	805,895
Low: Dep mean	.904	.07	.026	.846	.082	.017	.053
High: Dep mean	.928	.057	.015	.832	.097	.018	.052
P-value	.717	.96	.478	.944	.744	.904	.663

Note: Reduced form estimates for Equation 11 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 3 and 5). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

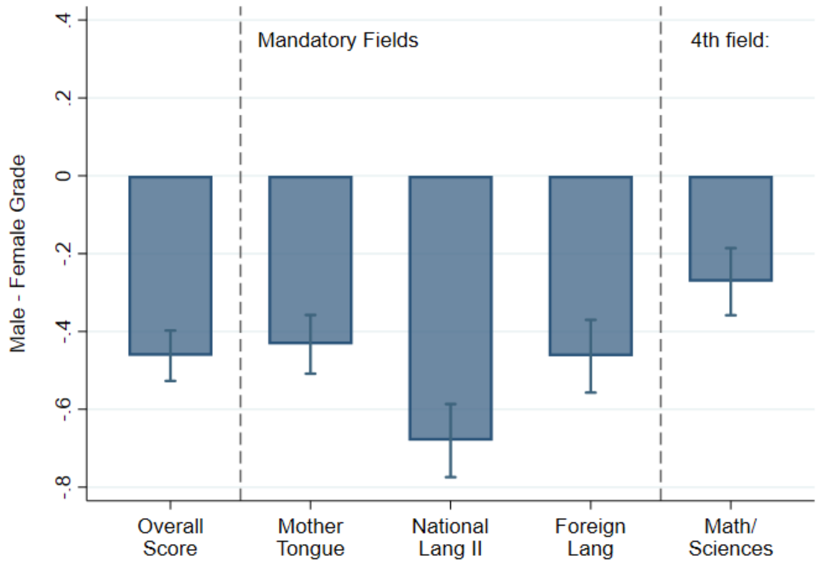
Table A19: Complementarities: IV estimates

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Low: Avg Share	0.766	–0.538	–0.228	0.752	–0.149	–0.199	–0.344
Male	(0.584)	(0.460)	(0.193)	(0.598)	(0.342)	(0.197)	(0.294)
High: Avg Share	0.545	–0.474	–0.071	0.663	0.002	–0.208	–0.453
Male	(0.385)	(0.351)	(0.139)	(0.488)	(0.301)	(0.160)	(0.280)
Low: MOP F^{eff}	3.69	3.69	3.69	3.78	3.78	3.78	3.78
High: MOP F^{eff}	5.04	5.04	5.04	5.00	5.00	5.00	5.00
Low: Obs	590,156	590,156	590,156	579,101	579,101	579,101	579,101
High: Obs	229,342	229,342	229,342	226,794	226,794	226,794	226,794
Low: Dep mean	.904	.07	.026	.846	.082	.017	.053
High: Dep mean	.928	.057	.015	.832	.097	.018	.052
P-value	.752	.912	.51	.909	.739	.971	.79

Note: IV estimate of Equation 10, split sample by median share male teachers in a municipality in 1990. Outcomes are pupils' application choices and labor market status (see Tables 3 and 5). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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.2 Performance Differences in Matriculation Exam

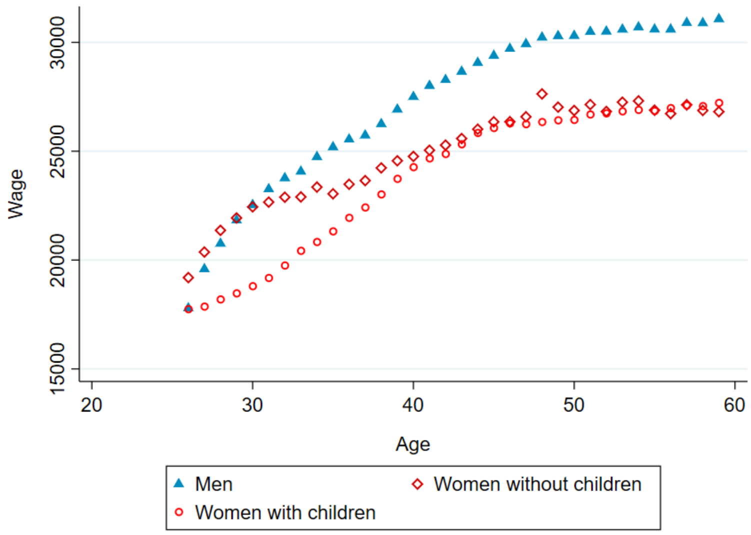
Figure A8: Performance Differences by Gender on Matriculation Exam by Field, Teachers



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 (see Figure 8). 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)

G.3 Career Attachment and Teachers' Compensation

Figure A9: Teacher Earnings by Age



Note: Wage earnings by age for the panel of active primary teachers. Earnings are winsorized below the 1st and above the 99th percentile. (back)

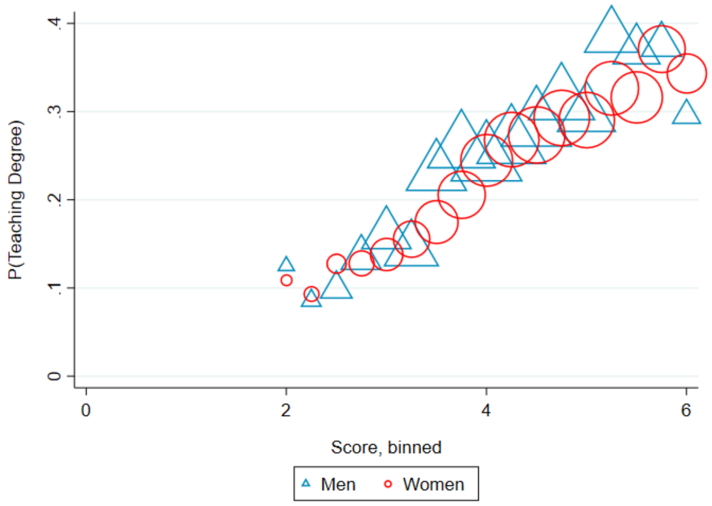
Table A20: Teacher Earnings

			No transfer	No Kids	Principal	
Male	3,141.128	3,074.115	2,579.230	1,056.822	0.036	2,878.487
	(26.712)	(45.933)	(25.957)	(84.609)	(0.002)	(66.893)
Experience		285.074				
		(1.453)				
Male * Experience		4.057				
		(2.508)				
Principal						7,966.748
						(226.359)
Adj. R^2	0.165	0.348	0.177	0.117	0.043	0.160
Obs	206,017	206,017	185,668	20,736	39,259	38,474
Dep mean	24,746	24,746	25,414	23,429	.021	25,927
Dep mean men	26,899	26,899	27,179	24,129	.045	28,087
Dep mean women	23,713	23,713	24,547	23,143	.01	24,938

Note: Regressions based on the panel of active primary teachers (ages 26-60). Outcome in columns 1–4 and 6 are yearly earnings, outcome in column 5 is an indicator for being principal. No transfer: Sample based on active teachers who did not receive any government transfer in a given year. No kids: Sample based on teachers who never had a child. Experience is measured as time since degree. Earnings are winsorized below the 1st and above the 99th percentile. All specifications control for year and municipality fixed effects. (back)

G.4 Compensating Differentials

Figure A10: Likelihood of Obtaining Primary Teaching Degree by Exam Score and Gender



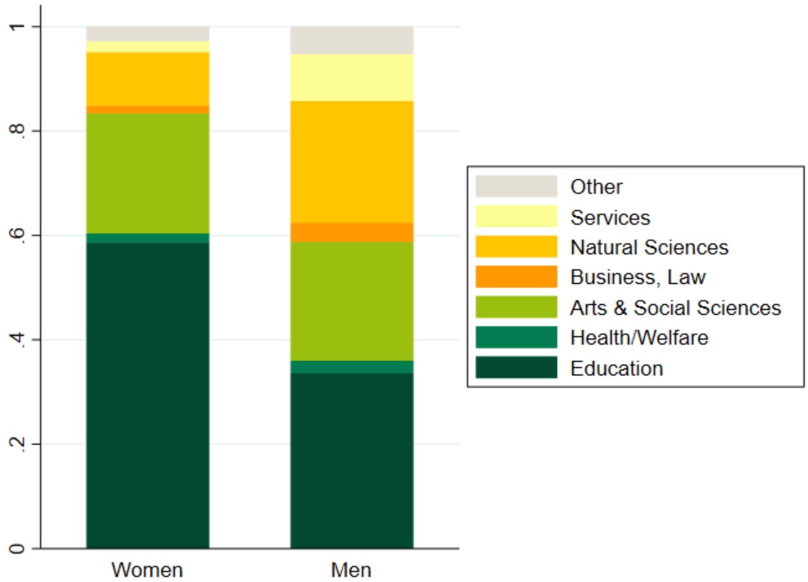
Note: Share of primary school teaching degree holders among all applicants to primary teaching by binned score in the matriculation exam. Weighted by number of applicants per bin. Sample contains all applicants to primary teacher education in the years 2000 - 2005, with data for 2003 missing and considers all degrees obtained by the year 2020. The exam score is based on a candidate's best performance if they had multiple takes. 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 A21: Likelihood of Teaching Degree by Score and Gender, Among Applicants to Primary Teaching

Score	0.072	0.073	0.069	0.070
	(0.003)	(0.004)	(0.004)	(0.004)
Male		–0.001		–0.006
		(0.036)		(0.043)
Male * Score		0.007		0.012
		(0.008)		(0.010)
Adj. R^2	0.100	0.100	0.082	0.084
Obs	20,560	20,560	17,934	17,934
Dep mean	.266	.266	.305	.305

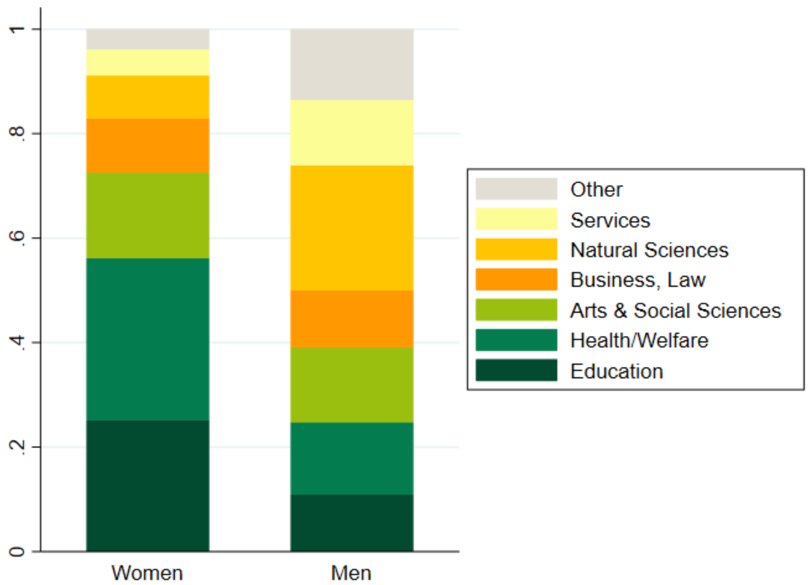
Note: Outcome is an indicator variable for whether an applicant holds a primary teaching degree in the year 2020. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers all degrees obtained by the year 2020. In columns 3 and 4, sample 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 9 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). (back)

Figure A11: 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. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing). (back)

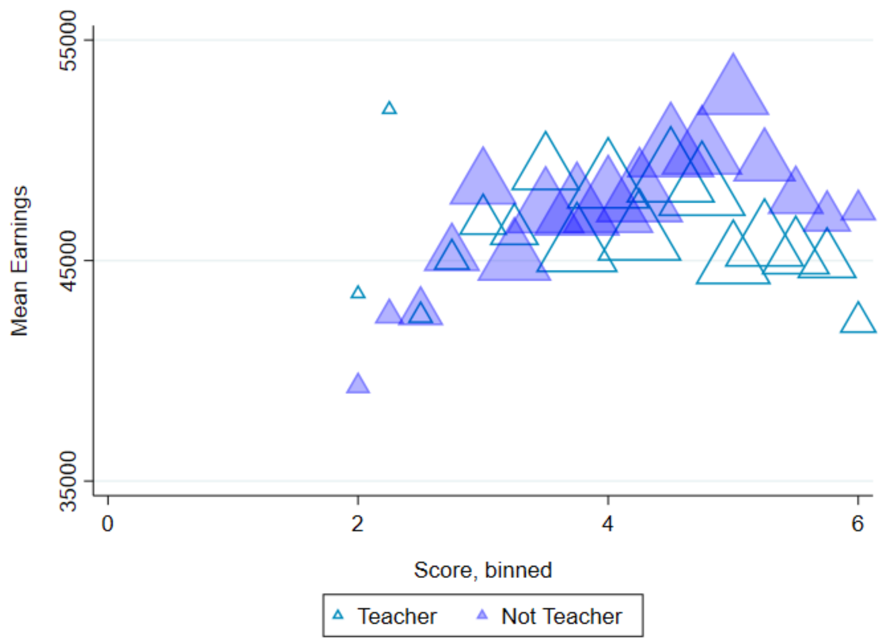
Figure A12: Actual Fields Obtained among Applicants Outside Primary Teaching



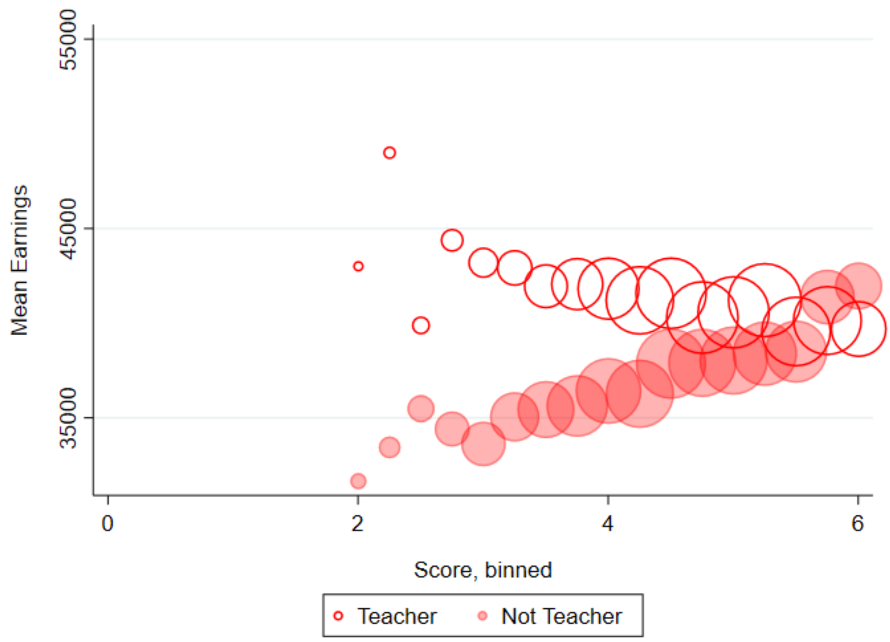
Note: Share of actual fields obtained among those who do not obtain a primary teaching degree by 2020. Sample contains all applicants to primary teacher education in the years 2000 - 2005 (data for 2003 missing) and considers most recent degree obtained in 2020. (back)

Figure A13: 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)

G.5 Pupil Grades

Table A22: IV Estimates: Pupil Grades in Middle School and Matriculation Exam

	Middle School					Matriculation Exam				
	Overall GPA	Overall GPA $<$ 99	Language	Math	Sample	Overall GPA	Language	Math	Math/ Science	Sample
Avg Share Male	–0.614	–0.651	–0.729	–0.008	0.039	–1.038	-0.715	0.558	-1.669	0.096
	(0.641)	(0.640)	(0.860)	(0.948)	(0.117)	(0.862)	(0.933)	(2.022)	(1.362)	(0.279)
MOP F^{eff}	15.71	19.79	19.83	19.82	15.28	16.02	16.02	16.92	16.03	15.28
Obs	800,811	674,913	670,046	669,352	825,095	403,052	403,052	323,869	403,084	826,181
Dep mean	7.616	7.608	7.548	7.485	.971	4.289	4.247	4.021	4.403	.488
Std effect	-.035	-.037	-0.038	0	.015	–0.066	-0.043	0.022	–.08	0.012

Note: Estimates of Equation 10. Outcomes from left to right:
Middle School: Overall GPA across all subjects; Overall GPA for the cohorts that start primary school before 1999; Language: the average across mother tongue and second national language grade; and Math. Language and Math are only available for the cohorts that start primary school before 1999. All grade information is based on pupils who ever apply to post-compulsory education and have a grade entered. Sample: Indicator for whether the pupil ever applies in the post-compulsory education data and has Overall GPA across all subjects entered.
Matriculation Exam: Overall GPA across mandatory subjects; Language: average across mother tongue, second national and foreign language; Math; Math/Science: the better score across Math and Sciences. All grades are for pupils who pass the matriculation exam by age 20. Sample: Indicator for whether pupil is in Overall GPA sample for matriculation exam, excluding those who have died by age 25.
All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level.

Table A23: Pupil Grades and Teacher Exam Scores

	Middle School Overall GPA				Matriculation Exam Overall GPA			
Overall GPA	–0.0080				0.0231			
	(0.0275)				(0.0401)			
Language		0.0044	0.0091			0.0785	0.0770	
		(0.0269)	(0.0284)			(0.0383)	(0.0396)	
Math or Science		–0.0146				–0.0646		
		(0.0215)				(0.0309)		
Math			–0.0039				–0.0206	
			(0.0120)				(0.0160)	
Science			–0.0167				–0.0302	
			(0.0198)				(0.0279)	
Math Background				–0.0242				0.0415
				(0.0606)				(0.0788)
Adj. R^2	0.188	0.188	0.188	0.188	0.082	0.082	0.082	0.082
Obs	800,750	800,750	800,750	800,750	403,021	403,021	403,021	403,021
Dep mean	7.616	7.616	7.616	7.616	4.289	4.289	4.289	4.289
Std effect X1	-.001	.001	.002	-.002	.004	.015	.015	.003
Std effect X2		-.003	-.001			-.014	-.007	
Std effect X3			-.004				-.007	

Note: Estimates of pupil outcomes on teachers' scores in the matriculation exam. Outcomes are pupils' Overall GPA across all subjects in middle school and Overall GPA in the matriculation exam (see Appendix Table A22 for a description of these outcomes). The score for Language is comprised of the (teachers') 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. Std effect X1-X3 indicate the standardized effect for each independent variable from (1) to bottom (3) for the respective column (i.e. in column 2, Std effect X1 indicates the standardized effect for Language score). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level.

H Robustness

H.1 Maternity/Paternity of teachers

Table A24: Teachers Becoming Parents

	Birth total	Birth fem	Birth male	Maternity leave
Share 60 * Post-Quota	0.064	0.033	0.031	–0.038
	(0.050)	(0.036)	(0.029)	(0.032)
Share 60	–0.122	–0.108	–0.014	–0.003
	(0.040)	(0.029)	(0.025)	(0.030)
Adj. R^2	.006	.006	.012	.005
Obs	4448	4448	4448	4448
Dep mean	.06	.03	.02	.02

Note: Estimates for Equation 9. 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). All specifications include year fixed effects. Standard errors clustered at the municipality level. Regressions weighted by population, means unweighted. (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	0.007	0.002	–0.003	–0.007	0.006	0.003
	(0.008)	(0.008)	(0.003)	(0.010)	(0.008)	(0.003)	(0.005)
Adj. R^2	0.070	0.030	0.138	0.025	0.023	0.003	0.013
Obs	825,095	825,095	825,095	811,393	811,393	811,393	811,393
Dep mean	.911	.066	.023	.842	.086	.017	.053

Note: Estimates of pupil outcomes on total exposure to female teachers having a newborn child while pupils are in primary school. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level.
(back)

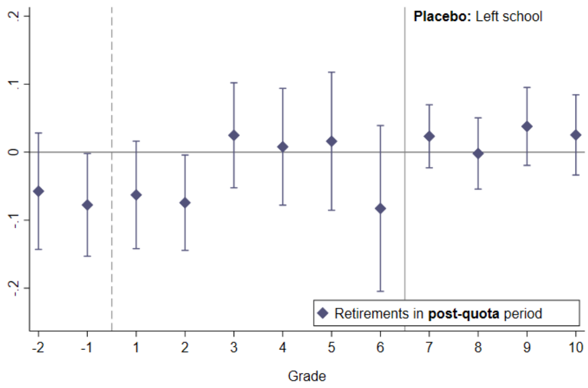
Table A26: Effect on Main Outcomes of Male Teachers Having a Newborn Child

	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)
Adj. R^2	0.070	0.030	0.138	0.025	0.023	0.003	0.013
Obs	825,095	825,095	825,095	811,393	811,393	811,393	811,393
Dep mean	.911	.066	.023	.842	.086	.017	.053

Note: Estimates of pupil outcomes on total exposure to male teachers having a newborn child while pupils are in primary school. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. Standard errors clustered at the municipality level. (back)

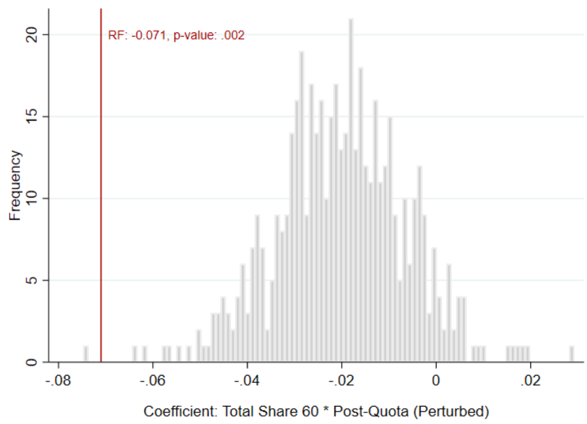
H.2 Placebo Check and Randomization Inference

Figure A14: Placebo Reduced Form by Grade: Applications to Post-Compulsory Education



Note: Grade level estimation of pupil level reduced form (see Equation 11). 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. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A15: 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 11. Outcome is binary indicator for pupils applying to post-compulsory education directly after middle school. Line indicates the reduced form coefficient in Table A15. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 * Post-Quota	–0.004	0.003		
	(0.005)	(0.009)		
Total Share 60	0.006	0.006	0.005	0.007
	(0.005)	(0.007)	(0.004)	(0.006)
MOP F^{eff}			15.28	15.28
Adj. R^2	0.042	0.049		
Obs	826,515	826,515	826,181	826,181
Dep mean	.005	.018	.005	.018

Note: Reduced form, and IV estimate of Equation 10. Outcomes are a binary indicator for pupils having left the sample at age 16 or age 25, excluding children with registered deaths until age 25. Pupils are defined as having left the sample if they do not appear in the register data at the respective age. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A28: Further Robustness

	Apply					Employed/Student				
	Main	No capital	No cities	Parent UB	Movers	Main	No capital	No cities	Parent UB	Movers
Avg Share Male	0.421	0.433	0.455	0.391	0.432	0.512	0.491	0.503	0.459	0.488
	(0.196)	(0.197)	(0.214)	(0.192)	(0.207)	(0.243)	(0.237)	(0.255)	(0.234)	(0.240)
MOP F^{eff}	15.28	15.59	13.86	15.28	15.45	15.37	15.69	13.96	15.37	15.57
Obs	825,095	758,380	648,931	825,091	821,930	811,393	746,393	639,044	811,389	808,106
Dep mean	.911	.911	.911	.911	.911	.842	.842	.842	.842	.843

Note: IV estimate of Equation 10. Outcomes are pupils' applications in the last year of middle school (see Table 3) and labor market status (see Table 5), in turn examining the main specification for comparison (column 1 and 6), dropping Helsinki (column 2 and 7), dropping the five most populous municipalities based on place of living at age 7 (column 3 and 8), controlling for parental unemployment status at age 7 (column 4 and 9), and assigning treatment based on the municipality in which pupils live in each year during primary school instead of only the municipality of living at age 7 (columns 5 and 10). The lower observation count in columns 5 and 10 is due to excluding pupils for whom I cannot observe a municipality in all years. All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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 A29: Reduced Form by Size of Municipality

	Apply directly	Apply late	Apply never	Employed/ Student	Un-employed	DI/ Pension	Other out of LF
Small: Total	–0.073	0.058	0.016	–0.083	–0.009	0.021	0.064
Share 60 * Post	(0.033)	(0.029)	(0.014)	(0.044)	(0.033)	(0.013)	(0.023)
Large: Total	–0.070	0.058	0.012	–0.087	0.010	0.021	0.053
Share 60 * Post	(0.028)	(0.025)	(0.012)	(0.033)	(0.026)	(0.011)	(0.019)
Total Share 60	0.018	–0.021	0.003	0.042	–0.000	–0.005	–0.035
	(0.021)	(0.019)	(0.009)	(0.029)	(0.023)	(0.009)	(0.015)
Adj. R^2	0.070	0.030	0.138	0.025	0.023	0.003	0.013
Obs	825,095	825,095	825,095	811,393	811,393	811,393	811,393
Small: Dep mean	.924	.055	.021	.849	.084	.016	.049
Large: Dep mean	.91	.067	.023	.842	.086	.017	.053
Share Small	.042	.042	.042	.042	.042	.042	.042
P-value	.891	.998	.726	.894	.429	.945	.504

Note: Reduced form estimates (Equation 11), 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 3) and labor market status (see Table 5). All specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: 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.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 A16a and A16b plot the residualized share male against the residualized treatment variable both for the municipal and the pupil level first stage.[25] A test for differences in slopes between treatment and control observations shows that these are small and not significant.

[25] I.e. for the first stage equation 8 at the municipal level, residualized treatment corresponds to the residuals of the following specification:

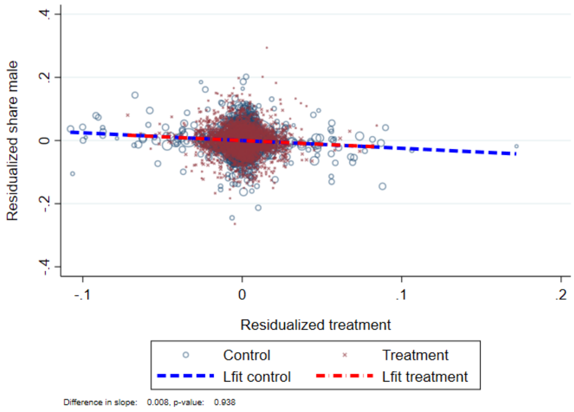
$$\textit{total share 60 post}_{mt} = \beta_0 + \beta_1 \textit{total share 60}_{mt} + \gamma_{mp} + \eta_t + \epsilon_{mt} \qquad (16)$$

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

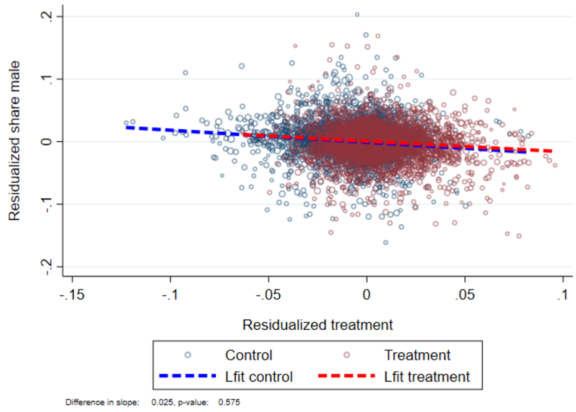
Second, following De Chaisemartin and d'Haultfoeuille (2020), we can examine the weights that observations receive. The focus here is to understand how treatment assignment based on actual treatment status maps into treatment assignment based on the *residualized* treatment variable. In our setting, 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 A17 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 A16: Treatment Effect Heterogeneity

(a) Municipality Level: Residuals of First Stage



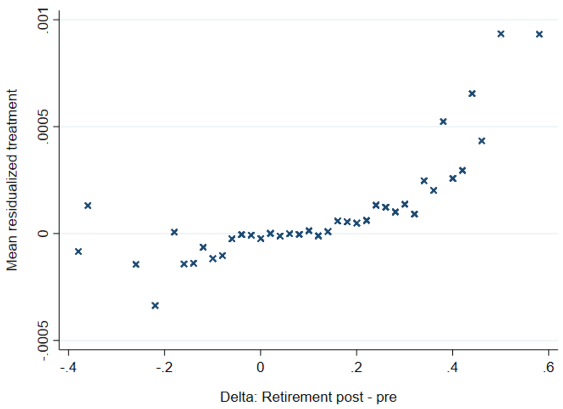
(b) Pupil Level: Residuals of First Stage



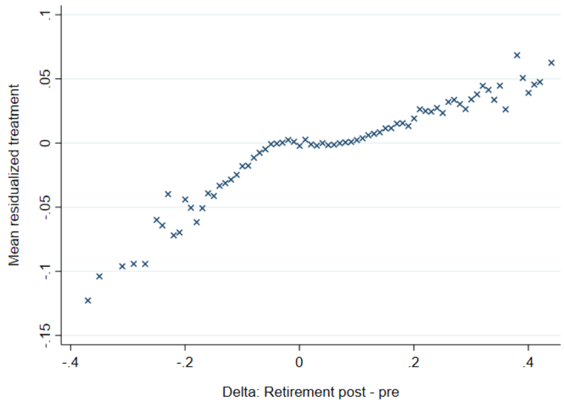
Note: Residualized share male against residualized treatment (see Equation 16), at the municipal level (upper panel) and pupil level (lower panel). Weighted by number of observations. (back)

Figure A17: Residualized Treatment

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



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

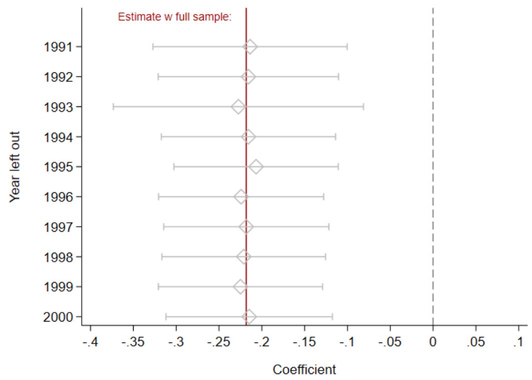


Note: Mean residualized treatment (see Equation 16) against actual treatment assignment (binned) at the municipal level (upper panel) and pupil level (lower panel). (back)

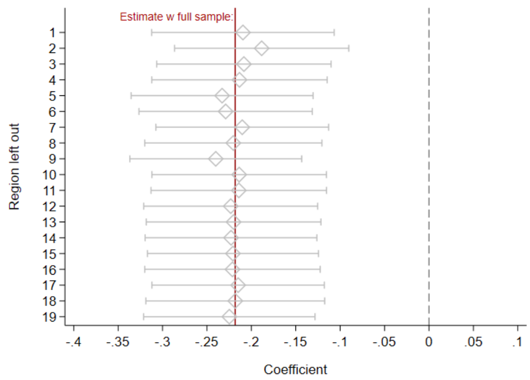
H.6 Leave-One-Out Estimation

Figure A18: First Stage: Leave-One-Out

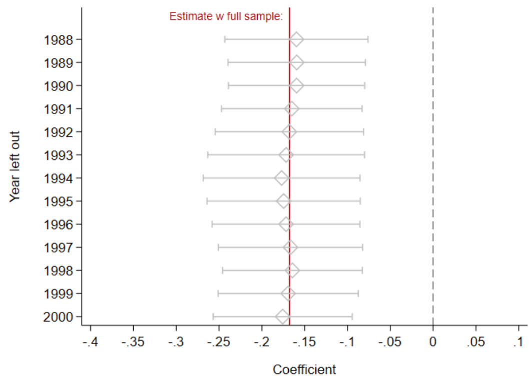
(a) Municipal Level: Leave-One-Out Years



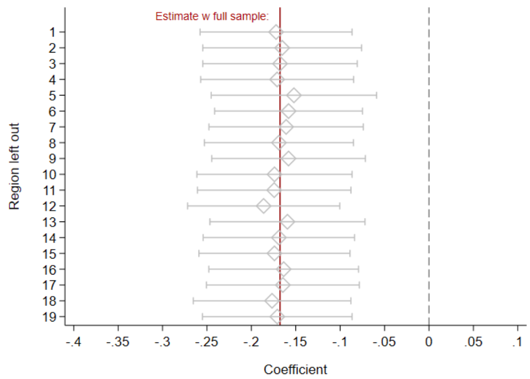
(b) Municipal Level: Leave-One-Out Region



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



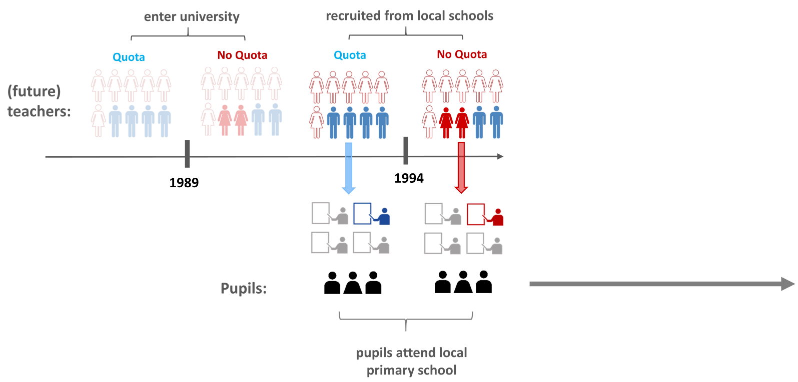
(d) Pupil Level: Leave-One-Out Region



Note: Leave-one-out estimation of treatment coefficient in municipal level first stage Equation 8 (Panel a and b) and pupil level first stage Equation 11 (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.
At the municipal level, all specifications include year and municipality-by-post-quota fixed effects. At the pupil level, all specifications include region-by-cohort and municipality fixed effects, as well as individual-level controls measured at age 7: gender, language (SE/FI/other), foreign origin, single parent HH, highest level of education in HH. (back)

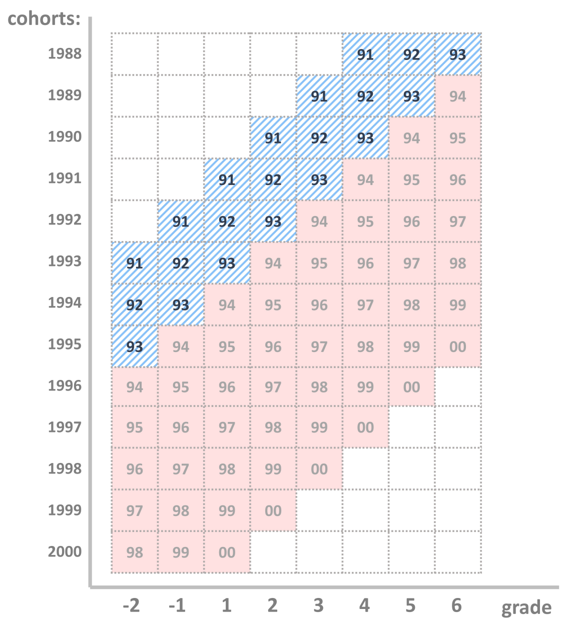
I Context

Figure A19: 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 A20: 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)