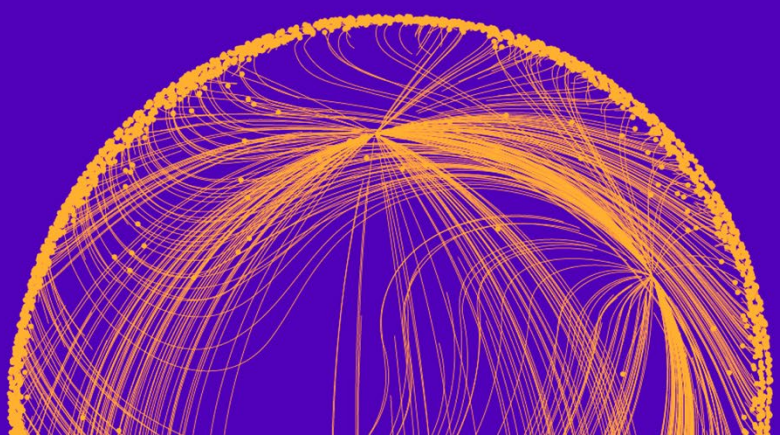
Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 30 · 2024

Teacher Selection in Finland

Ramin Izadi



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 30 · 2024

Ramin Izadi:
Teacher Selection in Finland

ISBN 978-952-7543-29-0 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, November 2024

Teacher Selection in Finland*

Ramin Izadi†

Aalto University

August 16, 2021

Abstract

Finland's success in international student comparisons is often attributed to the quality of its teachers. In this paper I examine the teacher selection process in Finland and highlight three key new findings. First, using rich administrative data for graduating cohorts between 1973-2012, I show that teacher graduates have consistently lower standardized test performance in comparison to other university graduates. However, in contrast to findings from other developed countries, they have been closing that gap during the last 40 years. Second, past test performance is a poor predictor of teacher aptitude, as measured by expert evaluators during entrance interviews for teacher training programs. This implies that the performance gap between teaching and other programs is not due to lack in applicant quality, but due to uncorrelated factors that influence the aptitude test performance in teaching. In other words, teacher training programs in Finland are not looking to enroll the academically best students. Third, relative to other university graduates with similar academic track record, teachers have high wages but low earnings, which helps to explain the popularity of teacher training programs in certain demographics.

Keywords: Teacher skills, teacher compensation, occupational choice, education.

JEL Codes: I21, J24, J31

*I thank Manuel Bagues, Matti Sarvimäki, Tuomas Pekkarinen, Marko Terviö, Mikko Silliman, and workshop participants in Helsinki GSE for thoughtful discussions and comments, and OP Group Research Foundation and Yrjö Jansson Foundation for generous financial support.

†Izadi: Aalto University, Helsinki GSE, ramin.izadi@aalto.fi.

1 Introduction

Teachers are perhaps the most important variable input that determines student outcomes in an education system. There is a growing concern in many developed countries that teacher quality is declining. Decreasing trends in teacher cognitive ability have been found for USA (Bacolod, 2007), UK (Nickell and Quintini, 2002), Australia (Leigh and Ryan, 2008) and for Sweden (Grönqvist and Vlachos, 2016). Eyes of policy makers have turned towards countries that consistently rank high in international student comparisons. Finland has excelled in the PISA country ranking for the first decade since it's inception in 2000. Public attention has focused on the quality of Finland's teachers and teacher training programmes. Part of the prevailing notion is that the Finnish education system is successful, because teacher training programs in Finland are highly selective and attract only the best high school applicants (Auguste and Miller, 2010).

This paper is a descriptive case study of the selection process of Finnish primary school teachers. They teach students from grades 1 to 6, usually staying with one cohort for several years, and typically teaching most subjects themselves. Teacher training programs in Finland award master's degrees and admit applicants through a process that includes an aptitude test with in-person interviews. First, I will look at the test score distribution of graduating teachers and compare them to other university graduates. What part of the test score distribution are teachers drawn from in Finland and how does it change over time? I use rich register data from years 1967-2012 to construct an index of academic ability based on standardized test performance in the academic track of high school [1]. I use it to track changes in the relative positions of teachers vs. other university graduates. Second, I relate the academic track test performance of teachers applicants to success in the aptitude test of teacher training programs. Is academic ability associated with success

[1] Secondary education has an academic and a vocational track in Finland. Only academic track administers standardized tests.

in the entrance interviews? I use data from the central university application register in years 2000-2014 with standard regression techniques to analyze this correlation. Finally, I evaluate teacher compensation, occupational exit, employment rate and promotion opportunites. Is teaching an attractive career choice? I use tax registers from years 2011-2015 to estimate a Mincer-type equation to find earnings premiums for teachers.

Through this analysis, I come to three main findings. First, in contrast to findings from other countries, I find that in a 40 year period, the average academic ability of teachers has increased by 20 percentiles in the distribution of high school academic track performance. In 1973, the average teacher graduate came from the 40th percentile of the academic track graduates, whereas in 2012 she came from the 60th percentile. Most of the gains were made before 1990. However, despite the gap closing over the period, the average teacher is consistently ranked below the average university graduate. This gap is present at least in mathematics, general studies, and language scores. The main caveat in this analysis is that due to increasing take up of the academic track, it is difficult to describe time trends relative to the population distribution.

Second, I find that academic ability is a poor predictor of success in the teacher aptitude test. Teacher training programs use an in-person aptitude test in conjunction with other criteria to select applicants. A panel of three educational experts interview the applicants individually and in groups to assess their suitability to become primary school teachers. Before 2007, applicants were pre-selected into these interviews based primarily on their high school test performance. After receiving an invitation, pre-selection points were reset and admission was solely based on the interviews. I observe the invitation decision as well as the final admission decision.

I find that relative to entrance exams in other university programs, high school performance has almost no predictive power for success in the aptitude test for teacher training programs. In particular, mathematical ability is irrelevant, whereas Finnish language is the only significant predictor. Even so, the R^2 is close to zero. In a counterfactual exercise, I show that by choosing higher performing applicants, teacher education programmes

could improve the academic ability of their admitted students at least to the level of the average master's student. This implies that the performance gap between teaching and other programs is not due to lack in applicant quality, but due to uncorrelated factors that influence the aptitude test performance in teaching. In other words, teacher training programs in Finland are not looking to enroll the academically best students.

The relatively weak correlation of academic ability with admission in teacher education programmes is not news among educational scientists in Finland. Quoting a notable Finnish educationalist:

> "*Academically best students are not necessarily the best teachers. Successful education systems are more concerned about finding the right people to become career-long teachers.*"
> — Pasi Sahlberg (The Guardian, 2015)

Finally, I show that the career offered by teacher training programs seems very attractive for risk averse individuals with a preference for low labor supply in the intensive margin. To start out, I use a Mincerian regression to show that teachers in Finland have 9% lower annual earnings than other university graduates with comparable test scores and potential work experience. This result is in contrast with Hanushek et al. (2019) who find a positive wage premium of 10% for Finnish teachers. The contradictory results are reconciled by differences between teachers and other professionals in total yearly hours supplied. Tenured primary school teachers in Finland have at least 12 weeks of paid vacations yearly compared to the average 5 weeks. They also report a shorter work week of 32 hours on average compared to the more typical 38-40 hours. Other things equal, lower total yearly work hours push down earnings and raise the hourly wage. Lastly, I demonstrate that teachers seem to stick to their career choice: only 6% of teacher graduates finish their careers in non-education occupations and 10% end up principals. Likewise, the risk for non-employment is negligible for all but the recently graduated. Low occupational exit is consistent with qualitative accounts of potentially valuable non-pecuniary benefits

associated with Finnish teachers, such as public respect and high teacher autonomy. These facts taken together, it doesn't seem surprising that Finnish teacher training programs can afford to select students from a large pool of qualified applicants.

The goal of teacher selection policies should be to enroll teachers that improve the human capital and later life outcomes of students (Jackson et al., 2014). Empirical research typically resorts to measuring teachers' ability to improve the test scores of their students, which is shown to have a positive association with later life outcomes (Chetty et al., 2014). Effective teacher selection then boils down to identifying ex-ante individuals who are likely to be effective teachers. This has proven to be a formidable task. The economic literature has been unable to find a strong link between observable teacher characteristics and teacher value added, although teacher test scores have been most consistently related to student outcomes (Dobbie (2011), Jackson et al. (2014)). [2]

In quest of answering these questions, economists too have turned towards countries whose students perform well in international comparisons. Hanushek et al. (2019) evaluate the importance of teacher numeracy and literacy skill on student achievement in a cross country setting. Finnish teachers top the country ranking in both subjects. Additionally, Finnish teachers place in the 60th percentile in the distribution of Finnish college graduates, which is among the highest in any country. Their paper exploits within-country variation between numeracy and literacy skills of teachers to establish a link between teacher skills in a particular subject and the corresponding PISA scores of students across countries.

Due to data availability, I will not look at student outcomes. Rather, the main contribution of this paper is to demystify Finland as the paragon of teacher selection by providing a first quantitative look at the selection process of Finnish teachers. In light of the findings

[2]Due to the inability to link teachers and students, this paper does not focus on student outcomes, but instead looks purely at teacher selection.

by Hanushek et al. (2019), it is perhaps surprising to find that teacher training programs in Finland are not trying particularly hard to enroll the highest achieving students. Although subject knowledge is shown to be important for teachers (Bietenbeck et al. (2018), Metzler and Woessmann (2012)), test scores are only used to screen out the academically poorest performing applicants in the pre-selection phase. This suggests that the Finnish model endorses complementarities between academic and non-academic ability in teacher selection.

2 Institutional Background and Data

2.1 Teacher Selection Process

This study focuses on certified primary school teachers (grades 1-6), because they all go through the same university requirements and can be clearly identified from the data. In 2016, 95% of primary school teachers were certified (Kumpulainen, 2017), so this covers almost all primary school teachers in Finland. The process of becoming a certified teacher follows a clear cut path: passing the high school matriculation exam, applying to the teacher training programme in one of eight institutions, participating in a two-stage entrance examination, being admitted, and completing a master's degree with teacher certification.

Admission into teacher training differs from most other master's programs in that it requires an aptitude test in addition to a written exam. I look at teacher graduates from 1973 to 2012. During this period there have been many changes in the specific admission criteria and the test itself. In the early years, admission criteria were centrally directed by the ministry of education. Applicants were evaluated by the "model citizen" standard: teachers should be cultured, healthy and reputable citizens with upright characters and blameless track records (Räihä, 2010). Accordingly, admission was based on a variety of evaluations, and supported by references and interviews. Tests included a medical exam-

ination, written exams and a test for musicality and speech impediments. Importantly, they have always included an in-person evaluation by the faculty, first by the principal, and later by three person admission committees. Later reforms have moved away from central direction towards university autonomy. At the same time, evaluation of character and "model citizenship" has made room for emphasizing specific traits. Today, only two tests remain that are geared towards evaluating two skills required of a teacher: ability to pass through the academic program and aptitude towards teaching.

From 1971 to 1996 the aptitude test included giving a 10-15 minute authentic teaching demonstration in front of a real class of 3rd/4th graders. The applicants performance was evaluated by a panel of three senior teachers. After 1996 the teaching demonstration was replaced with individual and group interviews (Räihä, 2010). In broad terms, the test has been in its present form since then, even though individual universities have experimented occasionally. Today the explicit purpose of the aptitude test is to gauge "motivation, commitment, interpersonal and communication skills, and introspective ability" (University of Helsinki Application Guide, 2019).

This system is not without its critics, who point out that the tests don't have solid scientific grounding for measuring teaching aptitude and rest mostly on the subjective opinions of the evaluators (Räihä, 2010). An effective aptitude test should predict teacher effectiveness and provide information that is otherwise unobservable. These tests have never been subjected to any quantitative scrutiny of such criteria. Unfortunately, the current data limitations don't allow for such analysis at the moment.

Between 1997 and 2006, applicants were pre-selected into the aptitude test based on their matriculation exam performance, extra-curricular activities and relevant work experience. After being invited to the aptitude test, pre-selection points were reset, so that admission was entirely decided by the expert evaluation of teacher aptitude. After the 2007 reform, points from the pre-selection where added to points from the aptitude test making their disentanglement harder. However, since the aptitude test itself has remained very similar, the results apply more broadly.

Teaching is a feminized field in Finland as it is elsewhere. Before 1989 gender quotas required that at least 40% of students admitted in teacher training programmes had to be male. The quotas were abolished in 1989. The abolition resulted in a permanent shift in the gender ratio of graduating teachers from 60% female to 80% female, as shown in Figure A.4 (Appendix A). Additionally, there have been a plethora of smaller changes in almost yearly basis, which may have affected the pool of applicants as well as who was admitted and who graduated. Furthermore, any changes happening in any other programs may influence teacher training programs through individuals' dynamic response to incentives. On top of that, general macroeconomic conditions and trends can also have an impact. My objective is not to address the effect of these changes, but to describe the outcomes. All in all, the results presented in the next section encompass any and all changes that influence individuals' choices leading up to graduation.

This brief overview of teacher selection shows that the Finnish system seems to view academic ability and non-academic aptitude as complementary qualities for teachers. While academic ability has never been the main selection criterion it still serves to exclude the academically poorest performing applicants from being considered.

2.2 Matriculation Exams

Secondary education has two tracks in Finland: academic and vocational. High school academic track completion is an application requirement for teacher training programs. At the end of the academic track a nationwide standardized test known as the matriculation examination is administered to the entire graduating cohort. Performance in this test determines graduation and influences university admission through a point system specific to each university program. Most programs have three admission quotas in fixed proportions: One admits based only on matriculation exam performance, one admits through an entrance exam, and the final one combines points from both. Because of their influence on admission, the exams are generally considered high stakes. In this study, I use these test scores to proxy individuals' academic ability at the time of application to

tertiary education.

Until 2005, candidates were required to take exams on at least four subjects. Finnish and Swedish were compulsory. The two (or more) remaining elective subjects were chosen from general studies, basic mathematics, advanced mathematics and various foreign languages. Quite regularly, about 90% have taken the general studies exam and 60% have taken either math. From 2006, the general studies exam was split into multiple field specific exams (physics, chemistry,...) and Swedish was made elective. Multiple changes to the curricula of different subjects have been made along the years. However, in addition to multiple choice questions, the exams have always included essays (in languages) and full answers (in mathematics and general studies). Finnish, basic mathematics and advanced mathematics have retained a relatively constant content during the entire period. These three are the subjects I will use in my academic ability index.

2.3 Data

In addition to the matriculation exam test scores, I use three primary sources of data that correspond to the three sections in the results. For the first section describing the evolution of teacher test scores I use the *Register of Completed Education and Degrees* as the base data set. This registry contains the universe of all degrees higher than compulsory education completed in Finland between 1970 and 2012. I add to this data matriculation exam scores from 1967-2012. I focus on graduates from master's degree programs, with the exception of including teachers who graduated before the teacher training program was upgraded into a master's program in 1979. Furthermore, I exclude those high school graduates who did not participate in the Finnish language exams. This group consists almost exclusively of the Swedish speaking minority, who also have a largely separate tertiary education market. The final sample consists of about 30,000 teacher graduates and 350,000 master's degree graduates from other fields. For anchoring test scores to university admission I use the *Centralized Application Register* as the base data set. This register records every application into university programs made between 1992 and 2014.

The data is sufficiently detailed for years 2000-2014 excluding 2003. After excluding again non-Finnish applicants, I am left with 1.7 million observations (individual applications). The first column in Table 1 reports summary statistics for this sample.

Table 1: Summary Statistics for The Centralized Application Register

	Full sample	Non-teachers	Teachers 2000-2006
Women	0.60	0.63	0.81
Age	23.0	22.3	24.9
Admission rate	0.203	0.191	0.396[3]
N	1,710,021	177,768	10,563
	Counterfactual exercise		
		Non-teachers	Teachers
Anchored score of admitted		70.5	59.8
Anchored score of top		91.1	89.1

Note: In the counterfactual exercise, the "anchored score of admitted" refers to the average matriculation exam percentile of the admitted students in their respective high school cohorts. The "anchored score of top" refers to the average score of the top applicants where "top" is chosen by ordering the applicants according to their percentile score and choosing the K highest scoring applicants, where K is the number of seats in that program. All applicants are included without restrictions.

For the second section analyzing the association of academic ability with teacher aptitude, I make further limitations to the application data by including only programs that have at least 800 yearly applicants, which corresponds to the smallest teacher training program. Furthermore, I include only applicants who participate in entrance exams and programs that choose over 90% of their cohorts through entrance exams. This ensures that admission success of teacher applicants is not compared to applicants in other programs who were admitted directly by virtue of their matriculation exam performance. These sample choices turn out to be quantitatively inconsequential: the association between test

[3]Note that this is the admission rate for the pre-selected applicants. The overall admission rate, taking into account all applicants for teacher training programs in this period was 0.11.

scores and admission probability in non-teacher programs is always strongly positive. To maximize power, I use the largest possible sample. After these restrictions, I am left with 180,000 applicants in non-teacher programs between 2000 and 2014 (excluding 2003) and 10,500 applicants in teacher training programs between 2000 and 2006 (excluding 2003). As detailed in section 2, 2000-2006 was the period when points awarded for matriculation exam performance in the teacher pre-selection phase did not carry over to the aptitude test, which allows me to analyze the aptitude test separately from pre-selection. The last two columns in Table 1 report summary statistics for these samples.

In the third section describing teacher compensation, I add *Tax Records* for 2011-2015 to the degree registry and exam score data. The underlying measure of earnings I use in the mincer equation is the sum of labor earnings and entrepreneurial income. Finally, I use occupational categories for years 1995, 2000 and 2004-2012 to track the occupations of teacher graduates. Occupations are categorized at the five digit level, which I map into the custom categories that I show.

3 Results

3.1 Evolution of Teacher Test Scores

3.1.1 A Measure of Academic Ability

Matriculations exams measure academic ability in multiple dimensions [4]. My objective is to order university graduates unambiguously by their test score performance. To facilitate this, I reduce dimensionality to unity by constructing an anchored index: weights for each

[4] To some extent, they can also measure other characteristics, such as motivation and personality. See Izadi & Tuhkuri (2021)

subject × *grade* combination. A further reason to use an anchor is the ability to assign weights to missing test scores that arise due to selection into subject exams. Figure A.5 in Appendix A displays the raw scores (in percentiles) of teachers and other graduates conditional on taking the exam.

Anchoring test scores is typically used not only to reduce dimensionality, but also to give meaningful interpretations to the relative differences between scores. Years of schooling or income are typical left hand side variables (Bond and Lang, 2018). Because my interest is to get a measure of academic ability based on high school matriculation exam performance, an intuitive starting point is to ask what is the relative value of success in different high school subjects from the perspective of admission committees.

Consider an applicant i to program p. The anchoring regression regresses the probability of being admitted to program p on the additively separable combination of subjects and their grades:

$$E(admitted_{ipt}|applied_{pt}, grades_i) = \alpha_{pt} + \sum_s \sum_g \beta_{sg} D_{sgi} \tag{1}$$

$$s \in \{Finnish, Basic\ Math, Advanced\ Math\}$$
$$g \in \{NA, 0, 1, 2, 3, 4, 5\}$$

In the above linear probability model, α_{pt} is the *program* × *year* fixed effect. This is the level at which admission decisions are made. D_{sgi} is an indicator for getting grade g in subject s. Table 2 shows the fractions receiving each grade conditional on taking the exam.

Matriculation exam performance influences admission through two channels. Firstly, performance in the exams is likely correlated with factors that help applicants to also succeed

Table 2: Fraction of Candidates Receiving Each Grade.

Grade	0	1	2	3	4	5
fraction	5.2	11.8	17.2	24.1	20.3	17.6

Note: Numbers are averages pooled across subjects (Finnish, basic math, advanced math, general studies) and years (1967-2013). Fractions are relatively stable across subjects and years. The numbers correponds to letter grades (I, A, B, C, M, L) in the Finnish matriculation system, where 0 (I) is fail.

in entrance exams. Secondly, most programs award some admission points directly based on performance in the exams. This creates some mechanical association between exit exam grades and admission even in the absence of any correlation of the first type.

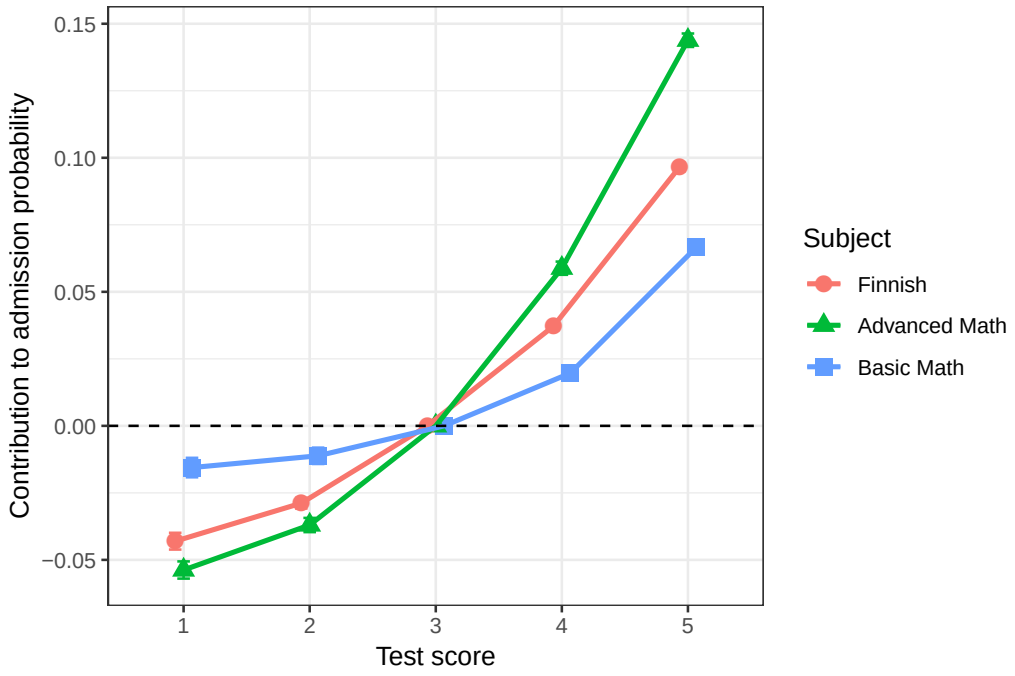


Figure 1: Anchoring Weights. Note: Displays β_{sg} coefficients from regression (1). Estimation sample includes all applicants in the university application register data from 2000-2014.

To maximize power, regression 1 is run for the entire 2000-2014 pool of applicants. Figure 1 displays the β_{sg} coefficients from this regression. Performing well in advanced mathematics is the strongest predictor for admission, followed by Finnish and finally by basic mathematics. An important caveat is that there is a lot of variation in the relative im-

portance of subjects between programs; mathematics being emphasized in STEM, and language in humanities and social sciences. The regression coefficients for a these fields are shown in Appendix A. Taking this into consideration, the coefficients from the pooled regression represent a weighted average from all programs and years.

3.1.2 Evolution of Teacher Test Scores

The anchored score for each individual is the fitted value from regression (1) subtracting the fixed effect. This score captures the contribution of the individual's test score performance on admission probability and can be interpreted as the ex ante (before entrance exams) quality of that individual as viewed from the perspective of an average admission committee. Next, I will apply the coefficients in Figure 1 to the entire high school academic track population. Weighting everyone's exit exam performance with the same weights ensures comparability across individuals in the same cohort. I further transform the anchor into percentile rankings within each high school cohort. Figure 2 shows the evolution of this metric over a 40 year period for master's degree graduates.[5]

It is immediately clear from Figure 2 that during the observation period, in each year the average teacher graduate is below the overall average of university graduates. While the overall average percentile has trended downwards, the teacher average has increased. The main caveat in this analysis is that individuals' scores are not comparable across years. Three factors can contribute to this issue: First, matriculation exam take up increases significantly over the period. Panel A of Figure A.1 shows an increasing secular trend in the fraction of birth cohorts that matriculate. This selection problem generates grade inflation whereby a larger share of each cohort is awarded top grades. Figure A.3 attempts to address this by imposing a "Manski-style" upper bound on the test scores (Manski, 1990). There I assume that non-matriculating individuals in each cohort are drawn from

[5]Before 1983, teacher training was a bachelor's degree program. Those are included in the sample.

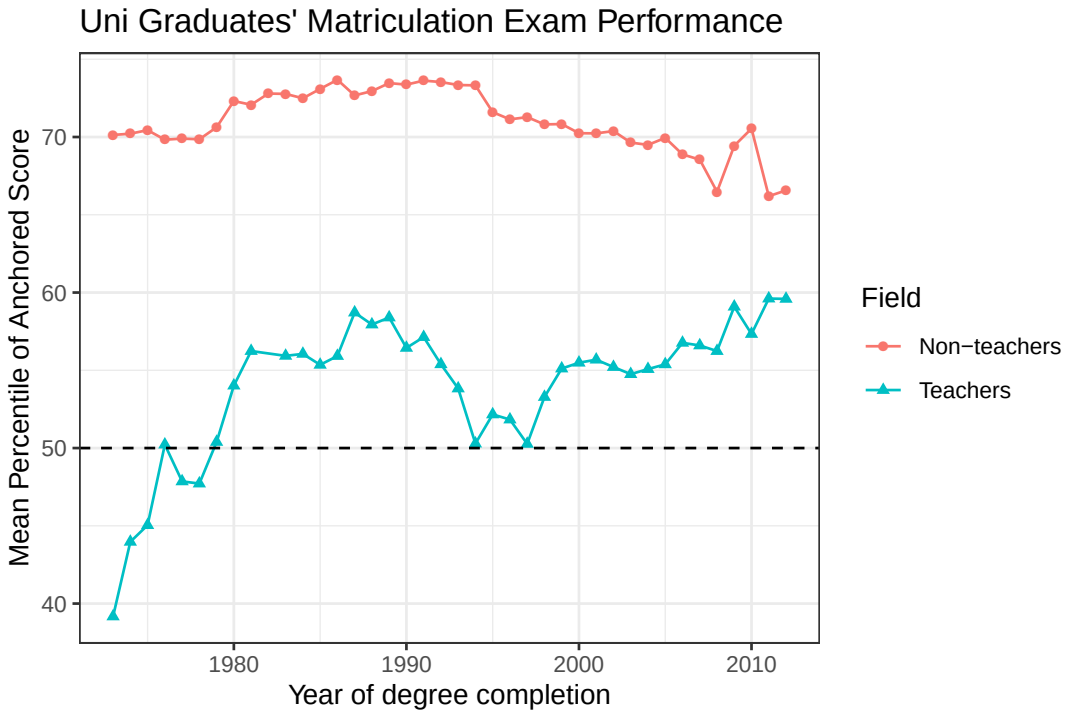


Figure 2: The Evolution of Teacher Test Scores. Note: In both series test scores are weighted using the same anchor from the pooled regression (1). An individual can appear only once in either series. If he has multiple degrees, only the highest degree is selected.

the bottom of the distribution (i.e. would have scored 0 had they taken up the exam). I then scale up the scores shown in figure 2 accordingly to produce Figure A.3. The result gives a starkly different impression of the trends: there is a secular decreasing trend in the test scores of both, teacher and non-teacher university graduates and no convergence. However, this mostly serves to demonstrate that university graduates in general perform very well relative to the overall population. The trend is unreliable because the upper bound assumption is more likely to be binding in earlier periods, when the academic track was less accessible to all high-ability individuals.

Second, changing selection into university can also influence the trends. However, panel B of Figure A.1 shows that the fraction of academic track graduates obtaining master's degrees has remained quite stable around 35% during the period. Hence, the relatively flat overall average of university graduates reflects the tendency that well performing students have always tended to obtain university degrees.

Finally, the population mean test score performance was also steadily increasing. Jokela et al. (2017) document this "Flynn-effect" for Finnish men. Because the difficulty level of the matriculation exam changes from year to year, without strong assumptions, there is ultimately no way of comparing the absolute levels of matriculation exam performance between two teachers of different matriculation cohorts.

The convergence of teachers' academic ability with the mean ability in Figure 2 is still noteworthy, because tertiary education programs are essentially competing for students of the same cohort. The Figure shows that teacher training programs are able draw students higher up from the distribution than they were in 1970's. This is likely to partly reflect the fact that the number of graduating teachers has increased less than the overall number of university graduates. This is shown in Appendix Figure A.2. Approximately the same amount of teachers were trained in the 1970's as now. Hence, teacher training programs can choose their students from a much larger pool than earlier. Overall it is difficult to say whether the trends in the Figure reflect such supply side factors more than the changes in the demand of applicants.

The early improvements in teacher test scores my also be related to the reforms made in teacher training programs during the 1970'. During this period teacher training was entirely moved from specialized training institutes into universities, until in 1983 teachers were required to complete a master's degree. This could have increased the prestige and appeal of the profession (Pennanen, 1997).

3.2 The Role of Academic Ability in Teacher Selection

There are three phases in the teacher selection process: application to the program, invitation to the aptitude test, and admission. Each phase contains a subset of individuals from the previous phase. Figure 3 shows the raw association between matriculation exam test scores and passing through each phase. The first panel confirms that only a small

fraction of the best performing academic track students apply to teacher training.[6] Despite the steep gradient, applicants are drawn from across the distribution. The middle panel shows that academic ability matters in the second phase: Only a small fraction of the worst performing applicants are invited to the aptitude test, whereas 40% of the best performing pass to the next phase (recall from section 2 that academic performance is not the only admission criterion). Finally, the last panel demonstrates that academic ability is not strongly associated with succeeding in the aptitude test. Excluding the first bin, the admission rate is between 35% and 45% in each test score bin for those who are interviewed. Next, I look at the invitation and admission phases in more detail by decomposing academic ability into its composite school subjects.

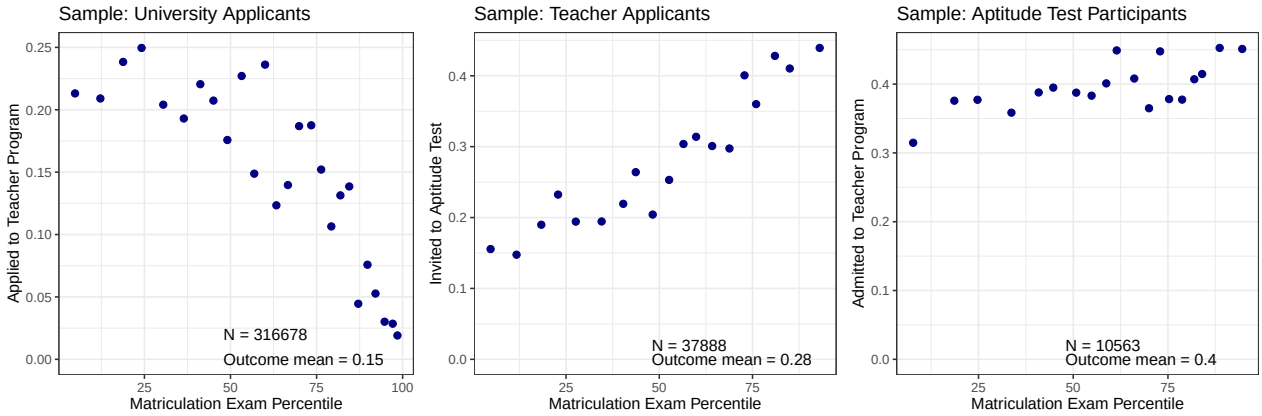


Figure 3: Teacher Selection Phases. Note: The y-axis in each panel is an average of the indicated dummy variable in that bin. The x-axis is the matriculation examination percentile derived from equation 1. Each bin has the same number of observations.

I estimate the association between subject specific academic ability and success in each phase by applying regression 1 to the relevant subsample of teacher training program applicants. Figure 4 displays side by side the β_{sg} coefficients from the regressions on three subsamples. The subsamples in the last two panels are the same as in Figure 3. For comparison, the first panel shows results for participants of written entrance exam takers

[6]This slope is inverted for many of the more prestigious programs, such as medicine and law.

from other programs.[7] Academic ability has clear correlation with admission in programs that use written exams, with advanced mathematics being the best predictor of admission. To demonstrate the effect of pre-selection on the pool of teacher applicants, the middle panel shows the results for regressing the invitation to participate in the aptitude test on matriculation exam test scores. The results confirm the importance of test scores, writing skills in particular, for pre-selection. This is consistent with pre-selection points being awarded directly for matriculation exam performance.

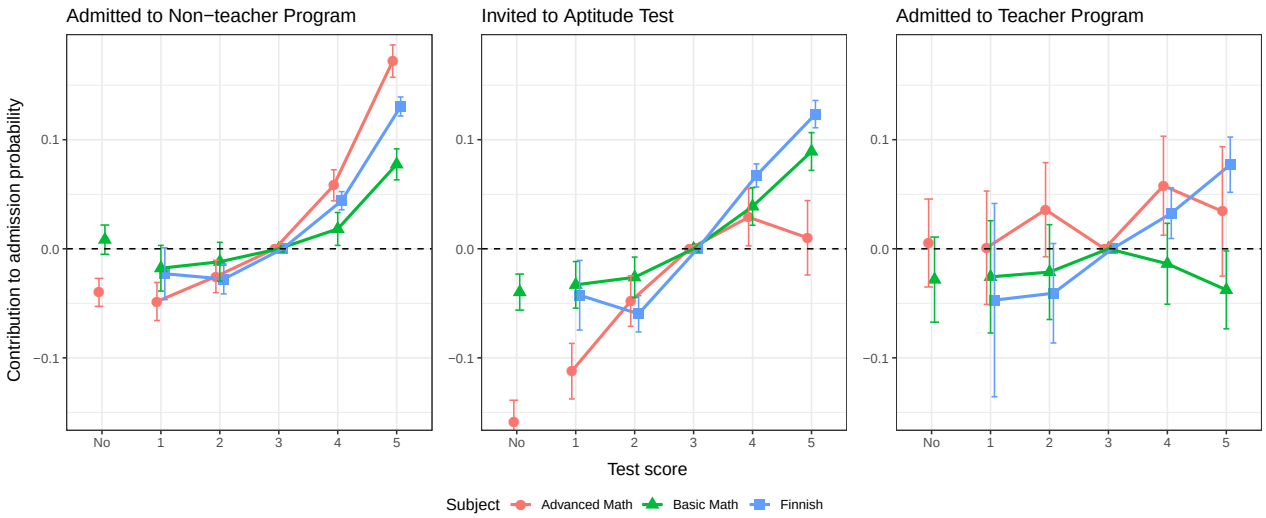


Figure 4: Academic Ability and Teacher Aptitude. Note: The left panel shows β_{sg} in the sample of applicants into non- teacher university programs. The middle panel uses invitation to the aptitude test as the dependent variable and shows β_{sg} in the sample of teacher training program applicants. The right panels shows β_{sg} in the sample of applicants participating in the aptitude test of teacher training programs. "No" is the coefficient for not taking the exam.

The last panel reaffirms that matriculation exam test scores and final admission to teacher training programs are only weakly associated. Contrary to other programs and pre-selection, Finnish language score seems to be the only significant predictor of admission. Due to smaller sample size, estimates for teacher programs are less precise. In the interviews, evaluators score applicants based on what they observe in-person. In particular,

[7]This subsample includes only exam taker of those programs that require written entrance exams, as opposed to Figure 1, which includes every applicant and program.

they don't see the applicant's prior test performance. However, questions in the individual interview are partly based on a written essay or motivation letter of the applicant. This could in part explain why language skills have predictive power but math does not. Additionally, observable behavior during the aptitude test, such as the ability to express oneself with coherence and lucidity, could plausibly be related to skills captured by the language test scores but not by math scores.

The main caveat of this part of the analysis is that we cannot be sure that teacher aptitude measures anything that is important to student outcomes. In other words, it relies on the hope that these expert evaluators, who are mainly faculty members and senior teachers, can identify potentially effective teachers. Even if it is not the case, this analysis serves as a case study into how teachers are selected in Finland. On the other hand the weak correlation between academic ability and teacher aptitude could just indicate that the aptitude test is white noise providing no valuable information. Since admission would then be random within the pool of test takers, that could lead to correlation patterns similar to the rightmost panel in Figure 4. However, given the statistically significant positive relationship between admission and Finnish language scores, this seems unlikely.

The overall pool of applicants in a given program naturally bound the distribution of admitted students. To get a sense of how binding exactly the pool of applicants is, consider a counterfactual exercise, where each program admits students using only weights from regression 1. The last row in Table 1 displays the average academic ability under this counterfactual. The result suggests that teacher training programs could improve the academic quality of admitted applicants, but choose not to. By admitting the academically best applicants, they could improve the average academic quality of their admitted students by 29 percentiles. By choosing otherwise implies that entrance exams provide (subjectively) valuable information particularly in teacher training programs.

The above analysis uses the additive linear probability model with grade indicators for ease of interpretation and visual presentation. As a robustness check, I repeat the analysis using percentile scores from the matriculation exams instead of grade dummies. The

percentile score is used to determine the grade as shown in Table 2, but it is not used in university admission decisions. Appendix A Table A.1 shows the results of a linear model as well as a probit model using the percentile scores. Again, all subjects are relevant predictors in other entrance exams, whereas in the teaching aptitude test, only Finnish is significant. A comparison of R^2 also shows that in the teacher training sample, the goodness of fit is an order of magnitude smaller. The model has close to zero predictive power on admission in teacher training, whereas for other programs the model explains about 7% of the variation in admission. [8]

Taken together, these results show that teacher aptitude, as measured by expert evaluation, correlates only weakly with academic ability conditional on being invited to participate in the aptitude evaluation. The admission context where these evaluations are made further imply that the aptitude tests give relevant information. Finnish teacher training programs are not out to enroll the best performing students.

3.3 Teacher Compensation

The previous section shows that teacher training programs attract a large pool of academically qualified applicants which affords the programs a high degree of selectivity. What makes teacher training programs so popular? To start out, I estimate a Mincer-type earnings equation with log earnings (ln y) regressed on gender (G), potential work experience (E), matriculation exam achievement (A), and a binary teacher indicator (T):

$$ln(y) = \alpha_0 + a_1 E + a_2 E^2 + a_3 G + a_4 A + \delta T + \epsilon \tag{2}$$

[8] The conclusions also carry through with a fully saturated probit model. Evaluating goodness of fit for the probit model with McFaddens pseudo R^2 gives similar results.

For each individual, y is measured by the sum of labor and entrepreneurial income. The equation is estimated in the cross section of university graduates pooled over years 2011-2015. Year fixed effects are included. The results are reported in Table 3, where the coefficient δ is the earnings premium for teachers. The result shows that on average, Finnish teachers earn 9% less than other university graduates with similar matriculation exam scores. This result is in stark contrast with Hanushek et al. (2019), who find that Finnish teachers have a 10% *positive* premium. They estimate the same equation using the PIIAC sample for Finland, which includes numeracy and literacy test scores. Different definitions of y most likely explain the discrepancy: I use gross annual earnings from the tax registry, whereas Hanushek et al. (2019) use self-reported hourly wages. [9]

The theoretical object of interest for this estimation is the compensating differential for teachers. The coefficient δ falls short of this interpretation, because observed earnings differentials are influenced by many factors besides non-pecuniary benefits/costs. Specifically, due to long vacations and shorter work weeks, the yearly earnings for teachers is likely to be low relative to their hourly wage rate [10]. Rough back of the envelope calculations across nine Master's degree fields suggest that teachers in Finland have hourly wages on par with lawyers and second to only medical doctors[11]. In summary, the discrepancy in the estimates reflects the fact that teacher profession offers relatively low earnings but high hourly wages.

[9] Other possible reasons for the discrepancy include data quality, my exclusion of non-primary school teachers and sample selection. Estimating one's own hourly wage is particularly difficult for teachers due to their complicated compensation structure and unregulated and unlogged total working hours (Hautamäki, 2015).

[10] Teachers average 32 hour work weeks with 21 hours of teaching. The remaining work time is allocated to preparation and administrative work (Hautamäki, 2015). Additionally, teachers enjoy 12-14 weeks of vacation each year.

[11] Average yearly earnings are divided by total hours of work. For teachers, I use 32 weekly hours multiplied by 40 weeks (assuming 12 weeks of vacation). For other fields, total hours equal 38 weekly hours multiplied by 47 weeks (assuming five weeks of vacation). These figures are likely to underestimate the true total hours for many non-teacher professions. As a consequence, the relative wage premium for teachers is possibly even higher.

Table 3: Mincer Coefficients

Dependent Variable:	log(Earnings)
Variables	
$Experience(a_1)$	0.076
	(0.0003)
$Experience^2(a_2)$	-0.002
	(0.00001)
$Gender(a_3)$	-0.273
	(0.002)
$Achievement(a_4)$	0.005
	(0.00003)
$Teacher(\delta)$	-0.094
	(0.002)
Outcome mean	10.66
Fixed-effects	
Year	Yes
Fit statistics	
Observations	1,873,788
R^2	0.07316
Within R^2	0.07283

Heteroskedasticity-robust standard-errors in parentheses

Figure 5: Sample includes all primary school teacher graduates who appear in the occupational data in years 1995, 2000, 2004-2012. Each data point (year) is calculated from the pooled cross-section across those years, so that individuals can appear in multiple years, but due to the range of data, never in all years. Earlier years are mostly populated by younger generations and later periods by older generations or graduates.

To make progress on evaluating the importance of non-pecuniary benefits, I describe the occupational paths of teacher graduates to show that teacher graduates in Finland stick to their career choice. Figure 5 shows the average occupation of a primary school teacher graduate relative to their graduation year [12]. Only about 5% of primary school teacher graduates are employed in non-education occupations after their graduation. Further-

[12]Means are estimated using pooled cross section from years 1995, 2000, 2004-2012. Data range prevents following any single teacher through their entire career.

more, this rate remains constant throughout the years with 10% of graduates ending their careers as principals. Likewise, the non-employment rate is negligible for all but the recently graduated. Overall, teacher training programs seem to offer a stable and low risk career with low earnings but high wage rate, and some opportunities for promotion.

Low occupational exit is also consistent with qualitative accounts of potentially valuable non-pecuniary benefits associated with Finnish teachers, such as public respect and high teacher autonomy (Tirri, 2014; Sahlberg, 2011). Taken together, the high wage premium, high employment rate and positive non-pecuniary benefits make for an obviously attractive package for risk averse individuals [13] with preference for low labor supply in the intensive margin.

4 Discussion

This paper looks at the relationship between teacher aptitude and academic ability. First I constructed a measure of academic ability and used it to rank university graduates from the last 40 years. I found that primary school teacher's consistently rank below graduates from other fields, but have risen in the same period over 20 percentiles in the distribution of high school performance. Second, I demonstrated that the discrepancy in test score performance between teachers and others is not due to lack of high performing applicants, but rather due to the particular nature of the aptitude test that is used to screen applicants in teacher training programs: Success in the aptitude seems to only correlate with test performance in Finnish language and not in mathematics. In addition, the correlation is weak relative to entrance exams in other programs. This leads to the rejection of many applicants that perform well especially in mathematics.

[13] See Lang and Palacios (2018).

Third, I presented evidence that teachers in Finland enjoy long and stable careers in education without fear of unemployment. Due to long vacations and short work weeks teachers have relatively high hourly wages despite earning less than their comparable peers. Looking at the whole package, a career in teaching can look attractive even without a strong preference for teaching. This could help explain the consistently high application numbers. In my view, an aptitude test can be particularly useful when the pool of qualified applicants is large and the "calling to teach" is not the only motivation to apply for teacher training. At best, it can be an effective way to screen motivated applicants. At worst, in a system with poor teacher compensation, it would provide little information if the programs can't attract anyone but those with the strongest preference for teaching.

This analysis provides unique insight into the teacher selection process in Finland. It seems that teacher education programs are not interested in getting the academically most successful candidates in their programs, instead using the aptitude test to select principally on non-academic ability. However, the fact that they pre-select applicants into the aptitude test partly based on academic ability suggests that they view the two as complements for effective teachers. Is the Finnish model successful in selecting good teachers? In other words does the aptitude test predict teacher effectiveness as measured by student outcomes? While Finland's PISA success may suggest so, it remains an open empirical question for an important research agenda.

There are two pieces of data missing to enable the study of this question. Currently, the available information on the success in the aptitude test is limited to admission. Since all teachers haves been admitted by definition, we need more fine grained scores or rankings to be able to distinguish high scoring teachers from low scoring teachers. This data could in principle exist in the application register of individual universities. Second, we need to be able link these teachers to their students. There is a gap for this kind of data in Finland. Recently, some progress has been made to obtain student-teacher match data in large Finnish municipalities, but we need data that reaches further to the past to obtain information on meaningful student outcomes.

Appendix A Figures and Tables

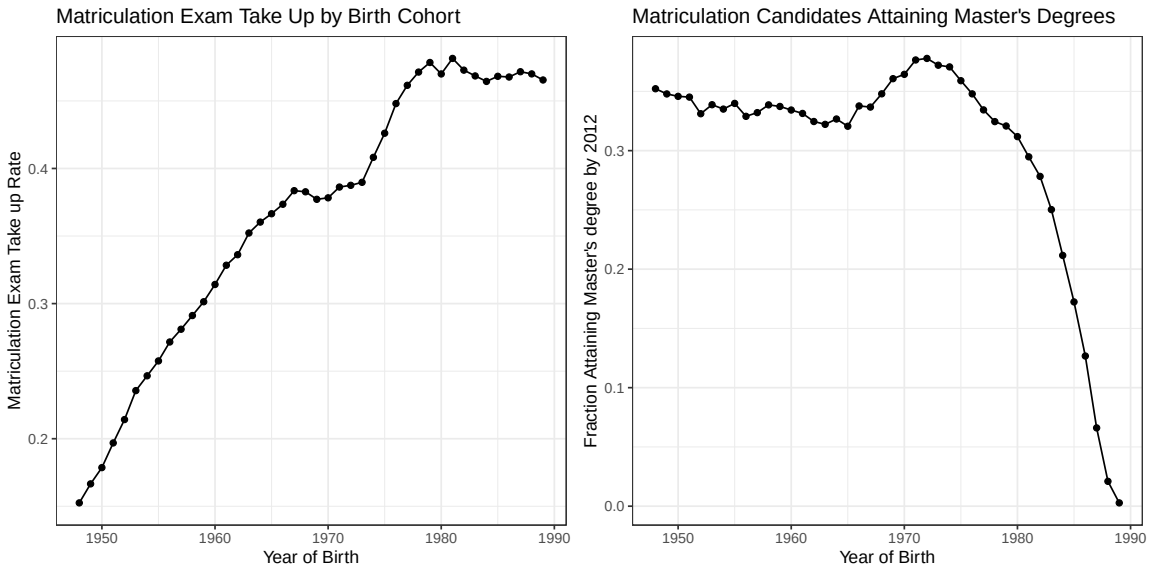


Figure A.1: Educational Attainment Descriptive Trends. The first panel shows the matriculation exam take up rate as a fraction of each birth cohort. The second panel shows the master's degree completion rate as a fraction of all matriculation candidates. The rate trends towards zero, because the would-be-graduates of later cohorts were still studying for their degrees in 2012 (the last year of observation).

Table A.1: The Impact of Test Scores on University Admission

	Dependent variable:			
	Admitted			
	OLS		*Probit*	
	Non-teachers	Teachers	Non-teachers	Teachers
Finnish	0.200***	0.128***	0.899***	0.342***
	(0.004)	(0.020)	(0.016)	(0.054)
Adv. math	0.291***	0.049	1.061***	0.127
	(0.005)	(0.035)	(0.020)	(0.091)
Basic math	0.143***	−0.008	0.700***	−0.019
	(0.005)	(0.024)	(0.024)	(0.063)
Took adv. math	−0.068***	−0.009	−0.253***	−0.026
	(0.004)	(0.021)	(0.015)	(0.057)
Took basic math	−0.067***	0.002	−0.366***	0.004
	(0.004)	(0.020)	(0.019)	(0.054)
Program x year effects	yes	yes	yes	yes
Observations	177,768	10,563	177,768	10,563
R^2	0.112	0.029		
Adjusted R^2	0.112	0.027		
Residual Std. Error	0.371 (df = 177623)	0.482 (df = 10541)		

*p<0.1; **p<0.05; ***p<0.01

Note: This table reports the coefficients for two continuous variable versions of equation (1). The specifications include indicators for missing exam results in math ("Took math"). The indicator is interacted with the corresponding continuous variable. The coefficients of the interactions are shown in "Adv. math" and "Basic math". Main effects for math are omitted (they are not estimable). For missing test scores, zero is imputed. The particular imputed value does not change the estimation results.

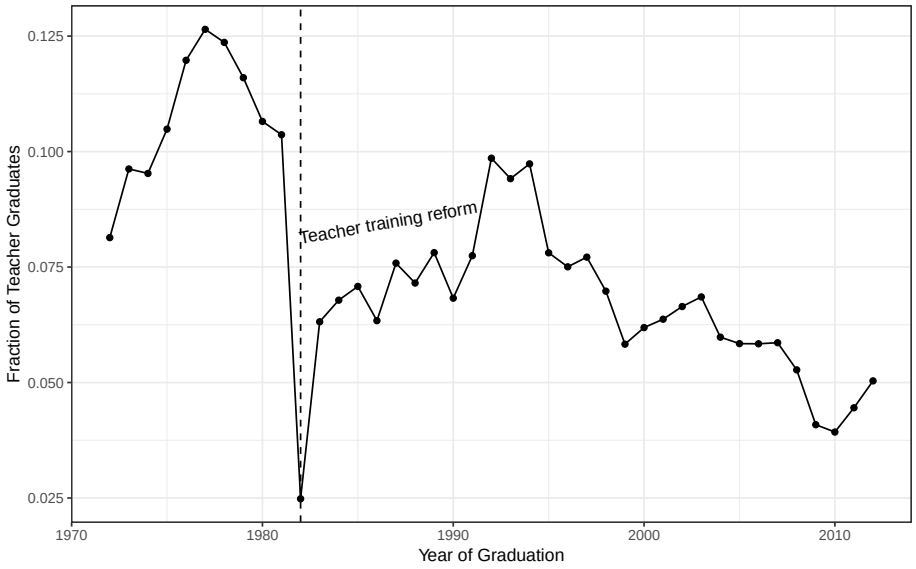


Figure A.2: Teachers as Fraction of All Master's Degrees. Note: The figure shows the proportion of teacher graduates from all master's degree graduates in years 1972-2012. There is a sharp drop in 1983, when the reform of 1979, which converted teacher training into master's degree programs, produces its first graduates.

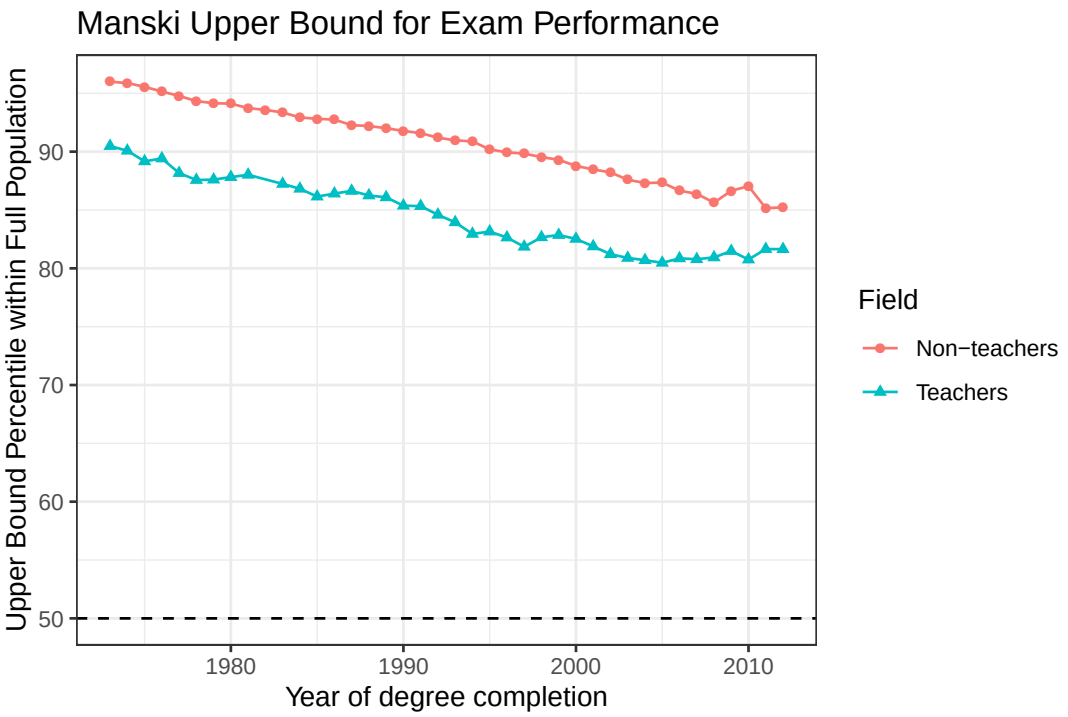


Figure A.3: The Evolution of Teacher Test Scores: Upper Bound. Note: Non-matriculating individuals in each cohort are assumed to come from bottom of the distribution (i.e. would have scored 0 had they taken up the exam). The anchored test scores in Figure 2 are then scaled up by recalculating the percentile ranks with the imputed missing tail of the distribution included in the sample.

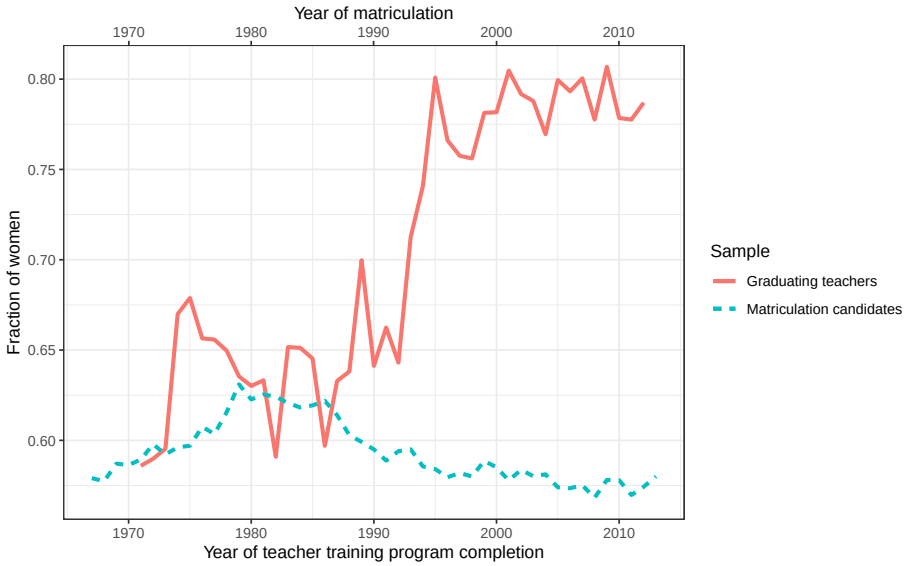


Figure A.4: Fraction of Graduating Women. The red solid line shows fraction of women graduating from teacher training programs. The turquoise dashed line shows the fraction of women in each cohort of high school academic track graduates.

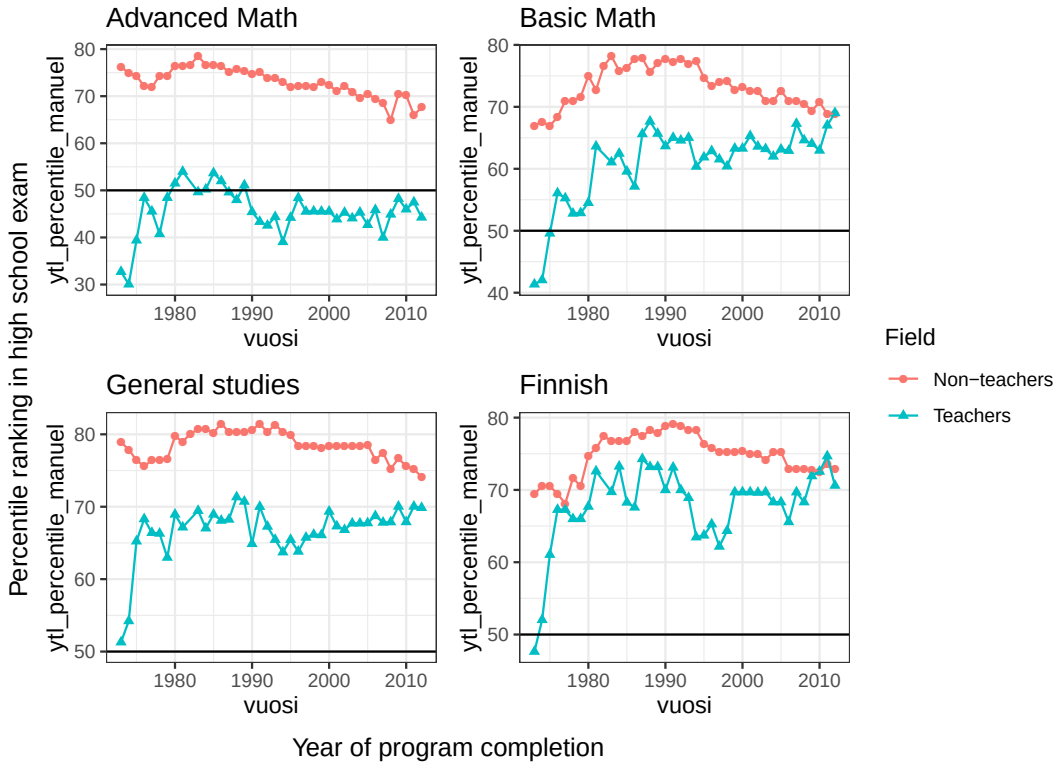


Figure A.5: Uni Graduates' Test Scores. Note: Shows the average raw percentile rank of the indicated group in each high school academic track subject conditional on exam participation. The sample includes university graduates from 1978-2012.

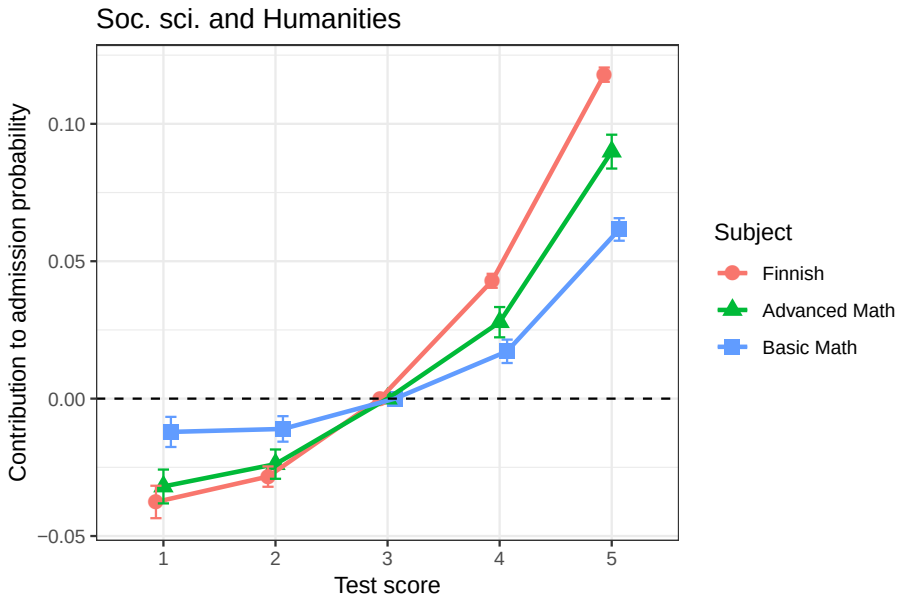


Figure A.6: Anchoring Weights: Humanities and Social Sciences. Note: Displays β_{sg} coefficients from regression (1). Estimation sample includes all applicants to Humanities and Social Sciences. I pool university application register data from 2000-2014.

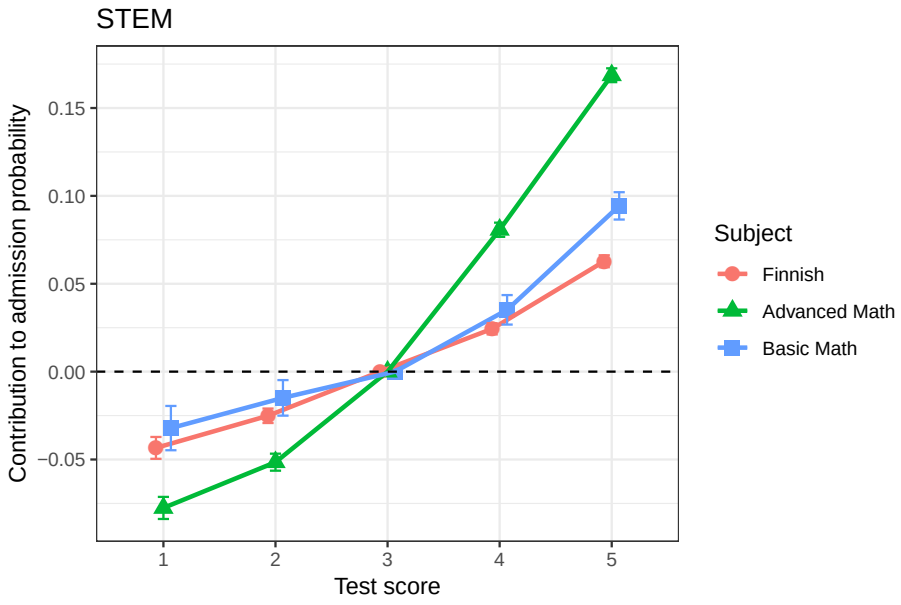


Figure A.7: Anchoring Weights: STEM. Note: Displays β_{sg} coefficients from regression (1). Estimation sample includes all applicants to fields of Health, Technology and Science. I pool university application register data from 2000-2014.

References

Paul Kihn Auguste, Byron and Matt Miller. Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching. *McKinsey and Company*, 2010.

Marigee P Bacolod. Do alternative opportunities matter? the role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89(4):737–751, 2007.

Jan Bietenbeck, Marc Piopiunik, and Simon Wiederhold. Africa's skill tragedy does teachers' lack of knowledge lead to low student performance? *Journal of human resources*, 53(3):553–578, 2018.

Timothy N Bond and Kevin Lang. The black–white education scaled test-score gap in grades k-7. *Journal of Human Resources*, 53(4):891–917, 2018.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–79, 2014.

Will Dobbie. Teacher characteristics and student achievement: Evidence from teach for america. *Unpublished manuscript, Harvard University*, 2011.

Erik Grönqvist and Jonas Vlachos. One size fits all? the effects of teachers' cognitive and social abilities on student achievement. *Labour Economics*, 42:138–150, 2016.

Eric A Hanushek, Marc Piopiunik, and Simon Wiederhold. The value of smarter teachers international evidence on teacher cognitive skills and student performance. *Journal of Human Resources*, 54(4):857–899, 2019.

A Hautamäki. Kokonaistyöaika opettajan ammatillisen kasvun ja koulun kehittämisen keinona: Yleissivistävän koulutuksen opettajien työaikajärjestelmät euroopassa.[overall working hours of teachers as a mean to develop school and teachers' professional growth. the working time systems of compulsory schools in europe. in finnish.] sitran selvityksiä 85, 2015.

C Kirabo Jackson, Jonah E Rockoff, and Douglas O Staiger. Teacher effects and teacher-related policies. *Annu. Rev. Econ.*, 6(1):801–825, 2014.

Markus Jokela, Tuomas Pekkarinen, Matti Sarvimäki, Marko Terviö, and Roope Uusitalo. Secular rise in economically valuable personality traits. *Proceedings of the National Academy of Sciences*, 114(25):6527–6532, 2017.

Tuomas Kahila and Mikko Ojanen. Pitkät lomat ja lyhyet työpäivät-luokanopettajan ammatin parhaat puolet: onnistuminen ja epäonnistuminen opettajan työssä, 2003.

Timo Kumpulainen. Opettajat ja rehtorit suomessa 2016. *Raportit ja selvitykset*, 2, 2017.

Kevin Lang and Maria Dolores Palacios. The determinants of teachers' occupational choice. 2018.

Andrew Leigh and Chris Ryan. How and why has teacher quality changed in australia? *Australian Economic Review*, 41(2):141–159, 2008.

Charles F Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

Johannes Metzler and Ludger Woessmann. The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics*, 99(2):486–496, 2012.

Stephen Nickell and Glenda Quintini. The consequences of the decline in public sector pay in britain: a little bit of evidence. *The Economic Journal*, 112(477):F107–F118, 2002.

Mikko Pennanen. Kasvatustiede yliopistollisena oppiaineena suomessa vuosina 1852-1995. *Kasvatustieteenn pro gradu tutkielma. Turun yliopisto. Myös verkko-osoitteessa: http://users. utu. fi/mikpen/gradu. html*, 1997.

Pekka Räihä. Koskaan et muuttua saa! luokanopettajakoulutuksen opiskelijavalintojen uudistamisen vaikeudesta. 2010.

Jonah E Rockoff, Brian A Jacob, Thomas J Kane, and Douglas O Staiger. Can you recognize an effective teacher when you recruit one? *Education finance and Policy*, 6 (1):43–74, 2011.

Pasi Sahlberg. The professional educator: Lessons from finland. *American educator*, 35 (2):34–38, 2011.

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