Teacher selection in Finland. Does academic ability predict teacher aptitude?

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 two 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.

Ramin Izadi

Aalto University

Working paper

June 13, 2019

Contents

1 Introduction **1**

2 Institutional background and data **4**

2.1 Teacher selection process 4

2.2 High school exit exams 6

2.3 Data 7

3 Results **9**

3.1 Evolution of teacher test scores 9

3.1.1 A measure of academic ability 9

3.1.2 Evolution of teacher test scores 11

3.2 Association of academic ability with teacher aptitude 13

4 Discussion **16**

Appendix A Figures **18**

Appendix B Data **21**

References **22**

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 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. 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 high school. I use it to track changes in the relative positions of teachers vs. other graduates. Second, I relate the high school test performance of teachers applicants to success in the aptitude test of teacher training programs. Is academic ability associated with success 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.

Through this analysis, I come to two 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 performance. In 1973, the average teacher graduate came from the 40th percentile of high school 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.

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 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 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)

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)).

In quest of answering these questions, economist too have turned towards countries whose students perform well in international comparisons. Hanushek et al. (2014) 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. The 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.

The main contribution of this paper is to demystify Finland as the paragon of teacher selection by providing a first quantitative look at selection process of Finnish teachers. In light of the findings by Hanushek et al. (2014), 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 worst 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 exit 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 examination, 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: passing 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 high school test score performance, extra-curricular activites 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 also in Finland. In 1989 gender quotas were abolished in teacher training. Before that, at least 40% of admitted students had to be male. The abolition resulted in a permanent shift in the gender ratio of graduating teachers from 60% female to 80% female, as shown in figure 4 (Appendix A). Additionally, there have been a plethora of smaller changes in almost yearly basis, which may affect the pool of applicants as well as who is admitted and who graduates. Furthermore, any change 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 worst performing applicants to ever be considered.

2.2 High school exit exams

High school completion is an application requirement for teacher training programs. In this study, I use test scores from the high school exit exams to proxy individuals' academic ability at the time of application to tertiary education. Secondary education has two tracks in Finland: academic and vocational. Throughout this paper, I refer to the academic track as "high school". At the end of high school a nationwide standardized test 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 high school 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.

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, the exams have always favored 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

I use two primary sources of data that correspond to the two sections in the results. For the first section describing the evolution of teacher test scores I use the degree registry 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 high school exit 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 1983. 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 the anchoring regressions 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 has sufficiently detailed information for years 2000-2014 excluding 2003. After excluding again non-Finnish applicants, I am left with 1.7 million observations.

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

[1]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.

Table 1: Summary statistics

	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[1]
N	1,710,021	177,768	10,563
	Counterfactual exercise		
Anchored score of admitted	70.5		59.8
Anchored score of top	91.1		89.1

that choose over 90% of their cohorts through entrance exams. This ensures that admission success of teacher applicants are not compared to applicants in other programs who were admitted directly by virtue of their high school test performance. These sample choices turn out to be quantitatively inconsequential: the association between high school test scores and admission probability in non-teacher programs is always strongly positive. 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). Figure 1 summarises the three different samples formed from the centralized application register

3 Results

3.1 Evolution of teacher test scores

3.1.1 A measure of academic ability

High school exit exams measure academic ability in multiple dimensions [2]. 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 *subject* × *grade* combination. A further reason to use an anchor is the ability to give weights to missing test scores that arise due to selection into subject exams. Figure 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 test 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:

[2] To some extent, they can also measure other characteristics, such as motivation and personality.

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

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

In the above linear probability model, α_{pt} is the $program \times year$ 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.

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.

High school performance influences admission through two channels. Firstly, performance in the exit exams is likely correlated with factors that help applicants to also succeed in entrance exams. Secondly, most programs award some admission points directly based on performance in the exit exams. This creates some mechanical association between exit exam grades and admission even in the absence of any correlation of the first type.

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 importance of subjects between programs; mathematics being emphasized in STEM, and language in humanities

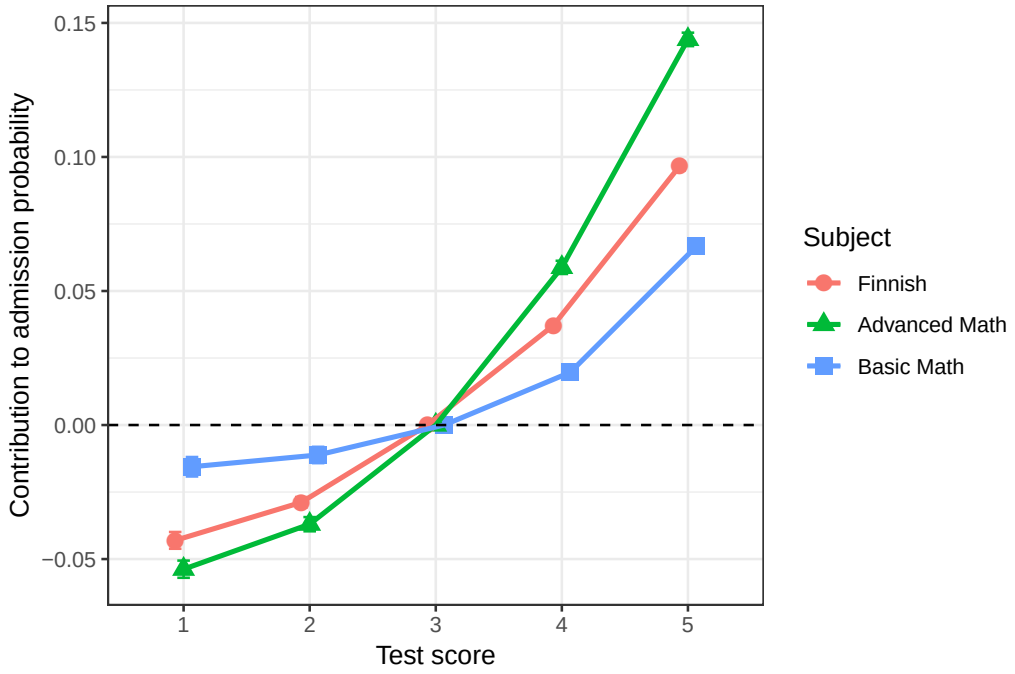

Figure 1: Selected coefficients from regression (1). Note: Estimation sample includes all applicants in the application register data from 2000-2014.

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 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 evolution of this metric

over a 40 year period for master's degree graduates.[3]

Figure 2: The evolution of teacher test scores. Note: Both series are weighted using the same anchor from the pooled regression. An individual can appear only once in either series. If he has multiple degrees, only the highest degree is selected.

It is immediately clear from figure 2 that during the observations period, the average teacher graduate is below the overall average of university graduates in each year. 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, due to changing selection into high school (there is an increasing secular trend in the fraction of birth cohorts that graduate from high school). The fraction of high school graduates obtaining master's degrees has remained quite stable around 30% 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.

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

The convergence of teachers' academic ability with the mean ability 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 also likely to partly reflect the fact that the number of graduating teachers has increased much less than the overall number of university graduates. 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 Association of academic ability with teacher aptitude

I estimate the association between teacher aptitude and academic ability by applying regression 1 on the subsample of applicants who participated in the aptitude test of teacher training programs between 2000 and 2006. Figure 3 displays side by side the β_{sg} coefficients from the regression on participants in the teacher training programs against the participants of written entrance exam takers from other programs [4]. Academic ability has clear correlation with admission in programs that use written exams, with advanced mathematics being the best predictor. However in the teacher aptitude test there seems to be little association with high school test scores and admission. Contrary to other

[4]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.

programs, Finnish language score seems to be the only significant predictor. Due to smaller sample size, estimates for teacher programs are less precise.

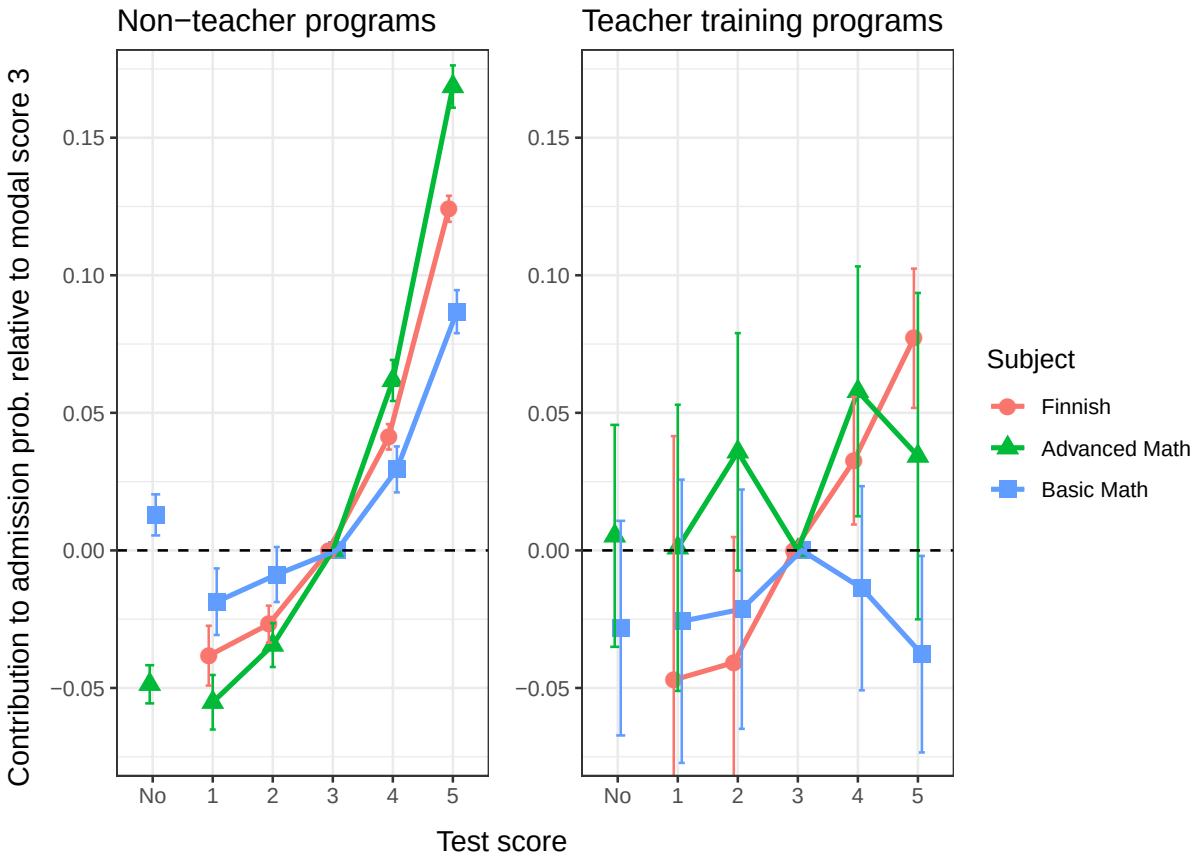

Figure 3: Academic ability and teacher aptitude. Note: The left panel shows β_{sg} in the sample of applicants into non- teacher university programs. The right panels shows β_{sg} in the sample of applicants participating in the aptitude test of teacher training programs. "No" is the coefficient for not taking the exam.

Evaluators score applicants based on what they observe in-person. In particular, 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 partly 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 right panel of figure 3. 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 high school exit 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 3 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. [5]

Taken together, these results show that teacher aptitude, as measured by expert evaluation, correlates only weakly with academic ability conditional on some minimum level of achievement. 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.

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 any other field, 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.

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

[5] 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.

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 teacher 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

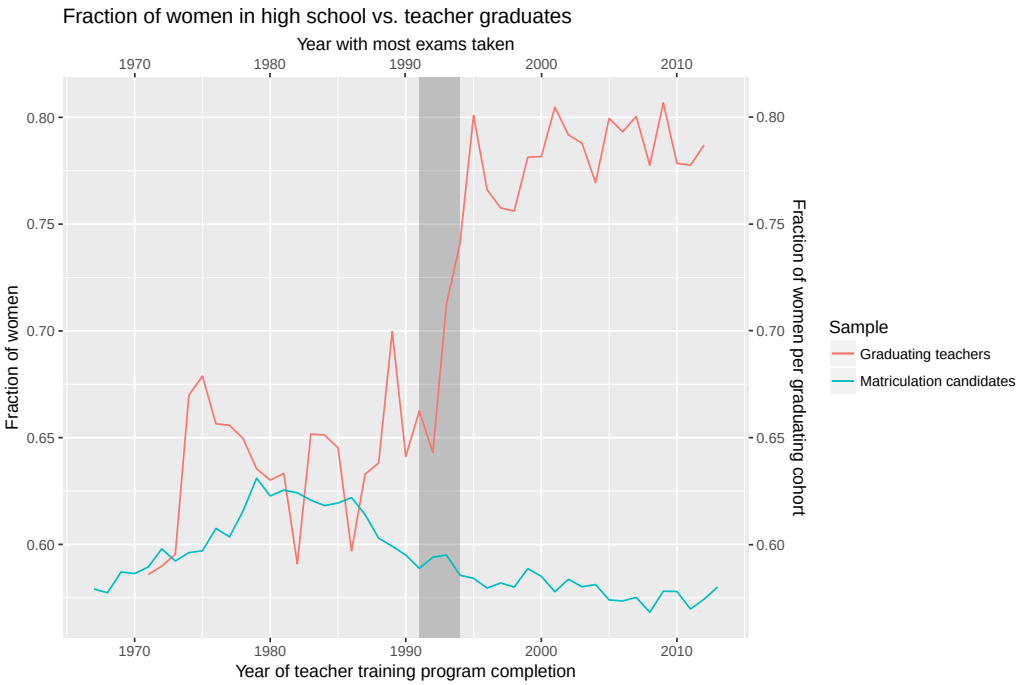

Figure 4: Fraction of women graduating from teacher training programs. The turquoise line shows the fraction of women in each cohort of high school graduates.

Table 3: Regressions results using test score percentile rankings as independent variables.

	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)		
Note:				*p<0.1; **p<0.05; ***p<0.01

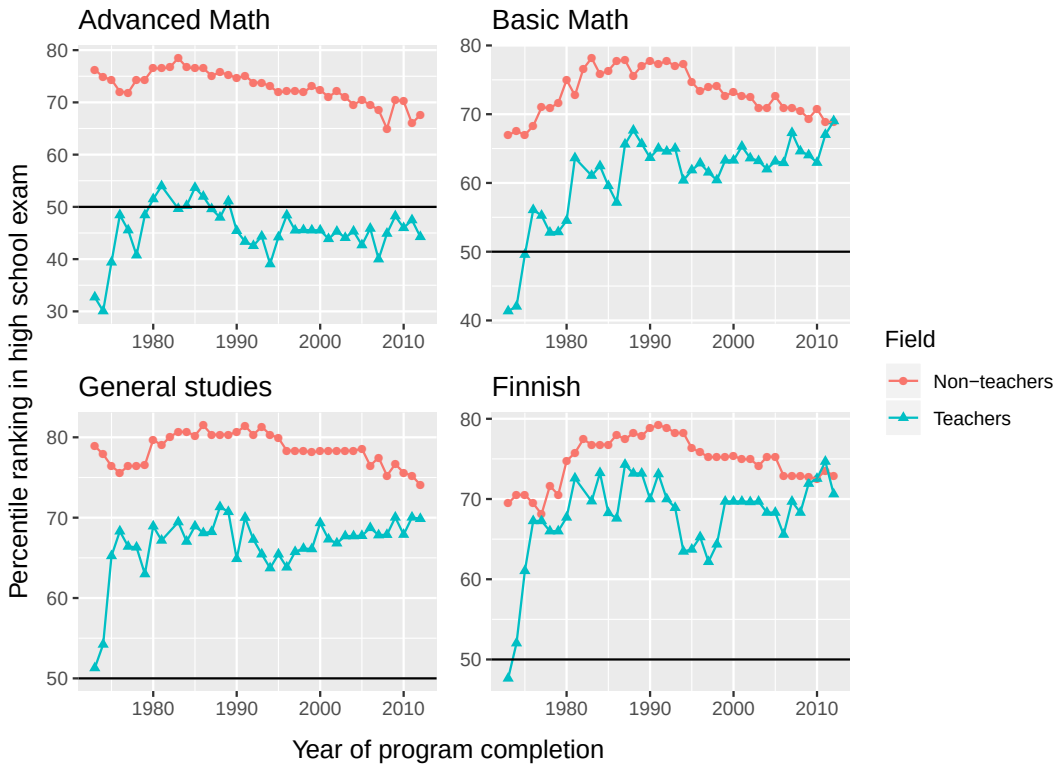

Figure 5: Note: Raw percentile ranking in each subject conditional on participation.

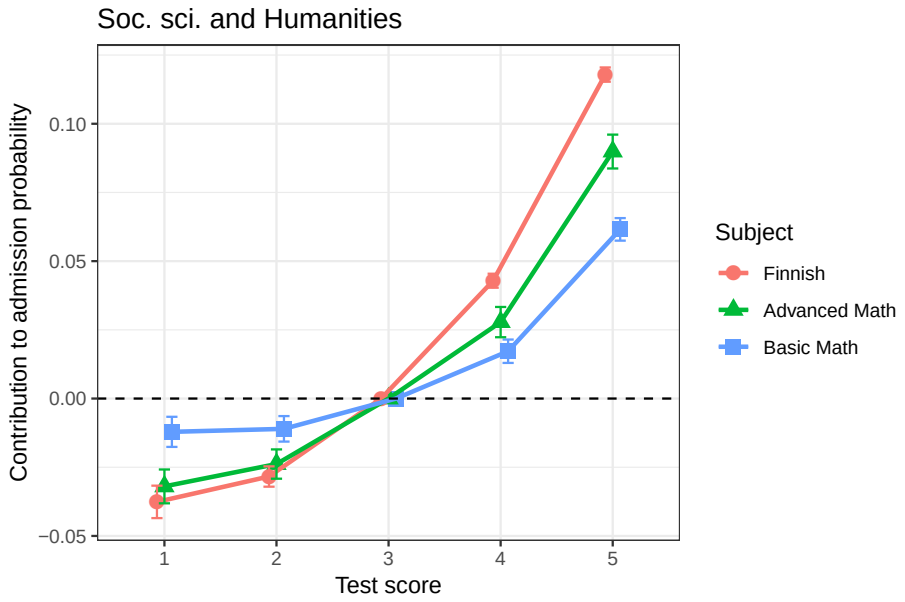

Figure 6: Note: Coefficients from the anchor regression estimated for the sample of applicants to Humanities and social sciences. I pool application register data from 2000-2015.

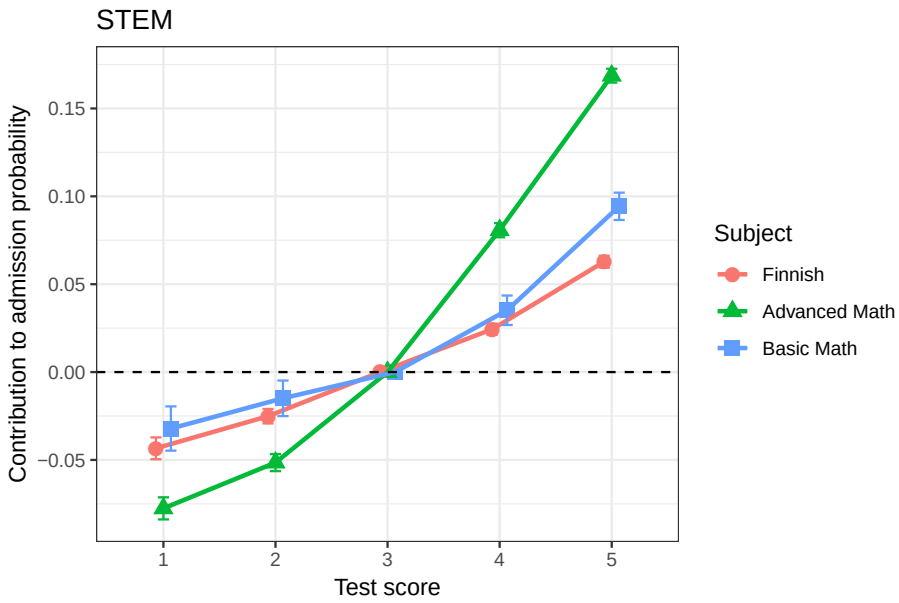

Figure 7: Note: Coefficients from the anchor regression estimated for the sample of applicants to STEM fields. I pool application register data from 2000-2015.

Appendix B Data

WIP

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. Technical report, National Bureau of Economic Research, 2014.

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

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.

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.