VATT Working Papers 108

Effectiveness of Private and Public High Schools: Evidence from Finland

Mika Kortelainen
Kalle Manninen

VATT INSTITUTE FOR ECONOMIC RESEARCH

VATT WORKING PAPERS

108

Effectiveness of Private and Public High Schools: Evidence from Finland

Mika Kortelainen
Kalle Manninen

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Helsinki 2018

Mika Kortelainen, VATT Institute for Economic Research, Helsinki, Finland,
E-mail: mika.kortelainen@vatt.fi, Phone: +358 40 304 5541.

Kalle Manninen, University of Jyväskylä.

We would like to thank Henrik Jordahl, Roope Uusitalo and the seminar participants of CESifo Economic Studies Conference on Public Sector Outsourcing for useful comments. The first author was supported by Strategic Research Council at the Academy of Finland Grant 303687.

ISBN 978-952-274-221-6 (PDF)

ISSN 1798-0291 (PDF)

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Helsinki, May 2018

Effectiveness of Private and Public High Schools: Evidence from Finland

VATT Institute for Economic Research
VATT Working Papers 108/2018

Mika Kortelainen – Kalle Manninen

Abstract

A number of papers have compared the effectiveness of private and public schools in different institutional settings. However, most of these studies are observational and do not utilize experimental or quasi-experimental design to evaluate the value-added or the effectiveness of private schools in comparison to public schools. This study focuses on private and public high schools in Helsinki, the capital city of Finland. We use two different methods to compare private and public schools, value-added estimation and regression discontinuity design (RDD). Although based on somewhat different assumptions, both methods allow us to evaluate the causal effect of private schools on the exit exam results in high school. We find that private schools perform marginally better than public schools, but the difference in performance is small and statistically insignificant according to both methods. Various robustness and validity checks strengthen our RDD results and the validity of the discontinuity design.

Key words: education, public sector outsourcing, regression discontinuity, school choice, value-added

JEL classes: H44, I21, I26, J24

1. Introduction

The question of the extent to which educational outcomes are affected by schools is fundamental for improving education policies. One interesting and policy-relevant question related to school effects concerns private schools and their effects on student performance. Importantly, private schools have recently become more common in many countries both at the primary and secondary school level. Even though a large number of studies have investigated the effect of private schools or compared the effectiveness of private and public schools, most earlier papers do not utilize experimental or quasi-experimental designs. In this paper, we utilize both a value-added estimation method and a quasi-experimental design to study the effectiveness of private and public high schools in Finland.

Finnish high schools provide an interesting setting to study private and public schools for several reasons. First of all, the institutions and data sets we utilize allow us to estimate a value-added model that has been utilized in a number of studies comparing schools (e.g. Raudenbush and Willms 1995, Reardon and Raudenbush 2009, Chamberlain 2013, Deming 2014). Secondly, Finland has a nationwide application system for high schools that matches the preferences of the applicants and preset student quotas for the schools in a centralized manner. The selection of students into high schools is based on students' revealed preferences and grade point averages (GPAs) according to the announced available slots in the high schools. This system offers a quasi-experimental regression discontinuity design (RDD) that we exploit using data on the students' revealed preferences and GPAs. While RDD has been used in many educational papers, to our knowledge it has rarely been utilized to compare the effectiveness of private and public schools. Thirdly, in the Finnish setting private schools are autonomous, but cannot make a profit, i.e. they are not-for-profit schools by law. The institutional structure might affect the selection of students and also how private schools organize teaching and thus possible school effects. In our empirical analysis, we concentrate on high schools in Helsinki, the capital city of Finland.

Our contribution to the relatively large literature on school choice is both empirical and methodological. Similarly to many other studies, we compare one type of choice (private school) to another (public school). However, few studies have examined the effects of different school types within the same urban region or city. For example, Berends and Waddington (2018) recently compared charter, magnet, private, and traditional public schools with each other in Indiapolis. Yet, similarly to several other studies concentrating on one urban area or city, their approach is based on an observational research design (fixed effect model). Moreover, while some papers have utilized voucher lotteries (e.g. Abdulkadiroğlu et al. 2018, Hahn et al. 2018) to study the causal effect of private schools, it has not been common to compare the results of observational design (e.g. a value-added model) to the results based on an experimental or quasi-experimental research design in this context.[1] Finally, although several papers have investigated high schools using Finnish data (e.g. Kirjavainen and Loikkanen 1998, Häkkinen et al. 2003, Kanninen 2013, Tervonen 2016), our paper is the first to compare private and public high schools. Moreover, only one previous study (Kortelainen et al. 2014) has estimated value-added models using data on Finnish schools.

Our results can be summarized as follows. Using both a value-added model and a regression discontinuity design, we find that private schools perform marginally better than public schools. Yet, the differences in estimates are generally very small and statistically insignificant. Our main results are robust to various kinds of modeling choices in RDD such as the choice of bandwidth and changes in the functional form. In addition, validity checks of the discontinuity design reinforce our main results and conclusions.

The rest of the paper is organized as follows. Section 2 describes the Finnish schooling system, private high schools in Finland and the high school application system, while Section 3 presents the data. In Section 4, we explain our empirical strategies based on value-added estimation and regression discontinuity design. Section 5 presents the econometric results and Section 6 concludes.

[1] For example, Hahn et al. (2018) use a random assignment of students into private and public high schools in Seoul to evaluate the effectiveness of private and public high schools. However, they do not compare their results to estimates given by a value-added model. Interestingly, Deming (2014) has recently utilized a public school choice lottery in Charlotte-Mecklenburg to test the validity of school value-added models.

2. Institutional background

2.1 The Finnish schooling system

In Finland education is compulsory for everybody between 7 and 16 years of age. In practice, almost everybody enters comprehensive school at age 7 and graduates at age 16. Comprehensive school consists of primary school (grades 1 to 6) and lower secondary school (grades 7 to 9). After graduation from comprehensive school around 50 % of each cohort enters a general upper secondary school, i.e. the academic track of upper secondary school[2]. In the following, we will simply use the term high school to refer to the academic track of upper secondary school.

The basic curriculum in a high school consists of general knowledge subjects such as Finnish, foreign languages, mathematics and various human and natural sciences. Studies in each of these subjects are broken up into courses. Generally, students have substantial freedom of choice as regards which courses to take. However, there are some compulsory courses. These include Finnish, Swedish, a foreign language (in most cases English) and basic mathematics.

Graduation from a high school requires passing a matriculation examination in addition to completing the required courses. Exams have to be passed in at least four different subjects. Of these, the Finnish (or Swedish for the Swedish-speaking minority) exam is compulsory. The other three can be chosen from a choice set determined by the courses completed during high school studies. The matriculation examination takes place in each spring and autumn. Traditionally, candidates would take their exams at the end of the spring term of their 3rd year, but the modern system is somewhat more flexible. Typically, students now sit their exams in two or more parts. Most still graduate after 3 years, but some graduate after only 2 years, where some take 4 years to graduate.

Success in high school is a major determinant of entry into tertiary education. In fact, an overwhelming majority of university entrants are high school graduates. Entry into university is based partly on matriculation exam grades and partly on entrance exams. Some university departments do not require participation in entrance exams for candidates

[2] Most of the rest enter vocational school or training.

with top grades in relevant matriculation exams. For those university curricula that do require entrance exams, the entry threshold in these exams in significantly lower for candidates with good matriculation exam results. This means that possible differences in high schools may translate into significant differences in entry into the most competitive university curricula.

2.2 Private high schools

In Finland most high schools are public and run by the municipalities, but there are also some private and government-run schools. Finnish private schools are funded by the government using the same principles as for public high schools, which are controlled and mainly financed by municipalities. Generally, private schools are given a state grant comparable to that given to a municipal school of the same size. It is important to note that by law in Finland private schools cannot make a profit and they do not usually charge tuition fees[3]. All the private schools we consider in our empirical analysis are not-for-profit schools.

While Finnish private schools cannot make a profit, there are some relevant differences between public and private schools. Maybe the most important difference is that private schools have more autonomy in arranging their activities and making decisions over resource allocation and personnel than public schools controlled by municipalities. In addition, private schools are more autonomous in defining their profile as long as they operate according to the national core curriculum. Related to this, many private high schools have an area of specialization (e.g. mathematics and science) in which they offer students more comprehensive instruction[4]. On the other hand, as students from both private and public schools have to take part in the same matriculation examinations at the end of high school, this might diminish possible effects of varying school profiles at least on test scores or on short-term outcomes.

[3] There are a few international schools that charge tuition fees. However, we do not consider these schools in our analysis.
[4] However, some public schools also offer specialized education.

Mainly for historical reasons, the relative share of private schools is much higher in Helsinki than in any other city or municipality in Finland. Since private schools are not common in other cities and the number of students attending private schools (around 10% on average) is generally low, we concentrate on comparing private (12 schools) and public high schools (14 schools) in Helsinki.

Both private and public high schools use the same admission system that we describe next.

2.3 High school admission system

In contrast to comprehensive schools, high schools in Finland are generally very selective. High schools usually use only one criterion in the school admission system, the grade point average (GPA) from comprehensive school. The GPA is based on grades in 10 different subjects[5] and grades are given on a scale from 4 to 10, which means that the GPA also lies between 4 and 10. In the school admission system, schools rank students according to their grade point average (GPA) from best to worst and students are admitted according to schools' rankings.

Generally, the joint application system to upper secondary schools is based on students' preferences and their GPAs and the selection mechanism can be seen as a deferred acceptance (DA) algorithm. Abdulkadiroğlu et al. (2014) describe the DA algorithm in a similar setting, while Virtanen (2016) has described the algorithm used in the Finnish joint application system. The basic idea is that students rank five or less track-school combinations according to their preferences[6]. The algorithm then matches schools' rankings and students' preferences in the following way:

Round 1: Each applicant applies to a track he/she has ranked first. Each track temporarily admits the highest-ranking applicants based on their GPAs until it is full. Other applicants are rejected. However, those temporarily admitted may also be rejected in later rounds.

[5] The subjects include mother tongue and literature, second national language, foreign languages, religion or life stance education, history and social studies, mathematics, physics, chemistry, biology and geography. The GPA is calculated using two decimal places.

[6] Instead of schools, students apply to track-school combinations, as some schools have several tracks (or tracks other than general tracks).

Round k > 1: Those rejected in round $k - 1$ apply to the track they ranked next-highest. These applicants and those who were temporarily admitted are ranked, and again, each track temporarily admits the highest-ranking applicants until it is full. The substitution list is also updated. Applicants who are temporarily admitted may be rejected again in later rounds. The algorithm continues until every applicant is matched to a track or every unmatched applicant has been rejected in every track he/she has applied to.

Applicants receive offers to secondary schools following the above algorithm. Students also receive information if they are on a substitution list for a school they have applied to. If a student has already accepted an offer but he/she qualifies for a higher-ranked school on his/her substitution list, he/she receives an offer for this school. This updating process continues until there are no slots available in any school or when there is no-one left on the substitution list. As an outcome of this process, the system determines for each school the GPA threshold of the last applicant to receive an offer (Virtanen 2016). However, there is also a replacement process that follows the same steps if for some reason there are places available after the updating procedure. Moreover, it is possible in some cases for students to obtain remaining places by directly contacting schools.

3. Data

The data we use is combined from three sources. The first source is the database of the Finnish Matriculation Examination Board, consisting of all matriculation exam results in Finland from 1990 to 2013. The second source is the National Board of Education database, which has data on all upper secondary school applications from 1997 to 2012. The records in the two databases were combined using the unique national identification number of each student. We also use the high school entrance threshold data of the Bureau of Education of the City of Helsinki, which is vital for the RDD analysis. The entrance threshold data contains the entrance thresholds for high schools located in Helsinki for the years 2000-2016.

The time period of the study, the years 2000-2012, was selected based on the temporal overlap of the data sources. The merging of the data was started by combining the matriculation examination data to the application register. After merging the entrance threshold data, the years of study were restricted to those students who applied to high schools between 2000 and 2012, since we only have information on the entrance thresholds starting from the year 2000.[7] In total, the final research data includes information on students from 12 private and 14 public high schools. There are observations on 34797 students from high schools in Helsinki, who applied to high school education between 2000 and 2012.

The matriculation exam data contains information on all the exams taken by each student / candidate, the time of each exam taken and on the grade or failure to pass. The database does not contain information on the date of matriculation or whether the exam was compulsory for the student taking it. There have also been some changes in the regulations during the period concerned. For these reasons it was impossible to say with complete certainty when each student graduated. Because of this, a matriculation date was constructed for each student from the data based on simple rules. The rules used were as follows. First, all students that had abandoned the matriculation process at some point and then restarted were removed from the database. Such students were rare and it would have been difficult to assign these students to a single school. Also, all students that had exam codes assigned to them that could not be resolved were removed.

There are examinations both in spring and autumn. The autumn examinations are different from those in spring, and most students choose to graduate in spring. For these reasons value-added analyses (presented later) were carried out on an annual instead of a semiannual level. Each student was assigned a date in spring as his/her graduation date. This date was chosen to be the first spring that the candidate had passed exams in at least four subjects. Passing was defined as either getting a passing grade in all four subjects or getting an overall passing grade by meeting the compensation rule. In our data compensation is achieved by receiving enough compensation points in three passing grades to compensate for a failed exam.

[7] In fact, most of the observations in the dataset are for students who applied to high schools during the time period 2000-2010, because graduation from high school usually takes three years. Only 1.4 per cent of the data is for students applying after the year 2010.

The grade average for each passing student was calculated to be the average grade of the exams taken up until and including the graduation date. The grades in the Finnish matriculation exam are actually Latin phrases, but the Matriculation Examination Board converts these into numbers using a scale in which the best grade is translated into 7, the second-best into 6, etc. The lowest passing grade is converted to 2 and fail is 0.[8]

In some of the regression models, we utilize control variables from the National Board of Education application data. For each applicant, this data set includes information on gender, mother tongue and the identification number of the comprehensive school from which the applicant graduated. We also use information on the grade point average as well as individual grades from the comprehensive school. Individual grades are not used in the main analyses, but are exploited in some validity checks.

Finally, it is important to note that a student may appear in the application data multiple times if he/she applied to secondary education multiple times. This poses no problems if the data for each time is consistent. For some students this was not the case. For those students, we used the latest information. After this selection, there were still some students with duplicate observations. We removed the duplicates, which had the same information on the application preferences. If this was not enough to produce a unique value for all the variables, the observation with the best comprehensive school GPA was chosen and after this the rows containing most information on all the variables.

4. Methods

4.1 Value-added estimation

We use two different econometric methods to compare the effectiveness of private and public schools, value-added estimation and regression discontinuity design. In this subsection, we briefly describe the value-added estimation method we utilize.

[8] For the RDD analysis, we also use the mother tongue grade and the sum of all matriculation examination grades in the data as alternative outcomes in the analysis.

The basic idea of the value-added model is to control for students' earlier or initial level of accomplishment when estimating school effects. In our context, this can be done by controlling for students' success in comprehensive school. The most natural control variable is the grade point average (GPA) in comprehensive school, since the high school admission system uses GPAs to select students (see Section 2.3).

The basic idea of value-added models is to use statistical techniques to control for the initial quality of student intake. While the literature includes various kinds of value-added models, here we follow the model suggested by Chetty et al. (2014), in which the method is described in more detail.[9] We use a notation, where each student *i* matriculating in year *t* is in some high school $j = j(i,t)$. His/her (average) score in the matriculation exam is $A^{*}_{i,t}$, which is determined according to the following equation:

$$A^{*}_{i,t} = \mu_{j,t} + \beta' X_{i,t} + \varepsilon_{i,t},$$

where $\mu_{j,t}$ is the effect (true value-added) of school *j* in year *t*, $X_{i,t}$ is a collection of covariates describing the characteristics of student *i* matriculating at time *t*, β is the coefficient vector associated with these covariates and $\varepsilon_{i,t}$ is the idiosyncratic error term of student *i*.

In our case, $X_{i,t}$ includes a third-degree polynomial in the comprehensive school grade point average, mother tongue, gender, year dummies and comprehensive school dummies. The comprehensive school dummies are included for two reasons. First, since grading is based on a somewhat subjective evaluation by teachers in comprehensive school, grade scales can differ among comprehensive schools. If this is true, the comprehensive school grade point average does not necessarily control for student quality well enough. The second reason to control for comprehensive school effects is an attempt to limit possible selection bias.

[9] Chetty et al. (2014) use the model to evaluate the value-added of teachers in primary schools, while we apply it to evaluate the effect of different types of high schools. A large number of papers have used value-added models to compare schools (see e.g. Raudenbush and Willms 1995, Reardon and Raudenbush 2009, Chamberlain 2013, Deming 2014).

The most interesting part of equation (1) is naturally the school effect $\mu_{j,t}$, which describes the impact of the high school on the student's exam results. This effect may be understood using a thought experiment. Suppose we could choose a number of students by random sampling and assign these students randomly to schools *A* and *B*. Then the difference in their matriculation exam results would be $\Delta_{A,B,t} = \mu_{\mathrm{A},t} - \mu_{\mathrm{B},t}$.[10] In the models we estimate, the effect of the average school is normalized to 0 in each year (i.e. $\bar{\mu}_t = 0$). This means that the value-added estimates are interpreted in comparison to the average value-added in a specific year.

One important extension to standard value-added models is that our model allows for the school effect to change over time. Note that we do not assume how rapidly such changes occur, but estimate them from the data. This allows for the effects of schools to decay rapidly or to be very persistent over time. However, what the model does not allow for is for school effects to increase or decrease limitlessly, or for the rate of change to vary over time (for a more detailed discussion about these assumptions, see Chetty et al. 2014). These are not strong requirements, at least when taking into account the time period we are looking at.

Estimation of the value-added measures is done in three steps. In the first step, the results of the matriculation exam $A^*_{i,t}$ are regressed on the covariates $X_{i,t}$ and high school fixed effects. This gives an estimate $\hat{\beta}$ for coefficients β. Further, using the estimated coefficient vector, we obtain the residualized exam score $A_{i,t} = A^*_{i,t} - \hat{\beta}' X_{i,t}$ for each student.[11]

In the second step, we then estimate the mean residualized score for each school-year combination by calculating $\bar{A}_{\mathrm{j},t} = \frac{1}{n_{j,t}} \sum_{i=1}^{N_{j,t}} A_{i,t}$, where $n_{j,t}$ is the number of matriculating students in school *j* in year *t* and the sum is over all *i* with high school $j = j(i,t)$. Finally, in

[10] For example, if the true effect of school *A* in year 2010 $\mu_{\mathrm{A},2010}$ is = 0.2 and the effect of school B is $\mu_{\mathrm{B},2010}$ = -0.5, this would mean that the expected exam results of the randomly assigned students would be 0.7 grade points higher in school *A* than in school *B* in year 2010.

[11] The residualized score is then the exam score purified of the effects of the covariates, or put another way, the exam score after controlling for the initial quality of the student.

the third step the value-added measure is constructed using the best linear prediction of the mean residualized score of a school conditional on *s* previous (or other) mean residualized scores, or $\hat{\mu}_{j,t} = \hat{E}\left(\bar{A}_{j,t} \middle| \bar{A}_{j,t-1}, \ldots, \bar{A}_{j,t-s}\right)$.[12] For more details of the estimation methodology and the assumptions behind the econometric model, see Chetty et al. (2014).

It is important to note that the value-added estimator is a consistent estimator of the "true" school effect $\mu_{j,t}$ only if the error terms are not correlated with school effects. However, this assumption is violated if students select into schools of different quality based on variables that are not captured by the model or the control variables. On the other hand, Deming (2014) and Chetty et al. (2014) have recently used randomized experiments and quasi-experiments to evaluate possible bias in typical value-added models. At least in these studies, selection bias was not an issue and standard value-added models worked surprisingly well in evaluating school and teacher effects.

4.2 Regression discontinuity design

As students can self-select to schools based on both observable (GPA) and unobservable variables (e.g. motivation), it is not clear how well the value-added model presented above can address the potential selection problem. Basically, the problem of selection bias could be solved by random assignment of students to schools. So we would ideally randomly assign students to private and public high schools. While this is not a possible exercise for us, we are able to utilize a quasi-experimental setting that mimics a randomized experiment.

Our quasi-experimental setting is based on regression discontinuity design (RDD) and high school entrance thresholds. Basically, because of the common application system to high schools (see Section 2.3), all the schools will have a school-specific entrance threshold. This threshold determines the minimum GPA that is required to get an offer from the school. So if a student's GPA is above the threshold for the (private) school he/she applied to, he/she will get an offer from that (private) school. Since entrance thresholds vary over years and

[12] This is a so-called jackknife estimator, which does not use information from the same period for prediction.

cannot be predicted exactly for a certain year, the Finnish system should provide a valid RDD (Virtanen 2016).

The idea in RDD is to compare students who were just accepted by a particular school to students who applied to the same school, but were just rejected. In other words, we want to compare students who were marginally over the GPA threshold of a specific school to students who were marginally below the threshold, but otherwise very similar. However, the design is more complicated in practice, because we have to take into account students' applications and preference rankings. As students can apply to five schools or less, they can basically appear above and/or below several different school-specific thresholds. To address these challenges, we exploit the approach developed by Abdulkadiroğlu et al. (2014) and construct so-called sharp samples for each individual school.

The sharp sample for an individual school *k* includes all the students who ranked *k* first in their application, those who did not qualify for their first-ranked preference and ranked school *k* second, those who did not qualify for their first two alternatives and ranked school *k* third etc. This continues until the student is accepted by one of his/her choices or when the student has been considered by every choice, but has not been accepted by any of them. Since the student can rank a maximum of five schools, he/she can appear in multiple sharp samples.[13] In addition, we pool individual school-specific samples into a private school (sharp) sample in order to study the general effect of private schools.

Although in the RDD setting crossing a specific threshold affect the treatment status, it is possible that there is imperfect compliance in obtaining treatment. This is also the case in our application if we consider private school enrollment as a treatment. Although private school-specific thresholds affect the probability of treatment, the probability of assignment to the treatment does not jump from 0 to 1, i.e. the design is fuzzy. This is because students can apply to several places and can in principle get offers from more than one school. Moreover, it is also possible that some applicants who are assigned to the control group (or are below the threshold) can receive the treatment.[14]

[13] To account for this structure of the sharp samples, in the statistical inference we cluster the standard errors at the individual level.
[14] For example, a school could use some other criteria besides or instead of GPA to select student(s).

To account for imperfect compliance, in our analysis we estimate both intention-to-treat effects (ITT) and local average treatment effects (LATE). ITT is obtained by estimating the effect of an offer on outcomes, in which case we do not take into account imperfect compliance. Instead, by estimating LATE we can take into account the fact that our design is not sharp, but fuzzy. To estimate LATE, we use two-stage least squares (2SLS) and use the offer as an instrumental variable for enrollment in private schools.

Before implementing the RDD analysis, we also have to make decisions on certain modelling choices. For example, one can estimate the treatment effect using either a parametric or nonparametric approach. Here we will use a more flexible nonparametric approach based on local linear regression. A local linear approach requires us to choose a bandwidth or window within the model is assumed to be linear. We use a mean square error (MSE) optimal selector as proposed by Calonico et al. (2017) to choose the bandwidth parameter(s). Besides bandwidth, we have to choose a kernel function for the local linear regression. As is quite typical in the RDD literature, we use a triangular kernel function, which has been shown to be optimal for local linear estimation at a boundary. However, in practice the choice of kernel function does not usually have much impact on the results.

Besides estimating the effect of private school enrollment using RDD, it is important to check that the design is valid and that the results are robust to modeling choices such as the choice of bandwidth. We follow standard RDD practice by implementing various kinds of validity and robustness checks. First, we investigate the density of the running variable, which should be continuous at the threshold. Second, as RDD assumes that predetermined variables should be continuous at the cutoff, we test this continuity assumption. Finally, we also check whether the results are robust or stay similar if we use a local polynomial approach instead of a local linear approach.

5. Results

In this section, we discuss the results from our estimations using the value-added model and RDD estimation described in Section 4. We first present the descriptive statistics and ordinary least squares regression results in Section 5.1 and then the results from the value-added analysis and the RDD estimation in Sections 5.2 and 5.3.

5.1 Descriptive results

The descriptive statistics of the data are presented in Table 1. The majority of the students who had completed exams in the matriculation examination are women. We can also see that students from private high schools have on average a slightly lower comprehensive school GPA and average score in the matriculation examination (MEA) than those from public schools. However, the matriculation examination grade in mother tongue and the sum of all grade averages (MESG) are higher for private school students than for public school students. Also, the standard deviations of matriculation examination scores are higher for those attending private schools, while for the GPA the standard deviations are slightly lower than for public high schools. However, the differences between school types remain small.

Table 1: Descpriptive statistics of the data

Variables	Obs.	Mean	Std.Dev.	Min	Max
Panel A: Entire data					
Female	34,797	0.585	0.493	0	1
GPA	34,797	8.519	0.762	5.08	10
Mother tongue	32,524	4.324	1.370	0	7
MEA	32,540	4.310	1.117	1.125	7
MESG	34,797	22.261	9.222	0	70
Panel B: Private schools					
Female	12,034	0.534	0.499	0	1
GPA	12,034	8.471	0.751	5.08	10
Mother tongue	11,346	4.437	1.418	0	7
MEA	11,303	4.278	1.131	1.125	7
MESG	12,034	22.638	9.367	0	70
Panel C: Public schools					
Female	22,763	0.612	0.487	0	1
GPA	22,763	8.544	0.766	5.38	10
Mother tongue	21,178	4.264	1.339	0	7
MEA	21,237	4.326	1.109	1.143	7
MESG	22,763	22.063	9.139	0	66

Variable Female is 1 if the person is a female. GPA is the comprehensive school grade point average. MEA is the calculated average of the student's matriculation examination test grades. MESG is the sum of all matriculation examination test grades the student has participated in.

Before going to our main results, we first present the results of simple OLS models. The OLS results in Table 2 are presented for the purpose of comparability. The OLS estimates show the association between enrollment in a private school and the following outcomes: matriculation examination grade in mother tongue, average matriculation examination grade (MEA) and the sum of all matriculation examination grades (MESG).

The estimate of the relationship between studying at a private high school and MEA is marginally negative in the first model without control variables. However, after controlling for the student's starting level, comprehensive school GPA, and gender, the MEA estimate turns positive, but is only fractionally above zero. The estimates for mother tongue and MESG are also positive, and after adding the controls they have a clear positive relationship to attending a private school.

Table 2: OLS results: private school enrollment

Outcome	(1)	(2)	(3)
Mother tongue	0.173***	0.253***	0.276***
	(0.016)	(0.013)	(0.013)
	32,524	32,524	32,524
MEA	-0.049***	0.027***	0.028***
	(0.013)	(0.009)	(0.009)
	32,540	32,540	32,540
MESG	0.575***	1.181***	1.185***
	(0.105)	(0.076)	(0.076)
	34,797	34,797	34,797
Controls			
GPA	No	Yes	Yes
Female	No	No	Yes

Robust standard errors in parentheses and N below them.

*** p<0.01, ** p<0.05, * p<0.1

Even though all the estimates are positive and highly significant (with control variables), it is important to emphasize that the OLS estimates are not causal. Therefore, we next apply more sophisticated methods, including value-added estimation and RDD to evaluate the impact of attending private school and compare the results between these methods.

5.2 Value-added estimation

We estimate value-added for each private and public school in Helsinki for the years 2002-2013. In these estimations, we use average matriculation examination grade (MEA) as an outcome variable. As described in Section 3.1, the value-added estimation is conducted using the method proposed by Chetty et al. (2014). In estimating the best linear prediction of the mean residualized score, the time drifts are limited to eight drifts. After obtaining value-estimates for all schools and years, we compare the mean and the distribution of the value-added estimates of different school types.[15]

Figure 1 plots a histogram of the value-added estimates, while Figure 2 portrays kernel density estimates for private and public schools separately. According to our results, there is no statistically significant difference in the mean value-added estimates between private and public schools. The mean estimate for private schools is just marginally higher than for public schools at 0.006 (p-value 0.800).

[15] We do not report regression coefficients for the value-added models, because we are not interested in how control variables are associated with matriculation exam scores. However, those results are available from the authors on request.

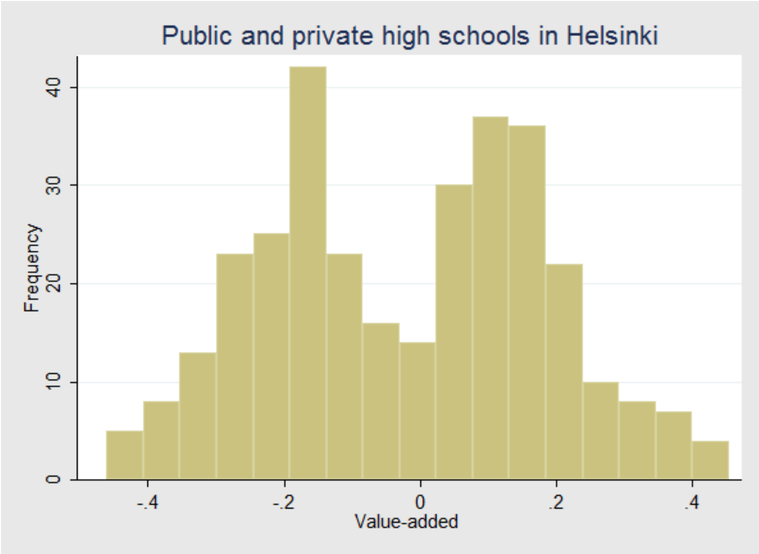

Figure 1: Distribution of value-added estimates for all schools.

Although there seems to be no difference in mean value-added between school types, the value-added distributions seem to differ between private and public schools (see Figure 2). When testing for the null hypothesis of identical distributions, we found that both the Kolmogorov-Smirnov test (p-value 0.010) and a more powerful Epps-Singleton test (0.000) reject the null hypothesis.

The deviation in the value-added estimates for private schools appears to be higher than for public schools, where they are more centered. This indicates some evidence of more homogenous quality in public schools compared to more deviated quality in private high schools, which might be caused by differences in organization and practices between the school types.

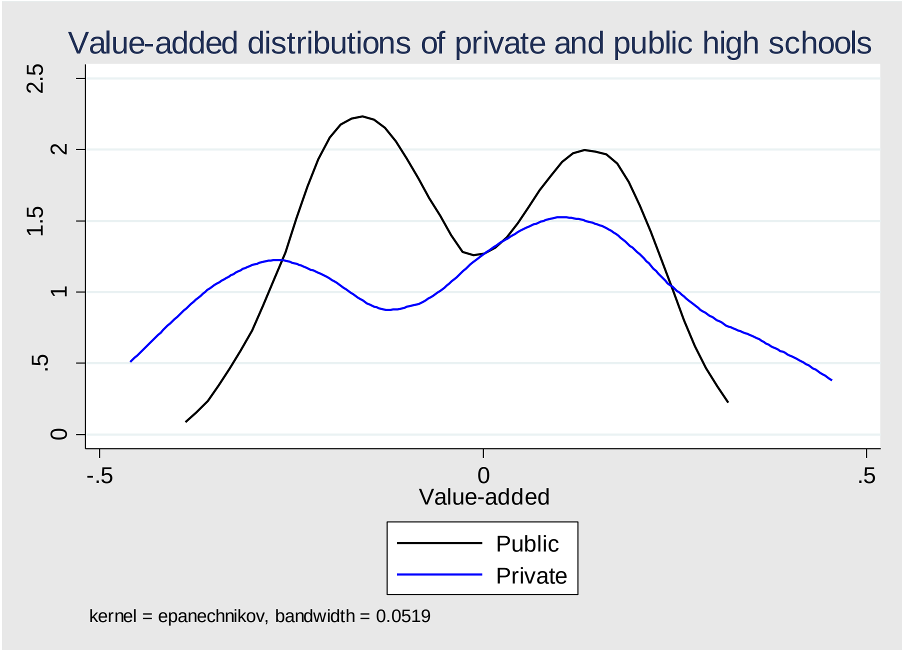

Figure 2: Kernel density estimates for value-added in private and public schools

5.3 Regression discontinuity design

We constructed sharp samples for each individual school using the approach of Abdulkadiroğlu et al. (2014) as described in Section 4.2.[16] To study the effects of private schools, the individual sharp samples are pooled. The final pooled sample has 12,723 observations. Because we have formed a pooled sharp sample to estimate the general private school effect, the counterfactual or control group include students who do not get an offer or study in a private high school, i.e. study in a public high school.

Note that the assignment variable is defined to be the student's distance from the school-year-specific threshold. The value of the assignment variable is the GPA minus the threshold value and it is multiplied by 100. A value of 10 means that the student is 0.1 grades above the threshold and -10 means that he/she is 0.1 grades below the threshold. All the observations which had a value of zero of the running variable were moved one unit

[16] Only the general tracks for all high schools are studied in the analysis, because the outcomes of students in specialized tracks might not be comparable to others and the threshold information from those tracks is difficult to connect to this analysis.

above or below zero based on the information about the school they were offered a place at. Since the student can rank the same school with different tracks in his/her application, we only consider the choice with the largest value of the assignment variable. Because students can appear in multiple sharp samples, we use clustering on an individual level in all of the RDD models.

Theoretically, in the application system the offers for a place in a private school are deterministic in the sense that crossing the entrance threshold to a private school indicates a change in the probability of receiving an offer from zero to one. In reality, the observed offers are not the same as the predicted offers as seen in Figure 3, which shows the observed offers at the cutoff point. To help with this issue, we can estimate a fuzzy RDD model where only those who have been observed to receive an offer are evaluated. This means that we would use the observed offer as the treatment variable in the fuzzy model. The reasons for the error can be measurement errors in the admission variables, schools not completely following the admission protocol or other measurement errors in the variables of the data (Virtanen 2016, Tervonen 2016).

In Figure 4 we can see that the effect of an offer on enrollment is not sharp, but only probabilistic. This can also be seen as the first stage of the RDD analysis.[17] The estimate for the first stage is highly significant and shows a 36 percentage point increase in probability (Table 3) of enrolling in a private high school after crossing the threshold compared to those who are just below the threshold. It should be noted that one factor influencing this result is that we do not observe the actual enrollment information, i.e. we do not have information on whether the student has enrolled in the school he/she has been selected for. Therefore we have to use the information about the high school from which the student has graduated in the matriculation examination data as a proxy for enrollment. This information is probably accurate enough, but it does not provide information about the track where the student has studied. Thus, we cannot ensure the student has studied in a general track, which might cause some bias in the first-stage estimates.

We next estimate the reduced form, i.e. the effect of an offer on the outcomes, including MEA, mother tongue grade and MESG. These estimates measure the overall effect of

[17] Later we run the fuzzy RDD with the treatment variable as the enrollment status of the applicant to account for this imperfect compliance problem.

crossing the entrance threshold for a private high school. This is called an intention-to-treat (ITT) effect. It estimates the effect of an intention to study in a private school on the matriculation examination outcomes. Table 4 presents the ITT estimates for all the outcomes and we can see that there are no significant discontinuities in any of the analyzed outcomes. Furthermore, the magnitude of the estimates is very small, but positive. The MESG estimate is the largest and slightly over 0.15 grades as expected, since on average the sum-of-grades variable by definition receives clearly higher values for every student. The effects are also plotted graphically in Figure 5. The results suggest that receiving an offer for a private school has no effect on matriculation examination grades.

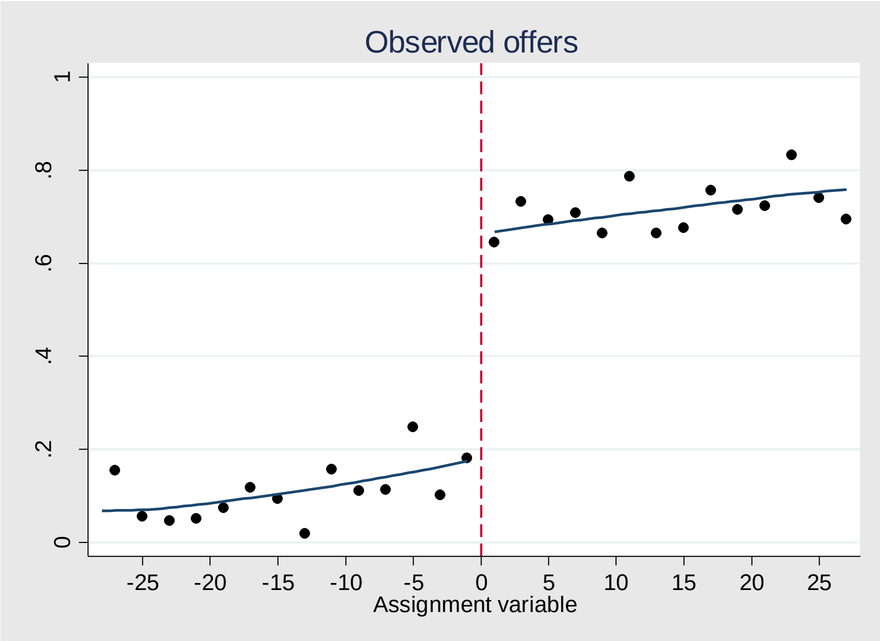

Figure 3: Observed offers

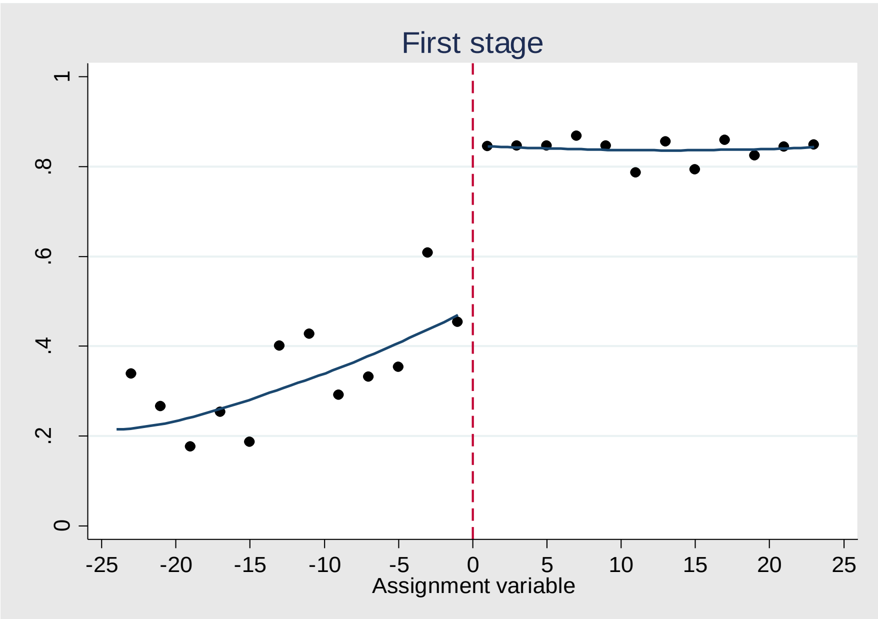

Figure 4: Enrollment in private school at the threshold

Table 3: First-stage estimate for sharp sample.

Outcome	
Enrollment in private school	0.361***
	(0.031)
	12,723

Standard errors in parentheses and N below them.
*** p<0.01, ** p<0.05, * p<0.1

After the reduced form, we consider two different fuzzy models, one with the observed offers and the other with enrollment in private schools as a treatment variable. In both cases we use predicted offers as instruments. The effect of enrollment on outcomes can be interpreted as the local average treatment effect for compliers, which is the group of students who enroll in a private school when offered a place in such an institution.

The 2SLS estimation produces similar results when compared to the reduced form. These estimates are presented in Table 5. All of the first-stage estimates are highly significant. For the fuzzy model with enrollment as the treatment variable, the estimates are of higher in magnitude and the MESG coefficient turns negative. In the observed offers model, the estimates for MEA and MESG are negative, but the mother tongue estimate is higher than the ITT result. However, all the estimates displayed in Table 5 are statistically insignificant.

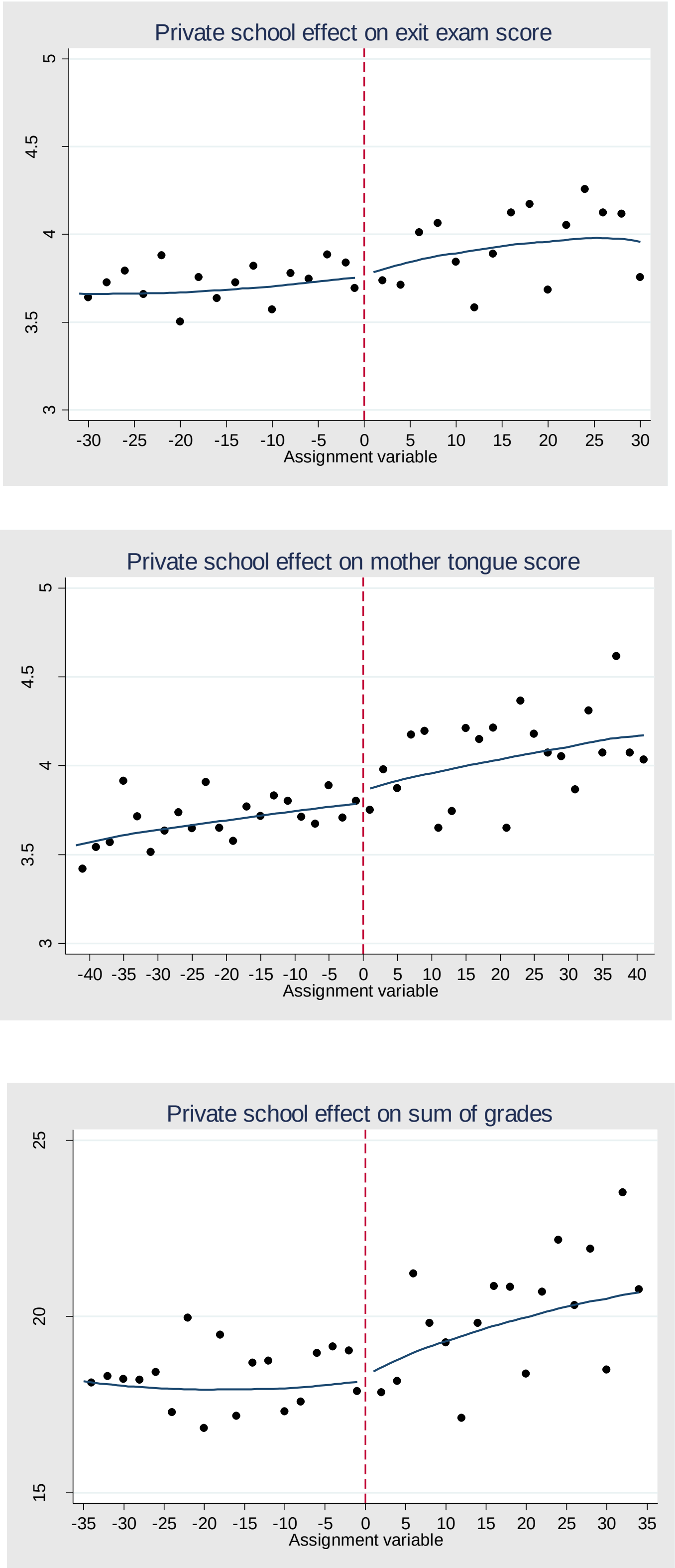

Figure 5: The effect of offer on outcomes (ITT).

Table 4: Reduced-form (ITT) estimates

Outcome	(1)
MEA	0.006
	(0.061)
	12,078
Mother tongue	0.074
	(0.071)
	12,019
MESG	0.157
	(0.386)
	12,723

Standard errors in parentheses and N below them
*** p<0.01, ** p<0.05, * p<0.1

Table 5: 2SLS estimates.

Instrument:	Enrollment	Offer
Outcome	(1)	(2)
MEA FS	0.378***	0.478***
	(0.028)	(0.027)
MEA 2SLS	0.029	-0.007
	(0.160)	(0.131)
	12,078	12,078
Mother tongue FS	0.372***	0.492***
	(0.028)	(0.026)
Mother tongue 2SLS	0.155	0.110
	(0.222)	(0.169)
	12,019	12,019
MESG FS	0.366***	0.487***
	(0.029)	(0.026)
MESG 2SLS	-0.247	-0.033
	(1.161)	(0.844)
	12,723	12,723

Standard errors in parentheses and N below them.
*** p<0.01, ** p<0.05, * p<0.1

The main assumption in the RDD model is the inability of individuals to precisely manipulate the assignment variable, in this case their GPA. If applicants could precisely set their GPA over the threshold value of the school they prefer, it could cause the RDD model to be invalid. This is implausible for two main reasons. Firstly, students cannot precisely choose their own GPA, although they undoubtedly do influence their own grades. There is always some degree of randomness in the school grades which form the GPA measure (Abdulkadiroğlu et al., 2014). Secondly, applicants do not know the entrance threshold for a specific school beforehand. They can use the thresholds from previous years to predict the threshold for the upcoming application year, but the thresholds rarely remain exactly the same from year to year. In fact, only 36 times out of the possible 336 times did the admission threshold remain the same as it was the previous year. Because of the randomness caused by these two factors, it is unlikely that one can precisely manipulate one's GPA in order to receive the treatment.

The continuity of the running variable can be tested using the McCrary test. However, in our case the discontinuities in the threshold can be caused by other factors. The comprehensive school GPA is calculated from grades for 10 theoretical subjects and possible additional foreign language course grades. Therefore, some of the GPAs are more likely than others, since some grade averages might be mathematically impossible to achieve for a student with for example 10 grades. This is highlighted by the reoccurring spikes in the histograms of the GPA distributions in Figures 6 and 7. The McCrary test rejected the null hypothesis of continuity in the sharp sample. However, this result could be a result of the reasons pointed out above.

We also test the continuity of other observable variables, which are determined before the treatment assignment. These variables include gender and the non-theoretical subject grades studied in comprehensive school. Table 6 presents the ITT estimates for these variables. The visual arts and the handicraft grades were not available in the data for the whole time period. Only the estimate for physical education is statistically significant, although the point estimate is not very large. According to these results, there are no significant differences in the observable characteristics of the applicants near the threshold.

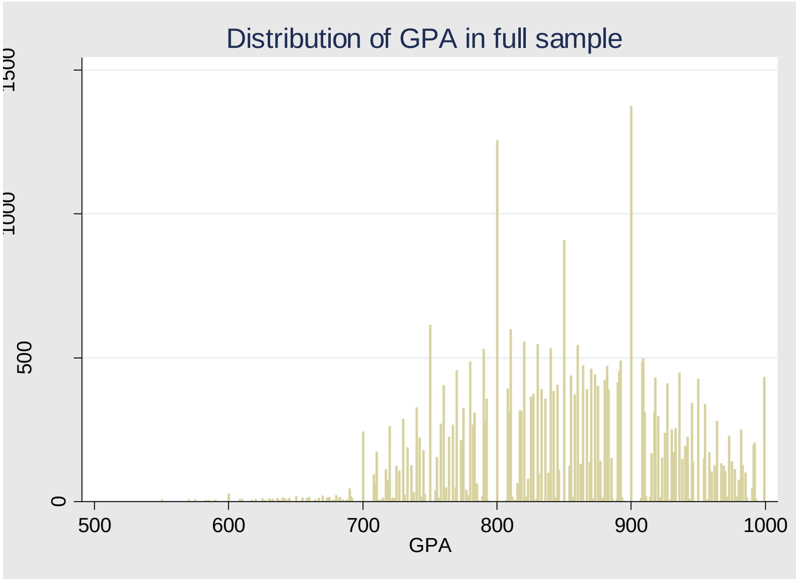

Figure 6: Distribution of GPA in the full sample.

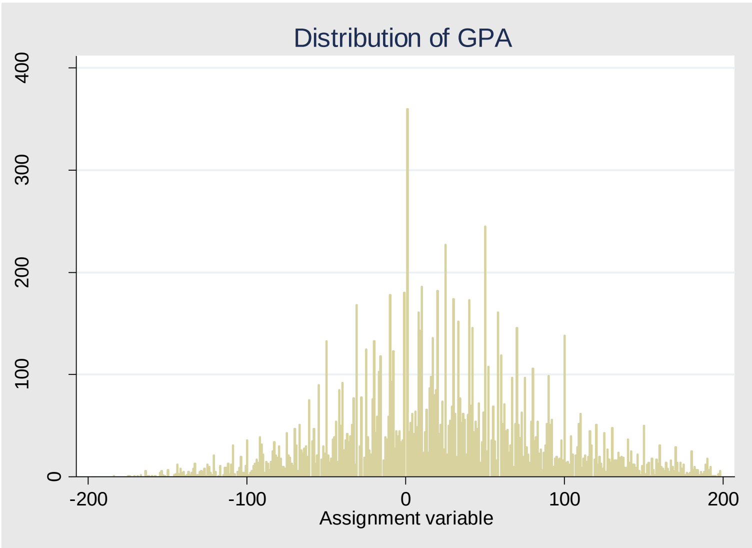

Figure 7: Histogram of GPA distribution in the sharp sample. Note that the observations which had a value zero of the assignment variable have been moved one unit right or left of the cutoff point based on the information for the school at which they were offered a place.

Table 6: Continuity of pre-determined covariates.

Outcome		Outcome	
Music	0.047 (0.046) 12,663	Handicraft	-0.067 (0.066) 7,091
Visual arts	0.062 (0.061) 7,114	Home economy	0.020 (0.040) 12,593
Physical education	0.119** (0.060) 12,651	Female	0.003 (0.026) 12,723

Standard errors in parentheses and N below them.
*** p<0.01, ** p<0.05, * p<0.1

The RDD analysis in this paper was conducted using local linear regression. As a robustness check, we also used quadratic polynomials in the local regression. The different polynomial order only had a minor effect on the ITT and LATE estimates and all of them were still clearly statistically insignificant.

To establish the sensitivity of the results to different bandwidth choices, we ran the local linear models using a series of different bandwidths ranging from 5 to 100. We did this with both the ITT and LATE specification using MEA as an outcome variable. These estimates are plotted in Figures 8 and 9. The lines in the graphs show the 95% confidence interval for each estimate. The optimal bandwidths are highlighted in red. The sensitivity checks confirm that the results are not dependent on the bandwidth choice since all the estimates are not significantly different from zero. There is some deviation in the estimates of the smaller bandwidths, but these estimates are very imprecise.

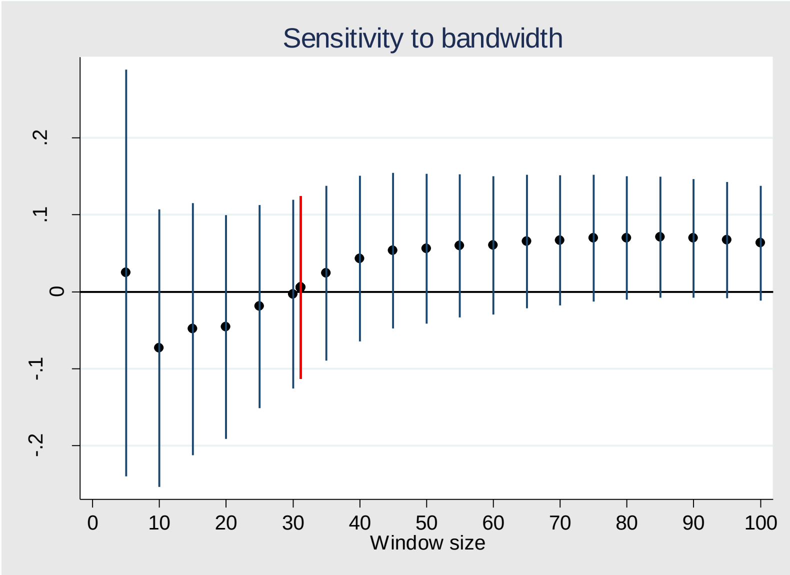

Figure 8: Sensitivity of ITT estimates of MEA to bandwidth size.

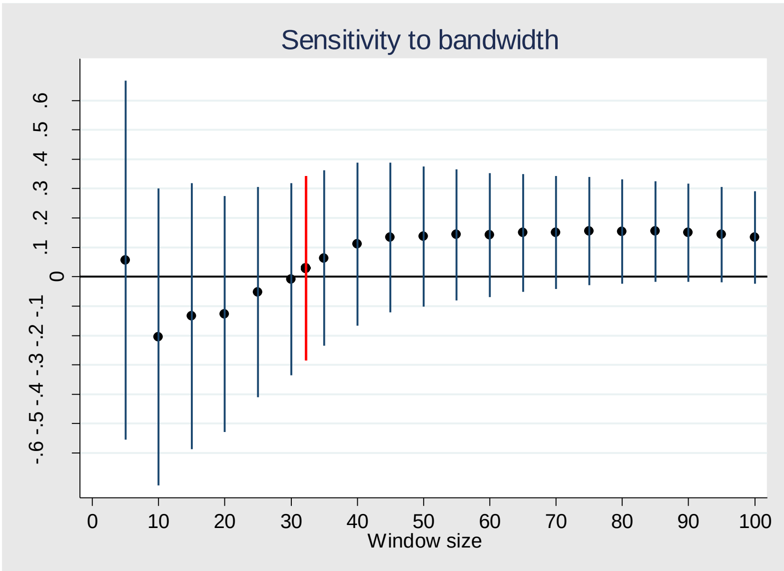

Figure 9: Sensitivity of LATE estimates of MEA to bandwidth size.

Finally, it could be argued that the number of matriculation exams taken by the student might vary between school types and this could influence the results. Therefore we tested the assumption that there is a discontinuity in the exam count at the threshold. The estimate for this was insignificant, thus indicating that the number of exams taken does not differ between school types and it has no impact on the results.

6. Conclusions

In this paper we investigated the differences in the effectiveness of private and public high schools in Finland. Although private high schools perform marginally better than public schools in terms of matriculation exam scores, the causal effects of private schools on exam results are very small and statistically insignificant. Importantly, these conclusions are consistent in the results obtained using two different econometric methods: value-added estimation and regression discontinuity design (RDD). Moreover, the various robustness and validity checks we implemented reinforce our RDD results and the validity of the discontinuity design.

While it is interesting that students' performance in high school does not depend on the type of school (private vs. public) in Finland, it should be noted that we have only looked at short-term effects on matriculation exam scores. It is possible, for example, that one can learn non-cognitive or other skills differently in private than in public schools. These possible effects might result in differences in long-term outcomes, even if we did not observe any short term effects on test scores. In future research we plan to study the effect of school choice on various long-term educational and labor market outcomes.

References

Abdulkadiroğlu, A., J. Angrist and P. Pathak (2014), "The elite illusion: Achievement effects at Boston and New York exam schools", *Econometrica* 82(1), 137-196.

Abdulkadiroğlu, A., P. Pathak and C.R. Walters (2018), "Free to choose: Can school choice reduce student achievement?", *American Economic Journal: Applied Economics* Vol 10 (1), 175-206.

Berends, M. and R. Joseph Waddington (2018), "School choice in Indianapolis: Effects of charter, magnet, private, and traditional public schools", *Education Finance and Policy*, forthcoming.

Calonico, S., M. D. Cattaneo, M.H. Farrell and R. Titiunik (2017), "rdrobust: Software for Regression Discontinuity Designs", *The Stata Journal* 17(2), 372-404.

Chamberlain, G.E. (2013), “Predictive effects of teacher and schools on test scores, college attendance and earnings”, *Proceedings of the National Academy of Sciences* 110(43): 117176-171782.

Chetty, R., J.N. Friedman and J.E. Rockoff (2014), “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates”, *American Economic Review*, *104*(9), 2593-2632.

Deming, D. (2014), “Using school choice lotteries to test measures of school effectiveness”, *American Economic Review: Papers & Proceedings* 104(5): 406–411.

Hahn, Y., L.C. Wang and H. Yang (2018), “Do greater school autonomy and accountability make a difference? Evidence from the random assignment of students into private and public high schools in Seoul”, *Journal of Public Economics* 118: 15-30.

Häkkinen, I., T. Kirjavainen and R. Uusitalo (2003), “School resources and student achievement revisited: new evidence from panel data”, *Economics of Education Review* 22 (3), 329-335.

Kanninen, O. 2013. Five essays on economics of education. PhD Thesis. European University Institute.

Kirjavainen, T. and H.A. Loikkanen (1998), “Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis”, *Economics of Education Review* 17 (4), 377-394.

Kortelainen, M., Pursiainen, H. and J. Pääkkönen (2014) ”Lukioiden väliset erot ja paremmuusjärjestys” (In Finnish), VATT Tutkimukset 179.

Reardon, S.F. and S. W. Raudenbush (2009), “Assumptions of value-added models for estimating school effects”, *Education Finance and Policy* 4 (4): 492–519.

Raudenbush, S. W. and J.D. Willms (1995), “The estimation of school effects”, *Journal of Educational and Behavioral Statistics* 20 (4): 307–35.

Tervonen, L. (2016), *Does attending an elite high school have an effect on learning outcomes? Evidence from the Helsinki Capital Region*, University of Helsinki, Master's Thesis.

Virtanen, Hanna (2016), *Essays on post-compulsory education attainment in Finland*, Aalto University publication series, Doctoral Dissertations 87/2016.