School choice in Amsterdam: Which schools are chosen when school choice is free?*

Nienke Ruijs Hessel Oosterbeek

Abstract

Advocates of increased school choice and competition assume that school choices are led by quality considerations. To test this assumption, this paper examines the determinants of secondary school choice in the city of Amsterdam. In this city there are many schools to choose from and school choice is virtually unrestricted (no catchment areas, low or no tuition fees, short distances). Using discrete choice models we find that published indicators of school quality are not consistent predictors of school choice. Instead, students appear to prefer schools that are close to their home and schools where many of their former classmates in primary school go to.

1 Introduction

This paper analyses the determinants of secondary school choice in the city of Amsterdam. Students in this city can choose any school that offers education at their academic level, tuition fees are low, and schools are easy to reach due to relatively short distances. These features create an ideal environment to examine whether school choices are affected by published information about school quality. We do so by estimating discrete choice models using data of individual school choices for the years 2007-2010. Analyzing individual school choices (instead of aggregate school enrollment data) allows us to compare the effect of published information about school quality on school choices with the effect of other factors, including traveling distance, whether the school was oversubscribed in the previous year, and the number of classmates from primary school that attends the secondary school.

Analyzing the effect of published school quality information on school choice is important from the perspective of the debate on school choice and competition. The main arguments in favor of these policies are that by having more choice students can find a school that better matches their specific needs and that exposure to competition gives schools an incentive to improve education quality.[1] The direct evidence of the impact of

*This version: October 2014. We gratefully acknowledge Herman Ozinga and Kees Waijenberg from the Amsterdam municipal office for supplying the data. We thank Adam Booij, Monique de Haan, Inge de Wolf, Pierre Koning, Geert Ridder, Jesse Rothstein, Adriaan Soetevent and Karen van der Wiel as well as seminar participants at various places for useful comments and suggestions. The authors are affiliated with the University of Amsterdam and TIER. E-mails: n.m.ruijs@uva.nl; h.oosterbeek@uva.nl.

[1] We will often phrase as if students are the decision-makers. We acknowledge that in reality their parents have an important say in this.

school competition on school quality and student achievement is mixed (Böhlmark and Lindahl, 2013; Cullen et al., 2006; Hoxby, 2000a, 2007; Hsieh and Urquiola, 2006; Rothstein, 2007).[2] A prerequisite for such an impact to emerge, is that school choices are at least partially driven by school quality. If this is not the case, improvement of school quality does not raise student inflow and is therefore not a dimension on which schools should compete.

In a recent study examining whether published information on school quality affects secondary school enrollment in the Netherlands, Koning and Van der Wiel (2013) found that higher quality scores are associated with a small but significant increase in enrollment. There are three reasons to reexamine the findings of that study. First, Koning and Van der Wiel analyze data from the whole country thereby also including more remote areas where most schools are de facto local monopolists. By analyzing data from a city where the number of schools at any academic level is abundant, we focus on a situation where students have more possibilities to respond to published quality information.

Second, due to data limitations Koning and Van der Wiel use enrollment in the third year of secondary school as a proxy for applications in the first year. Since on average 3.9% of Dutch secondary school students repeat a grade in a given year in the first three years of secondary school (CBS, 2012) , this implies that in the third year, 7.8% of the enrollees have been replaced by 11.7% repeaters from previous cohorts. These repeaters from previous cohorts have based their school choices on published information on school quality from earlier years. In addition, a non-negligible fraction of 23% of secondary school students are transferred to a higher or lower academic track by the time they get into the third grade or have moved to another school (Onderwijsinspectie, 2007). Moreover, since the most popular schools are oversubscribed and cannot admit all applicants, school choice in response to published information on school quality is best measured by applications instead of enrollment. Our data on school choice in the city of Amsterdam contains information on first year applications. Third, the analysis of Koning and Van der Wiel does not include students in the lowest levels of pre-vocational education, and thereby omits 25% of Dutch secondary school students. We extend the analysis to all academic levels.

In another recent study using Dutch data, Borghans et al. (2014) inquire the determinants of school choice of the parents of primary school students in the southern part of the southern province Limburg. They find that parents tend to avoid the about 5% of the schools that the Dutch Education Inspectorate assesses to be weak. Schools that have higher average scores on the national exit exam from primary school attract more

[2] In a famous study Hoxby (2000a) finds that more school competition (measured as the number of school districts in a metropolitan area) boosts student achievement in the US. Identification is based on variation in the number of school districts caused by the numbers of streams (rivers). Rothstein (2007) has shown that Hoxby's results are sensitive to the way in which streams are counted (see also Hoxby, 2007). Hsieh and Urquiola (2006) find no evidence that choice, triggered by the provision of vouchers, improved student achievement in Chile. Using a comparable reform in Sweden, Böhlmark and Lindahl (2013) find, however, a positive effect of choice on student achievement. Using data from admission lotteries from the Chicago public school system, Cullen et al. (2006) find that students who lost the lottery – and therefore have a restricted choice set – are not harmed in their achievement.

students.[3]

Related to our paper is also the work of Hastings et al. (2009) who use data from primary and middle school choices from a school district in the US. In that district, a school choice program was introduced where parents could list their top three schools. Using exploded mixed logit models, they find that the weight parents place on school characteristics is heterogeneous. Parents, especially those of high SES, tend to prefer schools with high test scores. Parents also tend to prefer schools in which the majority of the students has the same race as they have. These results imply that minority parents face a trade-off between high performance and ethnic composition. Further, the distance to schools and the availability of transportation are relevant determinants of school choice. While Hastings et al. focus on test scores as indicator of school quality, we look at the impact of published school quality information. Test scores are arguably a measure of gross school quality including the quality of the student inflow, the published school quality information that we analyze, is intended to be a measure of a school's value added (Subsection 3.2 provides details).

To match students to schools, the city of Amsterdam uses an algorithm that is quite similar to the so-called Boston mechanism. If the school to which a student applies, is oversubscribed and the student is not accepted at that school (loses the lottery), he can subsequently only choose from the schools that then still have vacant seats. This matching algorithm gives rise to strategic behavior where students do not report their truly preferred school (Abdulkadiroğlu and Sönmez, 2003; Calsamiglia and Guell, 2014). We examine strategic choices by including a variable capturing whether the school was oversubscribed in the previous year.

In addition to measures of school quality and oversubscription in the previous year, we also inquire the impact on school choice of the number of classmates from primary school that attends a secondary school. Peer effects are intrinsically difficult to identify. De Giorgi et al. (2010) exploit that peer groups do not fully overlap to identify peer effects in the choice of college majors. Lacking such a source of variation and in the absence of random assignment of peers as in Sacerdote (2001), we restrict ourselves to testing the null-hypothesis of no peer effects in school choice.[4]

The main results of this paper are threefold. Unlike Koning and Van der Wiel (2013) we find that school quality and changes in school quality are inconsistent predictors of school choice in Amsterdam. Second, we find some evidence that when a school conducted an admission lottery in the previous year, students are deterred from choosing that school, suggesting that strategic behavior occurs. Third, our results reject the null-hypothesis of no peer effects in school choice. When a larger share of a student's primary school peers chooses for a certain secondary school, the student is more likely to pick that school as well. This is even true when we correct for the general popularity of the specific secondary school among the students in the student's primary school. Although this result does not

[3] Note that the published quality measures and the assessment of the Inspectorate attempt to measure schools' value added. The nationwide test is instead uncorrected for the quality of the inflow.

[4] Angrist (2013) gives a critical assessment of the peer effects literature.

prove the importance of peer effects in school choices, it is consistent with it.

This paper proceeds as follows. The next section describes the context of secondary school choice in Amsterdam. Section 3 describes the data. Section 4 provides details of the empirical strategy that we employ. Section 5 presents and discusses the empirical findings. Section 6 summarizes and concludes.

2 Context

2.1 School choice in the Netherlands

The Netherlands has a long history of free school choice. The constitution of 1848 guarantees the freedom to provide education. In 1917, this freedom was extended with the amendment that all schools receive state funding (Eurydice, 2009). In current practice, this means that privately-run schools (either with a religious background or subscribing to specific pedagogical approaches such as Montessori and Dalton) are publicly funded at the same level as publicly-run schools. In return, schools have to adhere to certain rules. In particular, they are subject to quality inspections by the Dutch Education Inspectorate.

For students, these regulations imply that they are free to choose the school they want; they are not restricted by measures such as catchment areas.[5] Dutch students make extensive use of this option: 70% of the students in primary education and 75% of the students in secondary education is enrolled in a publicly-funded privately-run school (CBS, 2009). Privately funded schools are virtually non-existent: in 2009, only 0.3% of the students in secondary education attended a privately funded school (Onderwijsinspectie, 2010).

The government funding of schools is to a large extent dependent on student numbers, in which the money follows the student. For disadvantaged students there are additional funding schemes. In primary education there is the system of weighted student funding. For disadvantaged students in terms of low parental education, schools can get additional funding up to 1.2 times the regular per student funding. In the paper, we will use the the weighted student funding to identify disadvantaged students.

2.2 Secondary education in the Netherlands

Dutch secondary education starts around age 12 and lasts four to six years. The length of secondary education depends on the school track: the Netherlands has a tracked secondary school system. The lowest tracks (pre-vocational secondary education, vmbo) last four years, and give access to vocational education programs. Within the pre-vocational track, there are four different levels, each giving access to different levels of vocational education programs. In this paper, they will be indicated with the numbers I to IV, with IV being the

[5] A small number of municipalities have put restrictions on primary school choice to foster desegregation (Ladd et al., 2011). In those projects, preferences are important in placement decisions as well.

highest level.[6] The intermediate track (senior general secondary education, havo) takes five years, and gives access to professional colleges. The highest track (pre-university education, vwo) takes six years, and gives access to university education (Eurydice, 2009).

Which school track a student should take is mainly decided at the end of primary education. It is partly determined by standardized tests (in most cases the nationwide exit test called the "citotoets"), and partly determined by the assessment of the primary school teacher. Not all secondary schools offer all school tracks, so the track advice is an important factor in secondary school choice. When offering more than one school track, schools are allowed to teach children of different tracks together. Depending on student achievement and school policies, students can change track during secondary education. Also, they can enroll in a higher track after finishing a lower track.

Subject to some conditions, students can choose which courses they want to take during their last years in secondary school. Secondary schools have to follow national curriculum guidelines. Students take centrally determined national exams at the end of secondary school. The national exams count for 50% of students' final grades, the other 50% is determined by school specific exams taken in the last two or three years of secondary education.

2.3 Secondary school choice in Amsterdam

Amsterdam is the capital of the Netherlands and is with 750,000 inhabitants its largest city. Each year, 5,500 to 6,000 students transfer from primary education to secondary education. In the city of Amsterdam, there are about 54 secondary schools, excluding schools for students with special educational needs. Not surprisingly, some schools are more popular than others. Each year some schools are oversubscribed and conduct lotteries to allocate the available places.

Starting in 2005, the secondary schools in Amsterdam run a centralized application and admission system. In the first round students can only apply to one school that offers the track of their academic level. Students that perform on the standardized test in line with the primary school teachers' advice cannot be rejected by the secondary school, which effectively means that it is hard for schools in Amsterdam to select students.[7]

When a school is oversubscribed, a lottery is conducted by the school itself. Some schools have oversubscription for some school tracks, but not for others. In that case, the lottery is conducted for each school track separately. Schools are allowed to use a limited number of priority rules: they can grant priority to siblings of current students, children

[6] Their Dutch names are bbl, kbl, gl and tl. Pre-vocational III (gl) will not be taken into account in the analyses, since only 2% of the students receive a pre-vocational III advice and only 10 schools offer this school track.

[7] The definitions for corresponding test scores are strictly prescribed. When a student has a lower score than can be expected from the teacher's advice, the secondary school should discuss the student with the primary school and/or conduct an extra standardized test. In these cases, the secondary school has some discretion in rejecting the student, which happens in about 5% of the cases. For most school tracks, the majority of the students (52%) has a test score in line with the primary school advice. An exception to this are the vocational levels with additional support. Here, all students are placed after discussing them with the primary school.

of staff members and to students from a primary school with similar special programs (for example, Montessori secondary schools can grant priority to children from a Montessori primary school). These priority rules need to be announced before the application date, so they are known to parents. When losing a lottery, students have to apply to one of the schools that still has places available after the first round. In the years that we study, around 5% of the students could not be placed on the school where they applied for in the first round because they lost the lottery. Note that the school where a student applies in the first round is not necessarily the actual first preference: there may be strategic behavior in choosing where to apply.

3 Data

3.1 Student information

Data come from the centralized application and placement system of the city of Amsterdam. This database has information on 21,117 Amsterdam students choosing a secondary school in Amsterdam in the four years from 2007 to 2010. The database provides information on student background characteristics, such as sex and ethnicity (but not income), and on primary school achievement, such as school track advice, score on the standardized test ("citotoets") at the end of primary school, and grade repetition in primary school.

For each student, we know at which school s/he applied in the first round and whether the student was enrolled at that school. For students who are not placed at the school where they apply in the first round, subsequent choices are also registered. This information is, however, not used here because the choice sets after the first round are not sufficiently clear.

Using information on students' school track advice and information on the school tracks offered by each secondary school, we can create the choice set for each student. Although schools outside of Amsterdam can also be chosen, we limit the choice set to schools in the city of Amsterdam. Schools outside of Amsterdam do not follow the same enrollment rules and their students are not registered in the Amsterdam enrollment system. Similarly, children outside of Amsterdam can choose schools in Amsterdam. Because their data is not consistently registered, these children are omitted from the analysis as well. Because the enrollment procedures are different for special educational needs schools, we drop those schools from the sample. Moreover, we drop some students who have missing values on key variables.[8]

[8]Overall, 7,432 students are dropped out of the initial sample of 28,549. 3,913 of them are dropped because they are living outside of Amsterdam or going to primary schools outside of Amsterdam. 1,488 students are dropped because they are going to special education needs schools. 103 students are dropped because their primary school is unknown, and 236 students are dropped because their address is not registered. Finally, 1,539 students are dropped because they choose schools that should not be in their choice set given their primary school advice and 153 students are dropped because they go to a few very small secondary schools, which do not have enough observations to take into account in this study. These are mainly religious schools, such as Jewish or Islamic schools.

To study the relation between a student's own school choice and that of his/her classmates, we calculated for each student the shares of classmates in primary school that apply to each of the secondary schools.[9] Because some secondary schools may always be more popular amongst students from a specific primary school (because of distance or pedagogy), we also constructed variables that express classmates' choices for secondary schools in deviation from primary-school-specific trends. This is akin to the population variation used by for example Hoxby (2000b).

Table 1 reports descriptive statistics on student characteristics. The first column shows that the pre-university track attracts around 21% of the students. Another 28% of the students enter secondary school at the combined senior general secondary/pre-university track or the senior general secondary track. The remaining 51% of the students start secondary education at the pre-vocational tracks. The second column shows that the share of boys is fairly constant across the different secondary school tracks. They are only somewhat underrepresented at the two lowest tracks. Column (3) demonstrates the segregation of secondary school tracks in Amsterdam along the lines of migrant status. While the share of students with both parents born in the Netherlands in the population is only around one third, their share in the highest track is 0.6 and this decreases monotonically to 0.14 in the lowest track. This carries over to column (5) which shows the share of students without a disadvantaged background by school track. Column (7) shows the monotonic relation between students' score on the exit exam in primary school and their track in secondary school. (The scale for this variable runs from 500 to 550.)

3.2 School information

Table 2 reports descriptive information on the choice set and the chosen school for the different school tracks. The first column reports the numbers of schools at each track from which students can choose. Many schools appear multiple times in this column because they offer more than one secondary school track.

Using information on students' home addresses, we calculated the distances from their house to Amsterdam schools in their choice set. These distances are calculated in a straight line based on coordinates.[10] Columns (2) to (4) provide means and standard deviations of the distances that students have to travel to the school they chose (column (2)), to the school offering their advised track nearest to where they live (column (3)) and the average over all schools at their advised track (column (4)). The mean distance to the nearest school is around 1 kilometer and this is very similar for the different school tracks. The mean distance to the school that was actually chosen is around 3 kilometers at all school tracks, whereas the mean of the average distance to all schools at a given track is 5 to 6 kilometers, again with little variation across school tracks.

[9] Since we do not have information on classes within primary schools, we define classmates as all other students in the final grade of a students' primary school.

[10] For a random sample of 100 students, we calculated the road distance as well. The road distance turned out to be very closely related to the distance in a straight line.

Table 1: Student Characteristics

	N	Share	Both parents born in the Netherlands		Belonging to disadvantaged group		Citoscore		
	students	of boys	Yes	Missing	Yes	Missing	Mean	SD	Missing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-uni.	4,374	0.50	0.60	0.04	0.14	0.12	544.1	7.4	0.01
Senior gen./pre-uni.	2,819	0.49	0.45	0.05	0.27	0.11	540.5	6.1	0.01
Senior gen.	3,133	0.50	0.39	0.05	0.34	0.12	537.7	6.4	0.02
Pre-voc. IV/senior gen.	2,028	0.49	0.28	0.06	0.47	0.10	534.4	6.2	0.02
Pre-voc. IV	2,881	0.50	0.22	0.05	0.54	0.10	531.5	6.7	0.03
Pre-voc. II	2,130	0.49	0.16	0.06	0.67	0.09	526.7	8.1	0.19
Pre-voc. I/II	1,039	0.48	0.14	0.05	0.65	0.15	524.7	9.0	0.33
Pre-voc. I	2,713	0.47	0.14	0.07	0.64	0.19	520.4	10.6	0.58
Total	21,117	0.49	0.34	0.05	0.42	0.12	535.7	10.1	0.12

Note: Columns (2) to (6) report shares. The 'No' category is omitted.

Table 2: Descriptive information on students' choice sets

		Distance (in kilometers)			Share of	Actual share of peers		Predicted share of peers	
	N	chosen school	nearest school	all schools	schools	chosen school	all schools	chosen school	all schools
	Schools	M (SD)	M (SD)	M (SD)	with lotteries	M (SD)	M (SD)	M (SD)	M (SD)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-uni.	27	3.09	1.04	5.00	0.08	0.12	0.03	0.11	0.03
		(2.13)	(0.71)	(2.90)		(0.12)	(0.06)	(0.11)	(0.06)
Senior gen./pre-uni.	21	3.17	1.20	5.60	0.10	0.14	0.03	0.13	0.03
		(2.15)	(0.78)	(3.08)		(0.14)	(0.07)	(0.12)	(0.06)
Senior gen.	23	3.08	1.14	5.57	0.10	0.14	0.03	0.14	0.03
		(2.16)	(0.73)	(3.04)		(0.14)	(0.07)	(0.13)	(0.06)
Pre-voc. IV/senior gen.	18	3.01	1.24	6.03	0.11	0.17	0.03	0.16	0.03
		(2.10)	(0.71)	(3.28)		(0.16)	(0.07)	(0.14)	(0.07)
Pre-voc. IV	23	2.85	1.14	6.03	0.16	0.18	0.03	0.17	0.03
		(2.08)	(0.72)	(3.20)		(0.17)	(0.07)	(0.15)	(0.07)
Pre-voc. II	22	2.98	0.96	6.28	0.05	0.13	0.02	0.12	0.02
		(2.29)	(0.59)	(3.40)		(0.15)	(0.06)	(0.12)	(0.06)
Pre-voc. I/II	21	2.85	1.00	5.82	0.06	0.12	0.02	0.11	0.02
		(2.23)	(0.66)	(3.16)		(0.15)	(0.06)	(0.12)	(0.05)
Pre-voc. I	22	2.82	1.05	6.00	0.06	0.12	0.02	0.11	0.02
		(2.12)	(0.69)	(3.15)		(0.13)	(0.07)	(0.11)	(0.06)

Note: Column (3) reports the distances to the nearest school within a students' choice set, column (4) reports the average distance to all schools within a students' choice set.

We constructed an indicator that equals one for tracks in schools that had a lottery in the year prior to the year of application.[11] Column (5) of Table 2 shows the share of schools that conducted a lottery, averaged over the four years of observation.

Finally columns (6) to (9) provide information about the share of classmates in primary school that go to the same secondary school (actual and predicted) and the shares of classmates that would go to the same school if they choose their secondary schools randomly. These shares show that classmates from the same primary school tend to choose the same secondary school (the actual share of peers in chosen schools exceeds the actual share of peers in all schools) even when corrected for the usual popularity of secondary schools among students from a given primary school (the actual share of peers in chosen schools exceeds the predicted share of peers in chosen schools).

As described in Section 2, Dutch schools are subject to quality inspections by the Dutch Education Inspectorate. Their quality information is based on several aspects, such as exam results and school visits. Since the 1990's, the inspectorate's quality assessment of schools is public information, published on a yearly basis. A national newspaper called *Trouw* makes the quantitative part of the information accessible to the broader public, by publishing exam grades and other characteristics, and by computing an overall quality score measured on a 5-point scale ("- -", "-", "0", "+", "++").[12] The inputs for the newspaper's quality scores are: final exam grades in the school track, the percentage of students getting a degree without grade repetition in the higher classes and the performance of the school in the lower years. In this last number, changes in school track play a role: students enrolled in a lower track than their initial advice reduce the score, students enrolled in a higher track increase it. The overall quality score is calculated and published for each school track separately, so within a school, there may be different quality scores. Very small schools or schools that miss values on the underlying scores do not get an overall quality score.

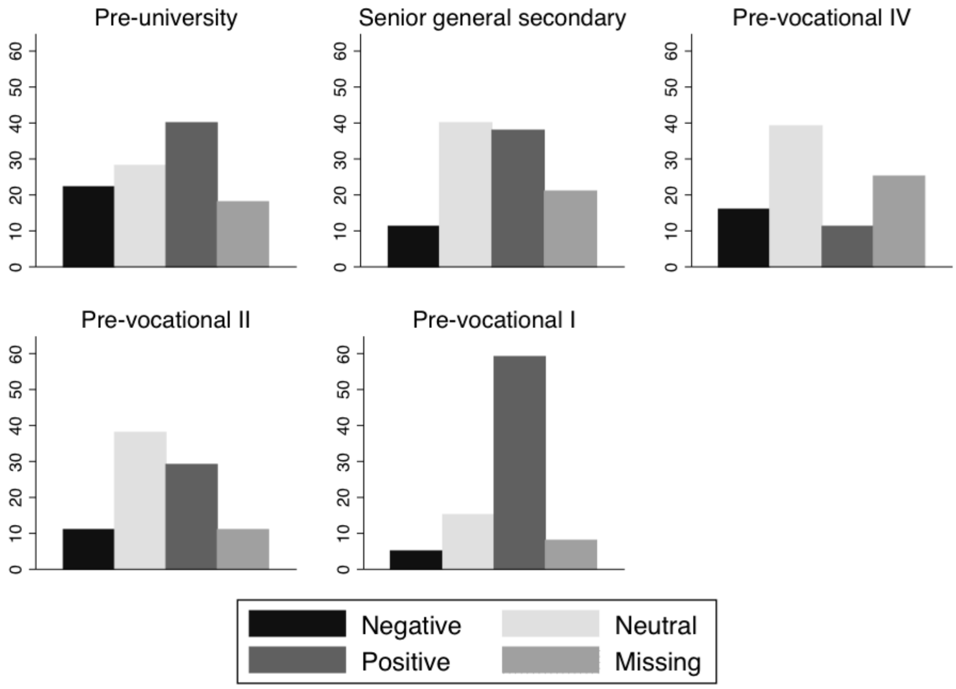
The quality information we use for each student is the quality information published in December before the student's school choice in April. In 2009 - the last publication relevant in our data - the 5-point scale was abolished and only the underlying numbers were published. Since the 5-point scale was computed in the same way every year, we are able to reconstruct the scores. In the analyses we will study the 2009 quality scores separately.

Since only Amsterdam schools are included in this study, there are not enough schools with the extreme quality scores "- -" and "++" in the sample. Therefore, we transformed the five-point scale into a three-point scale. As can be seen in Figure 1, there is substantial

[11]The information we use for the lottery indicator is from a booklet published annually by the municipality of Amsterdam. This booklet includes one page of information on each secondary school, together with information on the general enrollment procedure. It is handed out to all students in the last year of Amsterdam primary schools. From our contacts with schools, we noticed that the booklets have a few errors on the lotteries. Since students will visit the school before subscribing, and they will get the correct information there, we decided to adjust these cases to match the information from the schools.

[12]In 2001, a news magazine called "Elsevier" also started publishing school quality information based on the inspectorate information. We choose to use the Trouw data since they were the first to publish this information and since it has a wider circulation than Elsevier.

Figure 1: Variation in Trouw Quality Scores



Note: This figure shows the occurrence of different Trouw quality scores published in December 2006 to December 2009 for Amsterdam schools.

variation in the quality scores of the Amsterdam schools.[13] The quality scores are not stable over time. Table 3 shows the variation of the quality scores within schools in the four-year period of the study. For example, only 10 out of the 27 schools that offer the pre-university track keep the same quality score for four consecutive years.

4 Empirical strategy

4.1 Main analysis

To study the determinants of secondary school choice we use discrete choice models. Specifically we will estimate conditional and mixed logit models. These models follow from a random utility framework in which a student is assumed to choose the school that maximizes her utility. The utility that student i choosing in year t derives from secondary school s is defined as $U_{its} = V_{its} + \varepsilon_{its}$, where V_{its} is the observed portion of utility and ε_{its} is the unobserved portion of utility. ε_{its} is assumed to be independently, identically distributed extreme value, in which the independence means that the unobserved utility for one alternative is unrelated to the unobserved utility for another alternative (Train, 2009). We will use different specifications of V_{its}, thereby imposing different assumptions on the

[13] Four schools (with seven overlapping school tracks) are registered formally as two separate schools. These two pairs of schools cooperate very closely, and only get one quality score for the pair. Therefore, we used this quality score for both schools.

Table 3: Changes in school quality scores over time

School track	N Schools	N stable schools	Quality scores in current year	Quality scores in previous year Mis.	Neg.	Neut.	Pos.
			Pre-university				
Pre-uni.	27	10	Missing	9	0	0	0
Senior gen./pre-uni.	21	6	Negative	4	11	2	0
			Neutral	2	8	8	2
			Positive	0	1	9	24
			Total	15	20	19	26
			Senior general secondary				
Senior gen./pre-uni.	21	7	Missing	13	0	0	0
Senior gen.	23	9	Negative	1	0	3	1
Pre-voc. IV/senior gen.	18	7	Neutral	0	6	21	4
			Positive	1	4	5	22
			Total	15	10	29	27
			Pre-vocational IV				
Pre-voc. IV/senior gen.	18	4	Missing	14	0	1	1
Pre-voc. IV	23	5	Negative	0	3	6	2
			Neutral	4	8	16	2
			Positive	1	1	6	2
			Total	19	12	29	7
			Pre-vocational II				
Pre-voc. II	22	5	Missing	3	0	3	1
Pre-voc. I/II	21	5	Negative	0	4	2	0
			Neutral	2	5	17	8
			Positive	2	1	5	13
			Total	7	10	27	22
			Pre-vocational I				
Pre-voc. I/II	21	11	Missing	3	0	0	3
Pre-voc I	22	11	Negative	0	3	0	0
			Neutral	1	1	6	4
			Positive	1	0	7	36
			Total	5	4	13	43

Note: In this table, we include all relevant schools that have a quality score for a certain school track. We do not take into account the fact that some schools are not in the choice set for combined tracks (a student with combined senior general secondary/pre-university advice cannot go to a school that only offers pre-university education). The number of stable schools indicates the number of schools that have the same quality score over the four included years.

model. In all cases, we investigate school choice separately for each group of students with the same school track advice. The reason is that the students' choice sets differ by advised school track.

First, we analyze whether school choice can be predicted by student and school characteristics, in which $V_{its} = x'_{its}\beta + w'_{ts}\delta + z'_{it}\gamma_s$. Here x_{its} are characteristics that are specific for a student-school combination. In the main analysis this will only be the home to school distance. In additional analyses it will also include the share and the predicted share of a students' primary school peers going to that secondary school. w_{ts} are secondary school specific factors: the school quality and whether the school had a lottery for that school track in the previous year. z_{it} are student specific factors, specifically gender, ethnicity, test score and student weight. We include indicators for students missing values of these characteristics. The probability (p_{its}) that student i in year t chooses for secondary school s, is given by:

$$p_{its} = \frac{exp(x'_{its}\beta + w'_{ts}\delta + z'_{it}\gamma_s)}{\sum_l^m exp(x'_{itl}\beta + w'_{tl}\delta + z'_{it}\gamma_l)} \tag{1}$$

The parameters β, δ and γ are estimated using maximum likelihood estimation, in which the log likelihood function is given by $LL(\beta, \gamma_s, \delta) = \sum_{s=1}^S \sum_i y_{its} \ln p_{its}$, where y_{its} is a binary indicator equal to one if student i in year t chooses school s, and zero otherwise.

In specification 1, we do not take into account differences between secondary schools, apart from school quality and having a lottery in the previous year. It is likely, however, that unobserved characteristics (in our database) of the secondary school, such as the quality of the building and offering extracurricular activities, are important in secondary school choice. In specification 2, we therefore add secondary school fixed effects. d_{sj} is a secondary school dummy, which has $d_{sj} = 1$ if $s = j$ and $d_{sj} = 0$ if $s \neq j$. Adding secondary school dummies changes the interpretation of the quality scores: we no longer use the absolute scores of a secondary school, but study how school choices change when school quality scores change.

$$p_{its} = \frac{exp(x'_{its}\beta + w'_{ts}\delta + z'_{it}\gamma_s + \sum_{j=2}^J d'_{sj}\alpha_j)}{\sum_l^m exp(x'_{itl}\beta + w'_{tl}\delta + z'_{it}\gamma_l + \sum_{j=2}^J d'_{lj}\alpha_j)} \tag{2}$$

Models 1 and 2 are conditional logit models, which assume that the errors ε_{its} are independent of each other such that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative. This lack of correlation gives rise to the property of independence of irrelevant alternatives (IIA). IIA means that for any two alternatives j and k, the relative odds of choosing j over k are the same, such that the ratio is not dependent of the presence or attributes of other alternatives (Train, 2009). In practice, this means that two schools (of the same track) will be equally affected by the opening of a new competing school. This is not very likely: the school that is closest related in terms of distance or school policy will probably be affected more.

Mixed logit models do not have the IIA property. Therefore, we have also estimated mixed logit models for secondary school choice. As can be seen in equation 3, we estimated random coefficients for the student-school specific characteristics and the school specific characteristics (quality, lottery in the previous year and distance). $\phi(\beta, \delta|b, d, W)$ indicates the mixing distribution, in this case the normal density with means b and d and covariance W.[14]

$$p_{its} = \int \left(\frac{exp(x'_{its}\beta + w'_{ts}\delta + z'_{it}\gamma_s + \sum_{j=2}^{J} d'_{sj}\alpha_j)}{\sum_l^m exp(x'_{itl}\beta + w'_{tl}\delta + z'_{it}\gamma_l + \sum_{j=2}^{J} d'_{lj}\alpha_j)} \right) \phi(\beta, \delta|b, d, W)\, d\beta \tag{3}$$

Another advantage of the mixed logit model is that it allows for variation in preferences for school characteristics. It might be that some students put a high weight on the published quality scores, while other do not. Or some very risk-averse students may avoid a school that had a lottery in the previous year, while others do not care so much. In the mixed logit model, we do not only estimate the means of the preferences for school characteristics, but also their standard deviations, which allows us to see to what extent the preferences for secondary school characteristics vary across students.

4.2 Peer effects in school choices

The school choice literature typically assumes that students make their choices individually and ignores possible influences of peers. Casual observation suggests that this is unrealistic. Students who go from a primary school to a secondary school seem to be influenced by the choices of their classmates in primary school. They may coordinate their school choices, or some students may follow the choices of others. There may, in other words, be peer effects in school choices.

Peer effects are intrinsically difficult to analyze (Manski, 1993; Angrist, 2013). The following identification problems are present. The first is the reflection or simultaneity problem. In the presence of peer effects, it is not always clear how the causality runs - do peers affect the respondent, or does the respondent affect peers? The second is the self-selection problem. Group membership is endogenous - peers select themselves on similar characteristics. In the presence of peer effects, it is not always clear whether effects are driven by behavior, or by unobserved characteristics that are correlated with it. And related, there is the errors-in-variables problem. It is not always clear how to define a peer group. What constitutes a peer?

The economics of education literature has so far been unsuccessful in addressing all these problems in a satisfactory way. We follow Sacerdote (2001) and include the fractions of primary school classmates choosing for each secondary school in the logit model of students' secondary school choice. The results are subject to the reflection problem and the correlated errors problem and can therefore not be given a causal interpretation. The

[14] Instead of a normal density, Hastings et al. (2009) assume a negative lognormal distribution for distance, which imposes that all people dislike commuting. Without explicitly imposing this restriction, we consistently find that students have a negative preference for distance.

results are, however, informative about the degree of correlation in the choices of classmates in primary school.

To address the correlated errors problem, we split the the fractions of primary school classmates choosing for each secondary school into a component that is common for all cohorts from the same primary school (where we allow for a trend) and an idiosyncratic component where the choices of students' primary school classmates deviate from the choices of students from other cohorts in that primary school. If inclusion of the first component solves the correlated errors problem, the estimate of the effect of the idiosyncratic component in classmates' choices is only biased due to the reflection problem. According to Sacerdote (2001), this allows us to test the null hypothesis of no peer effects which predicts no relationship between a student's own school choice and the idiosyncratic school choices of her classmates in primary school.

5 Results

5.1 Main findings

Table 4 presents estimates from the conditional logit model without school fixed effects by secondary school track. The top part of the table reports the estimates of the impact of school quality scores on students' choices. For students with a mixed advice (e.g. senior general secondary/pre-university) the quality scores of both tracks matter and estimates of the impact of two quality scores are reported, where the first estimate pertains to the highest of the two tracks.[15] Neutral quality scores are the reference category. As described before, the quality scores of 2009 are reported separately, because only the underlying scores were published by *Trouw* in this year. The estimated coefficients for the quality scores provide little support for the hypothesis that school choice depends on published quality scores. Two coefficients for a positive quality score are significantly positive and three coefficients for a negative quality score are significantly negative. At the same time two coefficients for a positive quality score are significantly negative and two coefficients for a negative quality score are significantly positive. Just focusing on the sign of the coefficients, it turns out that 14 out of 22 coefficients have a sign that is consistent with the hypothesis that a positive (negative) quality score makes it more (less) likely that a school is chosen. We cannot reject that this number of correct signs is generated by chance (p=0.86). By and large we can therefore not reject the hypothesis that published quality score have no systematic impact on students' school choices.

Distance turns out to be a consistent and strong predictor of school choice. This is in accordance with results from other studies (e.g. Hastings et al., 2009; Koning and Van der Wiel, 2013). One way to interpret the size of the coefficients is in terms of odds ratios. The logit coefficient for distance to pre-university schools in Table 4 can be expressed

[15] Consider for example the estimates for a negative quality score for students with a mixed senior general secondary/pre-university advice. The estimate of -0.380 is the effect of a negative quality score of the pre-university track in a school and the estimate of 0.304 is the effect of a negative quality score of the senior general secondary track in a school.

Table 4: Estimates from conditional logit models without school fixed effects

	Pre-uni.	Senior gen./ pre-uni.	Senior gen.	Pre-voc. IV/ senior gen.
Negative quality	0.371***	-0.380***	0.106	0.001
	(0.111)	(0.127)	(0.084)	(0.101)
		0.304***		-0.307***
		(0.108)		(0.109)
Positive quality	0.477***	0.474***	0.122	-0.055
	(0.093)	(0.105)	(0.077)	(0.124)
		0.046		-0.321**
		(0.114)		(0.143)
Missing quality scores	-0.414***	-0.377**	-0.775***	-0.602
	(0.136)	(0.181)	(0.169)	(0.413)
		-0.705**		0.363
		(0.298)		(0.393)
Negative quality 2009	-0.011	0.696**	-0.058	0.108
	(0.229)	(0.268)	(0.383)	(0.459)
		-0.503		0.384
		(0.420)		(0.289)
Positive quality 2009	-0.293**	0.062	-0.274**	0.182
	(0.119)	(0.188)	(0.107)	(0.159)
		-0.100		0.181
		(0.162)		(0.208)
Missing quality scores 2009	-0.118	0.549	-0.006	-0.222
	(0.270)	(0.590)	(0.173)	(0.263)
		-0.768		
		(0.484)		
Distance	-0.535***	-0.552***	-0.520***	-0.553***
	(0.020)	(0.020)	(0.021)	(0.024)
Lottery	0.335***	-0.078	0.035	0.462***
	(0.067)	(0.064)	(0.067)	(0.137)
N students	4374	2819	3133	2028
N schools	27	21	23	18
Log Likelihood	10375.3	-5991.1	-7068.0	-3703.4
Pseudo R-squared	0.15	0.24	0.24	0.31

(Table continues on next page)

Table 4: (continued)

	Pre-voc. IV	Pre-voc. II	Pre-voc. I/II	Pre-voc. I
Negative quality	-0.222***	0.080	-0.115	-0.355
	(0.085)	(0.123)	(0.212)	(0.555)
			-1.037	
			(0.735)	
Positive quality	-0.189*	0.034	0.171	0.069
	(0.098)	(0.089)	(0.131)	(0.105)
			0.190	
			(0.169)	
Missing quality scores	-0.260**	0.160	-0.382	0.058
	(0.129)	(0.145)	(0.556)	(0.139)
			0.380	
			(0.587)	
Negative quality 2009	0.354**	-0.605	-0.763	0.385
	(0.169)	(0.482)	(0.578)	(0.404)
Positive quality 2009	-0.393**	-0.317*	-0.657**	0.238
	(0.159)	(0.170)	(0.259)	(0.218)
			0.039	
			(0.293)	
Missing quality scores 2009	-0.567***	-0.568**	-0.538	-0.375
	(0.191)	(0.258)	(0.650)	(0.377)
			-0.257	
			(0.777)	
Distance	-0.546***	-0.482***	-0.531***	-0.582***
	(0.022)	(0.017)	(0.023)	(0.017)
Lottery	0.633***	0.113	0.219	0.049
	(0.134)	(0.338)	(0.334)	(0.183)
N students	2881	2130	1039	2713
N schools	23	22	21	22
Log Likelihood	-5446.1	-4365.7	-2059.1	-5400.9
Pseudo R-squared	0.32	0.29	0.31	0.33

Note: Standard errors are clustered at the primary school level. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight. ***p<0.01 **p<0.05 *p<0.10.

as an odds ratio by taking the exponent: $\exp(-0.535) = 0.59$. This coefficient indicates that students with a pre-university advice are 41% less likely to choose for a school that is 1 kilometer farther away, as compared to an identical school 1 kilometer closer. The coefficients for the lottery dummy are positive and significant in three out of eight cases, but they have the "wrong" sign in the sense that they suggest that students are more likely to choose a school that had a lottery in the previous year. This result is robust for different specifications of the lottery and for using different samples: we get similar results when: i) we use the percentage of students losing the lottery in the previous year instead of the binary indicator; ii) we account for the school making an extra class in the year of school choice and; iii) we drop students who have priority for admittance to their school of first choice. Results are reported in Table A1 in Appendix A.

Table 5 presents estimates from the conditional logit model including secondary school fixed effects. This alters the interpretation of the quality scores: we no longer use the absolute scores of a secondary school, but study how school choices respond to changes in school quality scores. The estimation results are very similar to those in the previous table without school fixed effects. On the basis of the estimated coefficients for the school quality scores we can again not reject the hypothesis that quality scores have no systematic impact on students' school choices. Also the estimated effects for distance are very similar to those in the previous table. The inclusion of school fixed effects does, however, change the estimates of the coefficients of the lottery indicator. Now five out of eight coefficients have the expected negative sign, with two coefficients being significant at the 10% level and one coefficient being significant at the 5% level. There is thus some evidence that students avoid schools that changed from not conducting a lottery to conducting a lottery in the previous year.

Table 6 presents estimation results of mixed logit models. In addition to the mean of the preference weights for school characteristics the mixed logit model also produces estimates of their standard deviation. This allows us to see to what extent the preferences for secondary school characteristics vary across students. Again, the estimates indicate that students have a lower preference for schools that are farther from their home address. At the same time, however, we see that the aversion for distance varies significantly across students. Regarding school quality, we again see inconsistent coefficients over the different school tracks; we can still not reject the hypothesis that quality scores have no systematic impact on school choice. Some of the standard deviations for the quality scores come out significantly, indicating that the preferences for certain quality scores vary across students. Regarding the lottery indicator, it can be seen that only the mean coefficient for pre-vocational II comes out significantly. For the other school tracks, the coefficients are mainly negative but not significantly different from zero.

5.2 Peer effects

Figure 2 depicts for one single primary school from year-to-year how the students spread out across the available secondary schools. It shows that some secondary schools (such

Table 5: Estimates from conditional logit models with school fixed effects

	Pre-uni.	Senior gen./ pre-uni.	Senior gen.	Pre-voc. IV/ senior gen.
Negative quality	0.288**	-0.097	0.071	0.022
	(0.123)	(0.146)	(0.088)	(0.101)
		0.174*		-0.079
		(0.105)		(0.116)
Positive quality	0.397***	0.284***	0.006	0.014
	(0.102)	(0.108)	(0.076)	(0.129)
		0.065		-0.321**
		(0.114)		(0.153)
Missing quality scores	-0.252	-0.092	-0.069	-0.151
	(0.156)	(0.189)	(0.236)	(0.433)
		-0.225		0.168
		(0.332)		(0.409)
Negative quality 2009	0.321	0.236	0.634	-0.265
	(0.248)	(0.280)	(0.466)	(0.493)
		0.439		0.523*
		(0.492)		(0.308)
Positive quality 2009	-0.350***	-0.032	-0.210*	0.100
	(0.125)	(0.191)	(0.113)	(0.171)
		-0.069		-0.119
		(0.166)		(0.248)
Missing quality scores 2009	0.442	0.349	0.132	-0.334
	(0.378)	(0.630)	(0.182)	(0.274)
		-0.419		
		(0.473)		
Distance	-0.588***	-0.609***	-0.555***	-0.576***
	(0.023)	(0.023)	(0.021)	(0.024)
Lottery	0.066	-0.103	-0.113*	0.003
	(0.064)	(0.069)	(0.065)	(0.131)
N students	4374	2819	3133	2028
N schools	27	21	23	18
Log Likelihood	-10112.1	-5891.0	-6916.4	-3557.8
Pseudo R-squared	0.17	0.25	0.25	0.34

(Table continues on next page)

Table 5: (continued)

	Pre-voc. IV	Pre-voc. II	Pre-voc. I/II	Pre-voc. I
Negative quality	-0.153*	0.195	0.082	1.698***
	(0.091)	(0.121)	(0.212)	(0.646)
			-0.651	
			(0.976)	
Positive quality	-0.174*	-0.082	0.014	0.052
	(0.102)	(0.094)	(0.141)	(0.111)
			0.167	
			(0.195)	
Missing quality scores	0.022	-0.028	-0.367	0.027
	(0.163)	(0.161)	(0.555)	(0.145)
			0.109	
			(0.612)	
Negative quality 2009	0.578***	0.165	-0.861	0.419
	(0.189)	(0.520)	(0.598)	(0.408)
Positive quality 2009	-0.528***	-0.296*	-0.360	0.217
	(0.172)	(0.163)	(0.283)	(0.220)
			-0.215	
			(0.338)	
Missing quality scores 2009	-0.563***	-0.329	-0.085	-0.438
	(0.196)	(0.264)	(0.676)	(0.387)
			-0.874	
			(0.832)	
Distance	-0.557***	-0.495***	-0.552***	-0.588***
	(0.022)	(0.018)	(0.024)	(0.017)
Lottery	-0.291*	-0.678**	-0.151	0.026
	(0.175)	(0.287)	(0.320)	(0.187)
N students	2881	2130	1039	2713
N schools	23	22	21	22
Log Likelihood	-5272.8	-4240.6	-2003.0	-5344.0
Pseudo R-squared	0.34	0.31	0.33	0.33

Note: Cluster robust standard errors are in parentheses, standard errors are clustered at the primary school level. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight. ***p<0.01 **p<0.05 *p<0.10.

Table 6: Estimates from mixed logit models

	Pre-uni.		Senior gen./		Senior gen. pre-uni.	
	Mean	SD	Mean	SD	Mean	SD
Negative quality	-0.105	1.445***	-0.196	0.603	-0.006	0.660
	(0.184)	(0.301)	(0.193)	(0.450)	(0.225)	(0.753)
			-0.003	1.025**		
			(0.241)	(0.517)		
Positive quality	0.458***	0.889**	0.299***	0.854**	0.026	0.418
	(0.118)	(0.348)	(0.116)	(0.352)	(0.081)	(0.328)
			0.090	0.769**		
			(0.134)	(0.343)		
Missing quality scores	-1.122*	1.851**	-1.275**	2.486***	-2.431**	3.102***
	(0.617)	(0.726)	(0.513)	(0.523)	(0.976)	(0.777)
			-5.073*	5.077**		
			(2.635)	(1.970)		
Negative quality 2009	0.314	0.553	0.023	1.656***	0.139	1.973***
	(0.349)	(0.480)	(0.395)	(0.517)	(0.747)	(0.575)
			0.453	0.898		
			(0.648)	(0.554)		
Positive quality 2009	-0.192	1.164**	0.087	0.371	-0.218*	0.394
	(0.199)	(0.585)	(0.249)	(0.933)	(0.121)	(0.365)
			-0.076	0.495		
			(0.195)	(0.621)		
Missing quality scores 2009	-1.695	2.599**	1.739*	0.500	0.484	0.313
	(1.170)	(1.053)	(0.925)	(0.599)	(0.331)	(0.899)
			-4.547**	4.680***		
			(1.995)	(1.679)		
Distance	-0.660***	0.163**	-0.738***	0.174***	-0.649***	0.251***
	(0.031)	(0.071)	(0.035)	(0.059)	(0.031)	(0.042)
Lottery	0.028	0.450	-0.115	0.287	-0.107	0.035
	(0.086)	(0.399)	(0.073)	(0.246)	(0.068)	(0.126)
N students	4374		2819		3133	
N schools	27		21		23	
Log Likelihood	10197.6		-5936.0		-7007.5	
Pseudo R-squared	0.16		0.25		0.24	

(Table continues on next page)

Table 6: (continued)

	Pre-voc. IV/ senior gen.		Pre-voc. IV		Pre-voc. II	
	Mean	SD	Mean	SD	Mean	SD
Negative quality	0.037	0.027	-0.176	0.387	0.181	0.086
	(0.109)	(0.176)	(0.109)	(0.307)	(0.126)	(0.242)
	-0.128	0.020				
	(0.127)	(0.200)				
Positive quality	-0.095	0.899**	-0.184*	0.004	-0.272*	1.002***
	(0.178)	(0.413)	(0.106)	(0.113)	(0.144)	(0.351)
	-0.351**	0.156				
	(0.163)	(0.182)				
Missing quality scores	-0.133	0.550	-0.122	0.786	0.005	0.266
	(0.473)	(0.443)	(0.277)	(0.496)	(0.180)	(0.195)
	-0.356	1.520***				
	(0.565)	(0.552)				
Negative quality 2009	-0.628	1.253***	0.214	1.645**	-2.138	2.512
	(0.585)	(0.404)	(0.434)	(0.694)	(4.150)	(2.481)
	0.643*	0.215				
	(0.358)	(0.352)				
Positive quality 2009	0.134	0.464	-0.718*	1.075	-0.706	1.808
	(0.247)	(1.359)	(0.379)	(0.928)	(0.520)	(1.247)
	-0.165	0.893				
	(0.303)	(0.684)				
Missing quality scores 2009	-0.873	1.364	-1.077	1.259	-0.497	0.349
	(0.828)	(1.222)	(0.738)	(1.168)	(0.318)	(0.539)
Distance	-0.671***	0.232***	-0.634***	0.192***	-0.558***	0.152***
	(0.036)	(0.032)	(0.025)	(0.033)	(0.026)	(0.035)
Lottery	-0.021	0.282	-0.393	0.968	-1.804***	1.958***
	(0.146)	(0.289)	(0.285)	(0.589)	(0.666)	(0.522)
N students	2028		2881		2130	
N schools	18		23		22	
Log Likelihood	-3649.4		-5358.8		-4227.0	
Pseudo R-squared	0.32		0.33		0.31	

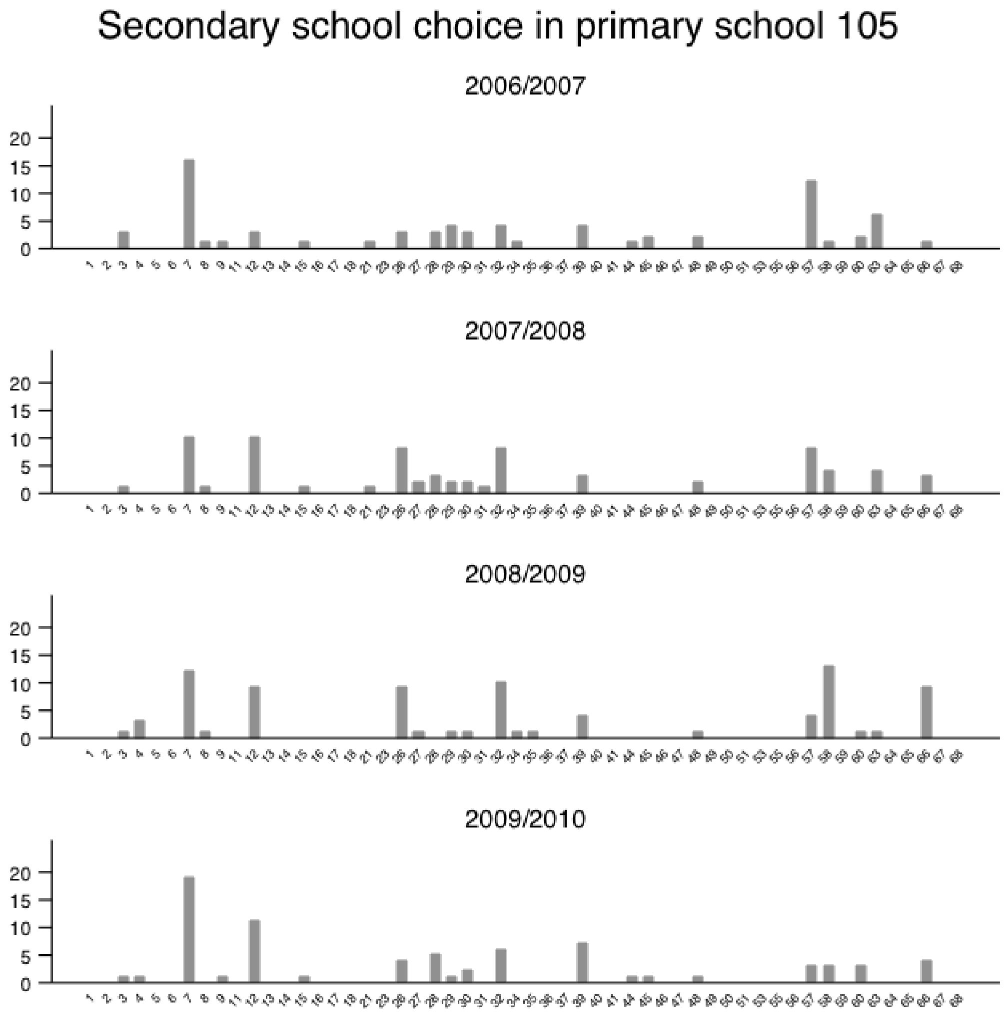
(Table continues on next page)

Table 6: (continued)

	Pre-voc. I/II		Pre-voc. I	
	Mean	SD	Mean	SD
Negative quality	0.011	0.690	-0.272	2.955***
	(0.336)	(0.945)	(1.387)	(0.615)
	-3.989	3.570		
	(4.330)	(3.042)		
Positive quality	-0.089	0.777	0.050	0.078
	(0.218)	(0.690)	(0.113)	(0.241)
	0.160	0.334**		
	(0.206)	(0.142)		
Missing quality scores	-0.364	0.232	0.012	0.087
	(0.612)	(0.611)	(0.155)	(0.506)
	-0.506	1.410**		
	(0.826)	(0.580)		
Negative quality 2009	-3.000	2.810*	0.244	1.333**
	(2.028)	(1.674)	(0.664)	(0.676)
Positive quality 2009	-0.575	1.138*	0.212	2.017***
	(0.431)	(0.658)	(0.234)	(0.712)
	-0.232	0.240		
	(0.380)	(0.319)		
Missing quality scores 2009	-0.593	0.599	-1.373*	0.158
	(0.754)	(0.414)	(0.778)	(0.487)
	-0.862	0.995		
	(1.000)	(0.737)		
Distance	-0.653***	0.178***	-0.615***	0.054
	(0.044)	(0.055)	(0.018)	(0.035)
Lottery	-0.852	1.946***	-0.017	0.258
	(0.690)	(0.668)	(0.202)	(0.677)
N students	1039		2713	
N schools	21		22	
Log Likelihood	-1984.8		-5324.8	
Pseudo R-squared	0.34		0.33	

Note: Cluster robust standard errors are in parentheses, standard errors are clustered at the primary school level. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight for pre-vocational I, I/II and II. For the pre-vocational IV to pre-university tracks, the missing indicators could not be included for computational reasons. When using Wald tests to compare the coefficients for the variables in this table in conditional logit models with and without the missing indicators, it turns out that the null hypothesis of equal coefficients cannot be rejected for the pre-vocational IV to pre-university tracks. ***p<0.01 **p<0.05 *p<0.10.

Figure 2: Secondary school choice in one primary school



Note: Frequency in the number of students choosing each school is depicted on the vertical axis, the horizontal axis are secondary schools.

as numbers 7, 12, 32 and 57) are in general more popular than other schools for students in this primary school. But we also see some strong fluctuations. For example school 58 is not a popular destination in 2006/07 and 2007/08, is the most popular destination in 2008/09 and is back to its original level again in 2009/10. The same is true for school 86. Similar graphs for other primary school reveal comparable patterns: students who leave the same primary school in the same year tend to choose the same secondary schools. To some extent this can be attributed to a general popularity of a specific secondary school among the students from certain primary schools. But there is also idiosyncratic clustering at – apparently random – destination schools.

Table 7 reports results from conditional and mixed logit models of school choice that also include the shares of classmates and of predicted classmates that choose each secondary school as variables in x_{its}. The table only reports the coefficients of these two variables. Consistent with the pattern we observed in Figure 2, all coefficients are significantly positive. Not surprisingly, we see a positive and significant coefficient for the predicted share of peers in the primary school. This coefficient captures the general popularity of secondary schools in the primary school. On top of the predicted share of peers, however, we see that the actual share of peers has a positive and significant coefficient as well. This indicates that when a higher share of the primary school peers chooses for a certain secondary school, the student himself is more likely to pick this school as well. We find this result for all specifications and also when we correct for secondary school times year fixed effects, which makes it unlikely that this result is driven by sudden changes in the attractiveness of secondary schools. The school times year fixed effects results are reported in Appendix B.

Due to the reflection problem and correlated unobservables these coefficients cannot be interpreted as causal effects of the school choices of classmates. The coefficients do not inform us about the social multiplier. Does the choice of one student influence all others, or is everyone affected by everyone? The coefficients also do not exclude that the correlated school choices are caused by a third unobserved factor, such as active recruitment by a specific secondary school in certain primary schools in a particular year. The results in Table 7 are therefore no proof of the importance of peer effects, they are, however, consistent with it.

To get an idea on the magnitude of the peer coefficients, we can compute willingness to travel coefficients if we interpret the peer coefficients as the actual preference to attend the same secondary school as primary school classmates do: $Wtt = -\beta_{peers}/\beta_{distance}$. For pre-university schools, we get a willingness to travel of $-3.464/-0.371 = 9.334$, indicating that students are willing to travel 9.3 kilometers for a 100% increase in the percentage of peers, or 943 meters for a 10% increase in the percentage of peers. Given that the average class size in primary schools is about 25, a 10% increase in the percentage of peers means an increase of 2 or 3 classmates. For the other school tracks, the willingness to travel for a 10% increase in the percentage of peers varies from 674 to 1429 meters.

6 Conclusions

We investigated which secondary schools students choose when school choice is free. It turns out that in the city of Amsterdam, publicly available information on school quality does not clearly predict school choice. Students do not avoid schools with a negative quality score, or schools that change from a neutral or positive score to a negative score. Neither are they more likely to apply to schools with a positive quality score, or to schools that change from a neutral or negative score to a positive score. Students' choices are consistent with a preference for schools close to their home, and for schools to which a larger share of their classmates in primary schools go to. There is some indication that the current application and admission procedure which follows the Boston system, triggers students to

Table 7: Estimates of peer coefficients; various logit models

	Pre-uni.	Senior gen./ pre-uni.	Senior gen.	Pre-voc. IV/ senior gen.	Pre-voc. IV	Pre-voc. II	Pre-voc. I/II	Pre-voc. I
Conditional Logit								
without school FE								
Predicted share	3.706***	3.063***	4.643***	3.062***	3.625***	2.025***	5.573***	3.642***
of peers	(0.586)	(0.684)	(0.595)	(0.653)	(0.625)	(0.723)	(1.192)	(0.642)
Share of peers	3.464***	4.259***	3.013***	5.127***	3.669***	4.554***	2.743***	3.131***
	(0.485)	(0.539)	(0.465)	(0.566)	(0.509)	(0.663)	(0.988)	(0.535)
Conditional Logit								
with school FE								
Predicted share	4.135***	3.537***	4.797***	2.659***	3.382***	1.618**	5.108***	3.530***
of peers	(0.564)	(0.681)	(0.593)	(0.640)	(0.621)	(0.707)	(1.233)	(0.635)
Share of peers	3.336***	4.282***	2.916***	5.131***	3.635***	4.506***	2.741***	3.135***
	(0.477)	(0.530)	(0.460)	(0.568)	(0.500)	(0.650)	(1.000)	(0.535)
Mixed Logit								
Predicted share	4.513***	4.022***	5.957***	4.417***	4.970***	2.786***	7.191***	3.918***
of peers	(0.644)	(0.785)	(0.699)	(0.862)	(0.760)	(0.838)	(1.705)	(0.783)
Share of peers	3.720***	5.721***	3.586***	6.117***	4.318***	5.654***	4.529***	3.808***
	(0.564)	(0.649)	(0.564)	(0.742)	(0.594)	(0.831)	(1.305)	(0.718)
SD (predicted	1.526	-0.012	2.897	-5.846***	2.063	3.545***	1.033	-0.820
share of peers	(1.699)	(3.647)	(2.954)	(0.965)	(2.792)	(1.122)	(5.083)	(2.728)
SD (share of peers)	-1.489	3.965***	-3.399	-1.657	-3.886**	4.977***	7.189***	-3.925***
	(1.174)	(0.750)	(2.294)	(2.134)	(1.717)	(0.779)	(1.730)	(0.810)
N students	4374	2819	3133	2028	2881	2130	1039	2713
N schools	27	21	23	18	23	22	21	22

Note: Standard errors are clustered at the primary school level. All models also include indicators for school quality and lottery, and distance. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight. For the pre-vocational IV to pre-university tracks, the missing indicators could not be included in the mixed logit specification for computational reasons. When using Wald tests to compare the coefficients for the peer characteristics and the indicators for quality, lottery and distance in conditional logit models with and without the missing indicators, it turns out that the null hypothesis of equal coefficients cannot be rejected for these levels. ***p<0.01 **p<0.05 *p<0.10.

avoid schools that had an admission lottery in the previous year.

Why are school choices not influenced by published school quality scores that each year attract a fair amount of attention? One reason might be that students and parents realize that the scores are changing over time and that a positive score for last year's performance is not necessarily informative about quality in the upcoming four to six years. A second reason might be that students and their parents are more interested in the reputation of a school or in a good school atmosphere than in school quality measures based on student achievement.Our analysis also lends support for a third reason, namely the apparent wish to enroll in the same secondary school together with classmates from primary school.

Using data that are arguably noisier and less suited to study the influence of school quality scores on school choice, Koning and Van der Wiel (2013) find some indication that better quality scores increase enrollment. Recall that their study uses enrollment in the third year of secondary school as a proxy for applications in the first year. If fewer students leave schools that receive good quality scores, or if these schools lose fewer students to the lower levels of the pre-vocational track (which are both likely scenarios), then the results of Koning and Van der Wiel are an biased towards finding an impact of quality scores on applications.

The results of our analysis have implications for the debate on school choice and competition. If the choices of students who can freely choose from a large set of schools are not affected by information about school quality, one important channel through which school choice can improve school quality does not seem to operate.

References

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747.

Angrist, J. D. (2013). The perils of peer effects. NBER Working Paper 19774, National Bureau of Economic Research.

Böhlmark, A. and Lindahl, M. (2013). Independent schools and long-run educational outcomes: Evidence from Sweden's large scale voucher reform. IZA Discussion Paper 6683, Institute for the Study of Labor.

Borghans, L., Golsteyn, B., and Zölitz, U. (2014). Parental preferences for primary school characteristics. *B.E. Journal of Economic Analysis and Policy*, Forthcoming.

Calsamiglia, C. and Guell, M. (2014). The illusion of school choice: empirical evidence from Barcelona. IZA Discussion Paper 8202, Institute for the Study of Labor.

CBS (2009). *Jaarboek Onderwijs in Cijfers [Education in Numbers Yearbook]*. The Hague: Statistics Netherlands.

CBS (2012). Voortgezet onderwijs; vertraging opgesplitst naar op- en afstroom [secondary education: delay split up to going to higher or lower school tracks].

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5):1191–1230.

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–275.

Eurydice (2009). *Organisation of the education system in the Netherlands 2008/2009*. Brussels: EU Education, Audiovisual and Culture Executive Agency.

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. Working paper. Formerly NBER Working Papers No. 11805 and No. 12145.

Hoxby, C. M. (2000a). Does competition among public schools benefit students and taxpayers? *American Economic Review*, 90(5):1209–1238.

Hoxby, C. M. (2000b). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4):1239–1285.

Hoxby, C. M. (2007). Does competition among public schools benefit students and taxpayers? Reply. *American Economic Review*, 97(5):2038–2055.

Hsieh, C. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90(8-9):1477–1503.

Koning, P. and Van der Wiel, K. (2013). Ranking the schools: How quality information affects school choice in the Netherlands. *Journal of the European Economic Association*, 11(2):466–493.

Ladd, H. F., Fiske, E. B., and Ruijs, N. M. (2011). Does parental choice foster segregated schools? Insights from the Netherlands. In Berends, M., Cannata, M., and Goldring, E., editors, *School Choice and School Improvement*, pages 233–253. Cambridge: Harvard Education Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Onderwijsinspectie (2007). *Aansluiting voortgezet onderwijs op het basisonderwijs [Connection between primary and secondary education]*. Utrecht: Onderwijsinspectie.

Onderwijsinspectie (2010). *De staat van het onderwijs. Onderwijsverslag 2008/2009. [The state of education. Education Report 2008/2009]*. Utrecht: Onderwijsinspectie.

Rothstein, J. (2007). Does competition among public schools benefit students and taxpayers? Comment. *American Economic Review*, 97(5):2026–2037.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Train, K. E. (2009). *Discrete choice methods with simulation (second edition)*. New York: Cambridge University Press.

Appendix A

Table A1: Alternative specifications of secondary school lotteries in the previous year

		Pre-uni.	Senior gen./ pre-uni.	Senior gen.	Pre-voc. IV/ senior gen.
	Conditional Logit without school FE				
(0)	Lottery	0.335***	-0.078	0.035	0.462***
		(0.067)	(0.064)	(0.067)	(0.137)
(1)	Percentage of students	0.011***	-0.001	0.004	0.006
	losing the lottery	(0.003)	(0.003)	(0.004)	(0.004)
(2)	Lottery controlling for	0.304***	-0.083	0.041	0.441***
	extra classes	(0.068)	(0.065)	(0.067)	(0.139)
(3)	Lottery without	0.263***	-0.180**	-0.025	0.283*
	priority students	(0.075)	(0.077)	(0.078)	(0.147)
	Conditional Logit with school FE				
(0)	Lottery	0.066	-0.103	-0.113*	0.003
		(0.064)	(0.069)	(0.065)	(0.131)
(1)	Percentage of students	0.004	-0.006*	-0.004	-0.003
	losing the lottery	(0.003)	(0.003)	(0.004)	(0.004)
(2)	Lottery controlling for	0.059	-0.113*	-0.116*	-0.075
	extra classes	(0.065)	(0.068)	(0.065)	(0.135)
(3)	Lottery without	-0.048	-0.188**	-0.146*	-0.013
	priority students	(0.075)	(0.080)	(0.076)	(0.151)
	N students (0-2)	4374	2819	3133	2028
	N priority students (3)	3462	2282	2577	1785

(Table continues on next page)

Table A1: (continued)

		Pre-voc. IV	Pre-voc. II	Pre-voc. I/II	Pre-voc. I
	Conditional Logit without school FE				
(0)	Lottery	0.633***	0.113	0.219	0.049
		(0.134)	(0.338)	(0.334)	(0.183)
(1)	Percentage of students losing the lottery	0.019***	0.008	0.005	0.001
		(0.004)	(0.007)	(0.006)	(0.005)
(2)	Lottery controlling for extra classes	0.640***	0.140	0.168	0.039
		(0.136)	(0.343)	(0.340)	(0.183)
(3)	Lottery without priority students	0.578***	0.100	0.217	0.049
		(0.131)	(0.340)	(0.334)	(0.183)
	Conditional Logit with school FE				
(0)	Lottery	-0.291*	-0.678**	-0.151	0.026
		(0.175)	(0.287)	(0.320)	(0.187)
(1)	Percentage of students losing the lottery	-0.007	-0.012*	-0.002	-0.001
		(0.005)	(0.006)	(0.006)	(0.006)
(2)	Lottery controlling for extra classes	-0.326*	-0.671**	-0.204	0.015
		(0.177)	(0.288)	(0.320)	(0.186)
(3)	Lottery without priority students	-0.278	-0.691**	-0.164	0.027
		(0.199)	(0.288)	(0.319)	(0.187)
	N students (0-2)	2881	2130	1039	2713
	N priority students (3)	2612	2099	1025	2668

Note: Each cell reports the coefficient of a lottery variable from a separate conditional logit model. All models also include indicators for school quality and distance. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight. Specifications (0) repeat the lottery coefficients from Tables 4 and 5. Specifications (1) include the percentage of students losing the lottery in the previous year instead of the binary indicator. Specifications (2) use the binary indicator and account for the school making an extra class in the year of school choice. Specifications (3) drop students who have priority for their school of first choice. Standard errors are clustered at the primary school level. ***p<0.01 **p<0.05 *p<0.10.

Appendix B

The school fixed effects models reported in Table 7 only capture time-invariant differences between secondary schools. However, changes in quality scores and having a lottery in the previous year will not be the only changes relevant for secondary school choice. Secondary schools may suddenly get more attractive because of more effort in attracting students or because of good publicity in the press. To see whether these unobserved yearly differences within secondary schools drive the positive peer estimates, Table B1 presents peer estimates including secondary school*year fixed effects. Since the secondary school*year fixed effects capture all time varying and time constant differences between secondary school, indicators for school quality and having a lottery in the previous year are dropped, leading to specification 4. As can be seen in Table B1, the peer estimates remain similar to the results without school*year fixed effects.

$$p_{its} = \frac{exp(x'_{its}\beta + z'_{it}\gamma_s + \sum_{j=2}^{J}\sum_{t=1}^{T} d'_{sjt}\alpha_{jt})}{\sum_{l}^{m} exp(x'_{itl}\beta + z'_{it}\gamma_l + \sum_{j=2}^{J}\sum_{t=1}^{T} d'_{ljt}\alpha_{jt})} \tag{4}$$

Table B1: Estimates of peer coefficients with school*year fixed effects

	Pre-uni.	Senior gen./ pre-uni.	Senior gen.	Pre-voc. IV/ senior gen.	Pre-voc. IV	Pre-voc. II	Pre-voc. I/II	Pre-voc. I
Conditional Logit								
school*year FE								
Predicted share	4.208***	3.586***	4.828***	2.847***	3.380***	1.915***	5.720***	3.644***
of peers	(0.555)	(0.713)	(0.607)	(0.675)	(0.645)	(0.701)	(1.234)	(0.629)
Share of peers	3.211***	4.194***	2.865***	4.921***	3.603***	4.261***	2.203**	2.964***
	(0.480)	(0.556)	(0.450)	(0.592)	(0.513)	(0.651)	(1.019)	(0.537)
Mixed Logit								
school*year FE								
Predicted share	4.157***	3.603***	5.402***	3.835***	3.971***	2.446***	6.707***	3.532***
of peers	(0.572)	(0.747)	(0.662)	(0.804)	(0.656)	(0.764)	(1.312)	(0.636)
Share of peers	3.433***	4.747***	3.242***	5.729***	3.878***	4.992***	2.952***	3.280***
	(0.508)	(0.586)	(0.517)	(0.691)	(0.517)	(0.709)	(1.057)	(0.571)
SD (predicted	-2.158**	-2.190**	-2.292	4.150***	2.909*	1.730	4.520***	0.370
share of peers	(0.861)	(0.909)	(1.432)	(1.322)	(1.605)	(1.657)	(1.448)	(0.355)
SD (share of peers)	1.623**	3.346***	3.968***	4.290***	2.956**	5.249***	4.671***	3.047***
	(0.724)	(0.557)	(0.703)	(1.270)	(1.449)	(0.762)	(1.447)	(0.734)
N students	4374	2819	3133	2028	2881	2130	1039	2713
N schools	27	21	23	18	23	22	21	22

Note: Standard errors are clustered at the primary school level. All models also include distance. Included student level controls are gender, Dutch, missing ethnicity, test score, missing test score, student weight and missing student weight for the conditional logit models and the same variables without the missing indicators for the mixed logit models. ***$p<0.01$ **$p<0.05$ *$p<0.10$.