Class Rank and Sibling Spillover Effects*

Alexandra de Gendre†
The University of Sydney

October 29, 2021

[Job Market Paper]

Abstract

Household responses to changes in school inputs can change drastically the returns to public investment. I study an important and understudied aspect of family decision-making: how parents and younger siblings adjust their behaviors in response to the relative academic achievement of an older sibling within her cohort. I propose a novel identification strategy for sibling spillover effects in school achievement which exploits variation in child rank in school. I use administrative records from the Netherlands, where pupils leave primary school with a national standardized test score and a tracking recommendation. Variation in class rank conditional on ability and cohort-by-school fixed effects is credibly exogenous in this setting, and isolates sibling spillovers driven by behavioral and psychological mechanisms, net of direct transmission of human capital. A 1 standard deviation (SD) increase in child rank decreases their younger sibling's test scores by around 2%SD. Higher-ranked older sibling experience faster own human capital accumulation, and same-sex older siblings have a stronger influence on younger siblings. I further show that migrant parents increase investments in school quality and speaking Dutch at home. Older sibling ranks also depress teachers' tracking advice particularly for migrant children, suggesting that teachers' form harsher expectations about younger siblings. My findings indicate that behavioral and psychological mechanisms contribute to sibling spillover effects, and school inputs are important drivers of within-family human capital spillovers.

JEL: I20, I24, D13
Keywords: sibling spillovers; rank; teacher expectations; parental investments

*I thank my advisor Olivier Marie, as well as Deborah Cobb-Clark, Hessel Oosterbeek, Dinand Webbing, Didier Fouarge, Inge de Wolf, and Rolf van der Velden for their feedback and suggestions. I thank especially Ulf Zölitz, Jan Feld, and Nicolás Salamanca for helpful comments and discussions. I also thank Fatima El Meslaki and her team at Statistics Netherlands for their invaluable support throughout this project. This research uses administrative records from *Statistics Netherlands CBS* (Project number 8308). The views expressed in this study are my own.

†School of Economics, The University of Sydney; IZA; LCC. alexandra.degendre@sydney.edu.au

1 Introduction

Household behavioral responses to public policies can affect their impact, and can change drastically their return on investments. While there is growing interest in documenting parental responses to public policies in education in particular, most studies ignore the presence of multiple children in the family and sibling spillover effects. One reason why this is the case is that we know little about within-family spillovers and sibling spillovers in particular, nor their underlying mechanisms. Understanding conceptually how sibling spillovers occur is one challenge, and estimating causal sibling spillover effects is another. Siblings may experience similar outcomes because of their similar genetic background or childhood experiences, but also because of direct peer-to-peer learning, parental decisions about resource allocation in the family, or because of teachers perceptions about siblings' ability. Although recent studies have produced convincing evidence on the causal nature of sibling spillover effects, their mechanisms remain a black box (Nicoletti and Rabe, 2019, Black et al., 2021, Karbownik and Özek, Forthcoming, Persson, Qiu and Rossin-Slater, 2021, Dahl, Rooth and Stenberg, 2020). In particular, there has been little work on behavioral, psychological and non-cognitive mechanisms operating within the family, while a vast and longstanding literature has shown the importance of socio-emotional and non-cognitive skills children's human capital development (Heckman, Stixrud and Urzua, 2006, Heckman, 2006, Currie and Almond, 2011, Almond, Currie and Duque, 2018).

I propose a novel identification strategy for sibling spillover effects by exploiting exogenous variation in older siblings' rank in school, in other words their *relative* position in class. The effect of older sibling's rank on their younger sibling isolates spillovers driven largely by behavioral and psychological mechanisms, net of the direct transmission of human capital. Because higher class ranks have been shown to improve non-cognitive skills such as motivation, self-confidence (Elsner and Isphording, 2017, Murphy and Weinhardt, Forthcoming, Elsner, Isphording and Zölitz, Forthcoming, Kiessling and Norris, 2020), higher-ranked older siblings' can be expected to increase competition in the family, or to motivate their younger siblings to study harder.

I investigate how children's relative position in class affects their younger sibling's school outcomes. My treatment of interest is the percentile rank of children in their school cohort based on their test score in a national standardized test at the end of primary school. I use high-quality administrative records from the Netherlands covering the universe of school-aged children between 2003 and 2016, a setting where children are tracked in secondary school into academic, general or vocational education based on their performance in this national standardized test, and on their teacher's tracking recommendation.

To identify the causal effects of children's rank on their siblings, I exploit idiosyncratic variation in cohort ability composition within schools, which generates variation in the relative position of a child in his school-cohort, keeping constant absolute ability. This identification strategy alleviates the two key endogeneity concerns which make it notoriously difficult to estimate any social interaction effects, and sibling spillover effects in particular the reflection problem, and the selection problem.[1] Because assignment to child rank takes place before younger siblings' outcomes, the reflection problem is not a concern in my setting. In addition, children's class rank is as good as random with cohort-by-school fixed-effects and controlling flexibly for absolute test scores using test point fixed-effects (Elsner and Isphording, 2017, Murphy and Weinhardt, Forthcoming). The key identifying assumption for my empirical strategy, which I test, is therefore that children's rank is also as good as random for their younger sibling.

I find that a 1 standard deviation increase in children's rank decreases their younger sibling's test score in Dutch by 2.1% standard deviation and in Math by 2.6% standard deviation. I also find that children's rank increases by 0.44 percentage point (2.7 percent compared to sample mean) the likelihood that their younger sibling receives a teacher recommendation to pursue the vocational track.

Combining my data with my this identification strategy allows me to explore unique mechanisms potentially underlying sibling spillover effects. I provide evidence on the role of siblings' own capital accumulation and role modelling effects, parental investments and teacher expectations for explaining sibling spillover effects.

To provide evidence on the importance of siblings and sibling interactions in response to older siblings' rank, I first investigate whether older siblings may accumulate more human capital following a higher rank in primary school, thus replicating in the Netherlands some of the findings of other studies on rank effects in education (Elsner and Isphording, 2017, Murphy and Weinhardt, Forthcoming, Denning, Murphy and Weinhardt, Forthcoming). Higher-ranked older siblings receive better tracking recommendations, attend secondary schools with a higher proportion of high-achieving pupils, and are more likely to graduate from high school in the Academic track and with a STEM profile. Based on previous literature on sibling spillover effects, we would however expect to find positive spillover effects on younger siblings' test scores. That I find negative spillover effects suggests that other mechanisms are at hand. I next investigate whether spillovers depend on the sex composition of sibling pairs. I find that same-sex pairs present stronger effects than mixed-sex pairs. My findings are consistent with recent evidence on the formation of gender norms in the family, which find that siblings' gender composition matter for the intensity and direction of sibling spillover effects (Brenøe, 2021, Altmejd et al., 2021, Dahl, Rooth and Stenberg, 2020).

I then focus on parental investments, by investigating the impact of child rank on three additional outcomes related to parental investments: parental school choice, language spoken at home and

1. The reflection problem is the inability to discern endogenous, contemporaneous effects arising in social interactions from exogenous effects arising from exposure to exogenous peer characteristics. The selection problem arises from the inability to separate contextual effects from peer background characteristics, typically caused by endogenous peer group formation. Lastly, the common shocks problem is the inability to discern contextual effects from endogenous effects. See Manski (1993*a*,*b*, 2000) for more details.

whether the younger sibling is on time with respect to his cohort defined by date of birth. In response to child rank, I find that migrant parents from Non-Western ethnic backgrounds (mostly Turkish and Northern-African) choose better-performing primary schools for the younger sibling, are more likely to speak Dutch at home by 3.7 percentage point and are less likely to have repeated a grade in primary school by 1.9 percentage point. I find similar effects although in smaller magnitude for families with a Western migrant background. In contrast, I find that native-Dutch parents invest in speaking another language than Dutch at home. These effects are informative about parental investments because speaking Dutch at home for migrant families implies time and effort, choosing a better school also requires time and search effort. That younger siblings are less likely to repeat a grade in primary school does not immediately indicate additional parental investments, but can be interpreted as the result of additional parental investment, e.g. in doing homework. These findings thus show that parents increase parental investments in younger siblings in response to their older child's rank in primary school.

Lastly, I investigate the role of teachers in spillovers of children's rank, by producing two additional sets of results in teacher tracking recommendations. I first exploit a policy change in admission procedures in secondary schools in 2014, which increased the weight of teacher assessment relative to standardized test scores in determining the track that a child follows in secondary school. Increasing the weight of teacher recommendations can increase reputational concerns of primary school teachers to secondary schools where their pupils eventually go, and increase parental pressures. Both reputational concerns and parental pressures can be expected to decrease in magnitude the negative impact of child rank on teacher's likelihood to recommend the vocational track. I show that after the reform child rank increases the likelihood that teachers recommend the Academic track instead of the vocational track, suggesting that policies such as school admission rules can reduce socioeconomic disadvantage through teacher tracking recommendations. Next, I compare sibling spillover effects for Dutch children and children with Western and non-Western migrant backgrounds. This heterogeneity is informative because children from non-Western migrant backgrounds are more likely to be disadvantaged and potentially suffer from negative teachers bias, while Western migrant are more likely to be privileged and experience positive teacher bias. I find that for families from Western ethnic backgrounds, child rank increases the likelihood that teachers recommend the prestigious academic track to their younger siblings; in contrast, child rank has no impact on teacher recommendations for children from Non-Western ethnic background, even though these children actually experience positive sibling spillovers of rank on test scores. These findings suggest that teachers rely on child rank as a noisy signal of ability to formulate tracking recommendations for younger siblings, and that teachers may contribute to sibling spillover effects.

Albeit small, the impact of children's rank in primary school is persistent on their younger sibling's educational trajectories beyond primary school. Younger siblings attend secondary school cohorts with lower test scores by 1.2%SD and thus have slightly higher relative ranks within their new cohort. They and are 0.57 percentage point more likely to study in the Vocational track and 0.66 percentage point less likely to study in the General track. Eventually they are 0.36 percentage point more likely to graduate from high school with a STEM specialization within their track.

This study expands the literature on sibling spillover effects in education by providing new evidence on the role of siblings, parents, teachers and siblings as underlying channel of sibling spillover effects. Previous studies have focused on two channels. The first channel is direct learning from one sibling to another affecting educational choices, such as direct transmission in the *quantity* of human capital. Nicoletti and Rabe (2019) studies the impact of children's human capital accumulation measured through test scores in school on their younger siblings' own test scores in secondary school, and estimates models with child fixed-effects. Karbownik and Özek (Forthcoming) explores the impact of children starting school on their siblings' test scores in secondary school, estimating models of regression discontinuity in school starting age. Altmejd et al. (2021) studies sibling spillover effect of accessing university, using university admission eligibility cutoffs in Chile, Croatia, Sweden and the United-States. In addition, two studies show evidence regarding educational choices and preferences: Joensen and Skyt-Nielsen (2018), which studies the effects of older siblings participating in an experimental educational program on their younger sibling's field of study in school, and Dahl, Rooth and Stenberg (2020), which provides evidence of spillovers in high school specialization. In contrast, my estimates exclude the direct effect of child ability on their younger sibling's performance, but rather identify residual effects driven by psychological and behavioral mechanisms. My findings regarding the sex composition of sibling pairs suggest new mechanisms related to role modelling, which complement the evidence of Dahl, Rooth and Stenberg (2020). The second channel discussed in the literature is parental reallocation of resources in the family. Black et al. (2021) use the difference in exposure between a first and a second-born sibling to a disabled third-born sibling as a measure of parental attention. Autor et al. (2019) show that sibling spillovers contribute to patterns of social disadvantage. Persson, Qiu and Rossin-Slater (2021) show that diagnosing one child with ADHD leads to increase treatment of ADHD for their siblings, suggesting that parents increase investments in detection for other children in the family. My findings on parental school choice, speaking the local language and grade repetition provide direct tests of parental responses and additional evidence regarding parental investment choices across children. In addition, this study provide the first evidence, albeit suggestive, on the role of teachers in sibling spillover effects through tracking.

This paper also contributes to a nascent literature which studies the causal effects of children's ordinal rank on educational outcomes. This study provides new estimates on the impact of child rank on their siblings' educational outcomes. To the extent that underlying ability distributions can be compared, these effects correspond roughly to a third in magnitude of the effects of own rank on later test scores that have been previously estimated in the literature (see in particular Murphy and Weinhardt, Forthcoming, Elsner, Isphording and Zölitz, Forthcoming). This study also provide the first evidence that teachers contribute to rank effects in education. Previous literature has found evidence of two channels: perceived own ability and academic self-efficacy (Elsner and Isphording, 2017, Murphy and Weinhardt, Forthcoming, Elsner, Isphording and Zölitz, Forthcoming, Goulas, Griselda and Megalokonomou, 2020, Kiessling and Norris, 2020) and increased effort, motivation and self-selection in the school system Elsner, Isphording and Zölitz (Forthcoming). Elsner and Isphording (2017) and Murphy and Weinhardt (Forthcoming) show that students' self-concept, effort in school, motivation and mental health are positively affected by own percentile rank. Elsner, Isphording and Zölitz (Forthcoming) show that in the short run, university students adjust their effort allocation between subjects and specialize into majors in response to their rank in tutorials. Compared to these studies, this paper provides new evidence consistent with teachers and schools responding to child rank.

Finally, this study also relates to an emerging literature on formation of teachers' expectations about children's ability. Teachers are one of the key inputs in the education production, and teachers with biased expectations may prevent students from reaching their full potential. While there is increasing evidence documenting teachers bias (e.g. Lavy, 2008, Alan, Ertac and Mumcu, 2018, Carlana, 2019, Papageorge, Gershenson and Kang, 2020), we know little about how teachers expectations form. This study provides suggestive evidence that teachers rely on noisy signals about children's ability when forming expectations, and use siblings as reference points.

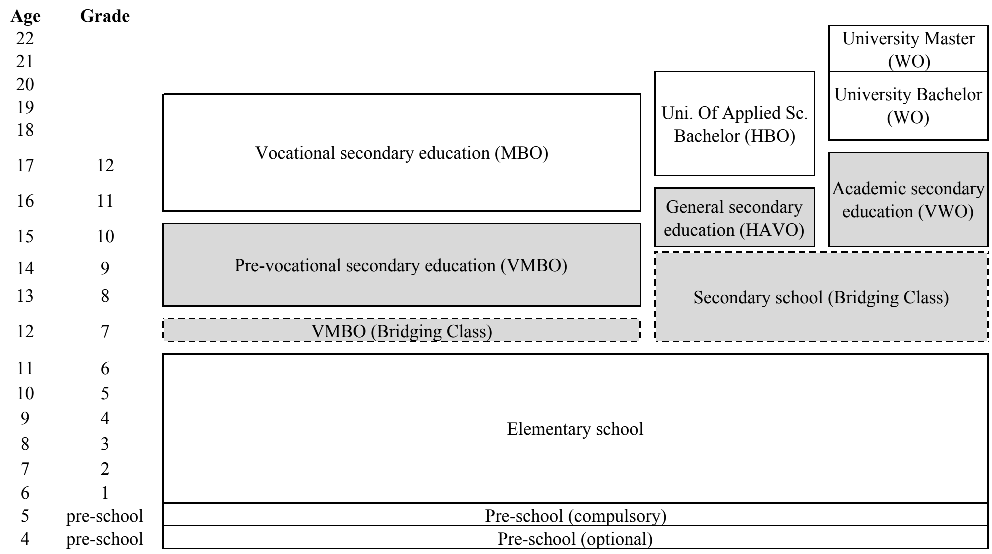
The remainder of the paper is structured as follows. Section 2 describes the institutional environment. Section 3 describes the data sources. Section 4 discusses the empirical strategy and shows the validity of the identification strategy. Section 5 presents the main results. Section 6 investigates underlying mechanisms from parents, teachers and siblings. Section 7 presents longer-run spillover effects of child rank on their younger sibling's outcomes. Section 8 presents additional evidence supporting the robustness of the findings. Section 9 concludes.

2 Institutional Setting

Figure 1 provides a simplified description of the Dutch educational system. Education in the Netherlands is compulsory from age 5 until age 16. Most children enter kindergarten at age 4 and start attending primary school two years later around the age of 6. Primary school takes six years to complete and ends around age 12. The academic year runs from September to June of the next calendar year. Upon exiting primary school at the end of June, pupils receive two key indicators of their academic ability: 1) an objective measure of academic ability in the form of an externally graded standardized test score, and 2) a subjective assessment given by the primary school teacher in the form of a recommendation to pursue one of three tracks in secondary school (vocational, general or academic). Below I present details on the test and teacher tracking recommendations.

The standardized test. Since its creation in 2005, over 80% of primary schools use the standardized test to 6th Graders as their primary school exit test. The standardized test is

Figure 1: *The Education System in the Netherlands*



Authors' figure. Simplified description of the Dutch educational system.

designed every year by the CITO Group (In Dutch, *CITO Groep*), an independent private firm specialized in educational testing (akin to the US-based Educational Testing Service). The test is designed to be a curriculum-free test of cognitive ability with sub-components covering skills in mathematics, language, general academic preparedness and geography.[2] The test is curriculum-free in the sense that it is meant to capture minimum knowledge that children are expected to have when leaving primary school, irrespective of the curriculum chosen by the school.

Every year, the CITO Group collects and grades the test externally and independently. Final test scores are a re-weighted sum of correct answers to each sub-component. They range between 501 and 550 points and are standardized such that the national annual average is always at 535 points with a standard deviation of 10. The CITO Group recommends a mapping from pupils' scores to tracks, to ease interpretation of standardized test score results. The CITO Group publicly recommends the following mapping from test scores to tracks; by construction the shares of pupils in each bin are stable: 1) vocational track (from 501 to 532), 2) either vocational or general track (from 533 to 536), 3) general track (from 537 to 539), 4) either general or academic track

2. Some schools do not take the Geography sub-components of the test. In my estimation sample, this represents roughly 8% of children

(from 540 to 544), and 5) academic track (from 545 to 550).

Pupils take the standardized test between February and May during 6th Grade before exiting primary school, and receive their score in June when they are in the process of choosing their secondary school from 7th Grade onward.

Teacher tracking recommendations. Teachers formulate their tracking recommendation in June at the end of 6th Grade, to help families in choosing secondary schools. This recommendation is based on teachers' own observation of pupils' in class, and is meant to help parents and children choose the most appropriate secondary school track given the child's ability. In this sense, teachers tracking recommendations reflect teachers' subjective assessment about the child's academic ability, which can be treated as teachers' expectations about a child's ability to graduate from each track. Teachers can also give mixed recommendations, e.g. Academic/General, General/Vocational or Academic/General/Vocational when they are unsure about the most appropriate track for the child.

Test scores only modestly predict teacher's recommendations because teachers' recommendations are not solely based on academic performance of 6th graders, but also on their behavior in class and socio-emotional development. In addition, a standardized test score also only modestly predict the track that children eventually attend, because school choice is free in the Netherlands, in the sense that parents and children can apply to any secondary school, irrespective of location, past grades and standardized test score, or teacher track recommendation.[3]

A reform in 2014 changed the relative weight of standardized test scores and teacher recommendations in determining the track for children. This reform of secondary school admission procedures decreased the importance of the standardized test score relative to the primary school teacher's track recommendation. Before the reform, children would exit primary school with a report card containing their their test score and their primary school teacher's track recommendation. Whenever a child had a lower score than expected and corresponding to a lower track compared to their teacher's recommended track, parents could indicate to secondary schools that the test was not representative of the child's ability. Because there were public concerns that this system could exacerbate disadvantage from parental disengagement with schools, from the academic year 2014-2015 onward, the track recommendation became the more important criterion in report card. In Section 6, I discuss how this reform changed teacher's incentives with respect to tracking

3. School funding is fixed per student, which implies that secondary schools have no incentives to orient students towards one type of track over another, and therefore the admission process for secondary schools is largely non-selective. This system of free choice and unconditional funding has led to a concentration of secondary schools; the majority of schools is very large and offers all tracks. Roughly 3% of secondary schools are, however, selective in the sense that they do require a standardized test score above some threshold, typically in order to offer a unique high-level academic track (Gymnasium). Although secondary schools decide eventually on the track that a pupil follows, parents and their child can influence this tracking outcome in two ways: 1) through discussion with the primary school teacher on the track recommendation the child receives, and 2) by choosing a school that only offers certain tracks.

recommendations, and I exploit this change to study teachers' behavioral responses to child rank when assessing younger siblings.

Secondary school and tracking. Students enter secondary school in 7th Grade in September, and start with up to 3 years of common education. After these years of common education, students are tracked into one of three high school tracks: 1) vocational, 2) general and 3) academic.

The vocational track ends in 10th Grade, and prepares students for a vocational secondary education program, which may or may not take the form of an apprenticeship. The general track ends one year later (11th Grade), and prepares students for professionally-oriented higher education. A general high school diploma constitutes the minimum requirement for enrolling into professional higher education programs, but does not make students eligible for university education. The academic track ends in 12th Grade. A diploma from academic high school constitutes the minimum requirement for enrolling into a research university. The academic track is generally considered to be more difficult than the general track, which is considered to be more difficult than the vocational track. After graduating from their high school track, students then pursue post-secondary education following their track.

3 The Data

3.1 Education Registry Data from Statistics Netherlands

For this study, I use population-level linked administrative data from the Netherlands, a setting which offers both high-quality education registry and teacher's tracking recommendations for each child. This unique setting allows me to estimate for the first time the role of child rank on their younger siblings' educational outcomes.

I use data from the Dutch education registry linked to administrative records, allowing me to observe detailed demographic characteristics of pupils and families. I link the CITO records with education registry to construct complete education histories of children exiting primary school between 2006 and 2016, including their standardized test score (and sub-components) and teacher tracking recommendations at the end of primary school, as well as their subsequent progression through the educational system. This includes their educational institution, tracking enrolment and persistence until graduation, graduation outcome, post-secondary enrollment and major choice.

I complete these records to population registers, to obtain detailed characteristics about all household members including their sex, date of birth, country of birth, migration background (1st- or 2nd-generation migration), ethnic background (classified into Western or non-Western background). Lastly, I also use housing records, to derive the exact household composition and residence at any point in time, as well as labor market and tax records of both parents. Labor market records contain all employment spells of parents unless they are self-employed.

These data are close to ideal for studying the effects of child rank in primary school on their younger siblings' test scores and teacher tracking recommendation, because 1) I observe up to 10 cohorts of pupils per primary school; 2) standardized test scores allow me to compare cognitive ability within and across schools and cohorts; and 3) the data include the main teacher's track recommendations for secondary school.

Still, the data have limitations. I do not observe unique classroom in the data, which implies that rank is calculated at the cohort level. Observing children within classrooms would allow me, for example, to compare the impact of class rank for two children in the same school-cohort with the exact same test scores, by exploiting differences in ability composition between classrooms within school-cohort. I also do not observe a unique teacher identifiers in the data, which prevents me from being able to use e.g. teacher fixed-effect estimates and construct measures of teacher bias or value-added.

I construct a dyadic dataset containing the universe of children who took the standardized test, with at least one younger sibling who took the standardized test in a later cohort. I describe in detail the data construction process in Appendix A. The full sample includes 1,063,560 unique children who took the standardized test between 2006 and 2016, paired in 531,780 sibling dyads.

3.2 Construction of Older Sibling Percentile Rank

I construct the absolute ordinal rank of each child following Elsner and Isphording (2017) and Murphy and Weinhardt (Forthcoming). The child with the lowest score in her school cohort c_s receives an absolute rank of 1, and the child with the highest score receives a rank of N_{cs}, with N_{cs} denoting the size of cohort c_s in school s. Because this absolute ordinal rank is sensitive to school cohort size, I transform absolute ordinal ranks R_{ics} into percentile of ordinal ranks $rank_{ics}$ which ensure that we capture the *relative* position of a child in the ability distribution of his school.

$$\text{rank}_{ics} = \frac{R_{ics} - 1}{N_{cs} - 1} \tag{1}$$

With this transformation, the child with the highest score in his cohort now receives a percentile rank of 1 and the child with the lowest score a percentile rank of 0. I construct percentile ranks for each subject and for the overall score. By construction, these percentile ranks have a mean of approximately 0.5 and a standard deviation of roughly 0.29.

3.3 Summary Statistics

Table 1 presents summary statistics of the full sample of 531,780 sibling dyads with complete information. On average, younger siblings are 50% likely to be female, and live with another 2.79 siblings at the time when they take their standardized test. Younger siblings are 2.8 years younger than their adjacent older sibling, and both siblings take the test 2.8 years apart, reflecting that grade repetition in primary school is rare in the Netherlands over the period and at this level of education.

Table 1 indicates that on average 10% of younger siblings reside in one of the four largest cities in the Netherlands (Amsterdam, Rotterdam, The Hague and Utrecht). Close to 19.3% of younger siblings are of migrant background, almost all of which are second-generation migrant. 83% of younger siblings live in households with both parents present. 10% of mothers and 8% of fathers in the sample had never worked as employee by the time younger siblings take the test; this can be because these parents are self-employed. Younger siblings grow up in households where the annual taxable income in the year prior to taking the test is roughly 61,000 euros (3.5 times the annual minimum wage for an adult worker). 81% of younger siblings attend the same primary school as their older sibling; this is likely because parents choose primary schools that are geographically very close to their home. Primary schools are generally small, with 25 pupils taking the test per cohort; 53% of schools have fewer than 22 pupils when the maximum class size is 30 pupils. Cohorts - including pupils who do not take the test - count 36.4 pupils, just enough to open two classrooms. 39% of siblings are likely taught by the same primary school teacher - a dummy variable that I construct marking siblings who i) are in the same primary school, ii) are in a 1-class only primary school, and iii) take the test less than 3 years apart.

Lastly, the table also presents summary statistics on the treatment variable, older sibling percentile rank, and on the main outcome variables, younger siblings' test scores. I use both the re-weighted test score of younger siblings - the test results observed by schools and families -, and the number of points per sub-components of the test (Dutch, Math, Academic Preparedness and Geography), which are excellent measures of child subject-specific ability. To easy interpretation of estimates, I re-standardize percentile rank to have a mean of 0 and standard deviation 1. The main treatment of effect then becomes a 1 standard deviation change in percentile rank, which corresponds to jumping from the last rank in the cohort to the bottom tier of the cohort.

Table 1: *Summary Statistics*

Variables:	(1) N	(2) Mean	(3) S.D.
Younger sibling is female	531,780	0.5	0.5
Number of children in household	531,780	2.79	0.90
Younger sibling year of birth	531,780	2000.6	2.8
Age difference between siblings in years	531,780	2.83	1.49
Difference in years between both siblings std. test	531,780	2.76	1.48
Younger sibling's lives in 1 of 4 largest cities	531,030	0.11	0.31
Younger sibling's migration background (base: native)			
First generation migrant	531,780	0.01	0.11
Second generation migrant	531,780	0.18	0.38
Younger sibling's ethnic background (base: native)			
Non-Western	531,780	0.14	0.35
Western	531,780	0.05	0.22
Number of parents of migrant background (0,1,2)	531,780	0.31	0.67
Younger sibling living with mother	531,780	0.98	0.14
Younger sibling living with father	531,780	0.85	0.36
Younger sibling in a two-parent household	531,780	0.83	0.38
Parental labour market and earnings:			
Younger sibling's mother never recorded as employee*	531,780	0.10	0.31
Younger sibling's father never recorded as employee*	531,780	0.08	0.27
Mother annual pre-tax income (in Eur $1,000)*	362,817	29.42	22.94
Father annual pre-tax income (in Eur $1,000)	405,110	57.67	64.02
Household annual pre-tax income (in Eur $1,000)	531,780	60.66	66.66
School characteristics:			
Both siblings in same primary school	531,780	0.81	0.39
Number of test-taking pupils in older sibling's cohort	531,780	25.02	16.02
Mean std. test in older sibling school-cohort	531,780	535.37	4.00
Younger sibling std. test scores:			
Std. test score overall (CITO weights)	531,780	535.46	9.73
Std. test score in Dutch	531,780	81.06	18.50
Std. test score in Math	531,780	47.93	14.60
Std. test score in Academic preparedness	357,636	29.80	6.03
Std. test score in Geography	412,239	62.43	12.59
Older sibling percentile rank:			
Overall rank	531,717	0.515	0.295
Rank in Dutch	531,717	0.516	0.298
Rank in Math	531,717	0.514	0.296
Rank in Academic Preparedness	500,143	0.505	0.294
Rank in Geography	440,906	0.527	0.298

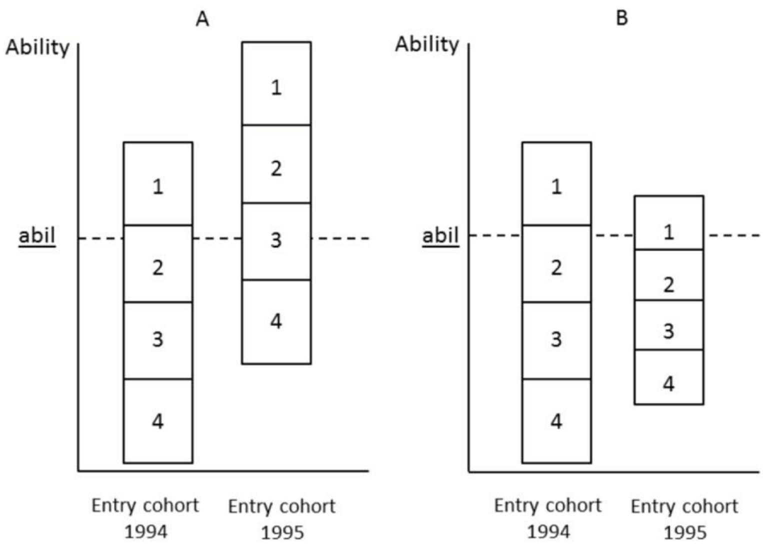
*This table presents summary statistics on the estimation sample of siblings dyads, based on education registry data from Statistics Netherlands over the years 2003 to 2016. * denotes parental labour market outcomes and income variables that are measured in the calendar year preceding child standardized test.*

4 Empirical Strategy

4.1 The Thought Experiment and Identifying Variation

The thought experiment to identify the causal effect of interest is to take two children with identical ability, but exposed to peers with different distributions of ability. With the same ability, these two children will be located at different relative ranks in their respective cohort. A child of a given level of ability can have any rank depending on their *relative* ability compared to their school-cohort. In the words of Elsner and Isphording (2017), "Conditional on attending a given school, the cohort composition is as good as random for a particular student". This thought experiment is graphically described in Figure 2 from Elsner and Isphording (2017). In this study, I want to isolate the impact of these ranks on their younger siblings, keeping constant older siblings' ability and absorbing mean differences in primary school cohorts of older siblings.

Figure 2: *Identifying Variation of Relative Rank Effects*



This figure shows schematically the source of identifying variation for rank effects arising from idiosyncratic variation in the ability composition of cohorts. From the perspective of any one child with given ability, given the school they attend, the cohort composition is as good as random and therefore so is their rank. Panel A shows that cohorts can vary in their means, at equal spread; Panel shows that cohorts can also vary in their variance, at equal mean. Thus rank effects can be identified from first and higher order moments of the ability composition of cohorts.

4.2 Estimation Equation

The isolate the causal effect of a child's rank on their next younger sibling's test score, I exploit idiosyncratic changes in the cohort composition of older siblings' primary schools over time, holding constant child *absolute* ability. A child's rank is her relative position in her primary school cohort based on her national standardized test score. Because cohorts in primary schools are small, for any given test score, ranks vary substantially.

I estimate the effect of a child's rank (indexed 1 in a pair) in primary school s in cohort s_1 on her next younger sibling's test score (indexed sibling 2 in a pair) in cohort s_2, with the following reduced-form equation:

$$\text{Test score}_2 = \alpha + \beta \text{Rank}_1 + \gamma g(\text{Test score}_1) + \delta X_2 + \lambda_{s_1} + \varepsilon_2 \tag{2}$$

where β is the coefficient of interest, corresponding to the effect of child rank Rank_1 in her school cohort, and λ_{s_1} is a fixed effect for this child's school-cohort. Test score_1 is the child's standardized test score, and $g(.)$ is a flexible functional form through which Test score_1 captures all prior inputs into up until the child's test score. In my preferred specification, I control non-parametrically for test scores using dummies for each point (point fixed-effects), and I show in Section 8 that my findings are robust to alternative functional forms for older sibling absolute ability. λ_{s_1} represents unobserved heterogeneity that is specific to the older sibling's cohort and primary school, including e.g., mean peer ability or teacher quality. I account for this source of endogeneity by including fixed effects for school-by-cohort of the older sibling. X_2 is a vector of individual characteristics of the younger sibling (which vary across specifications). α is an intercept corresponding to average test scores in the younger sibling's primary school cohort. Finally, ε_2 is an idiosyncratic error term clustered at the level of the school-by-cohort of the older sibling.

4.3 Identification

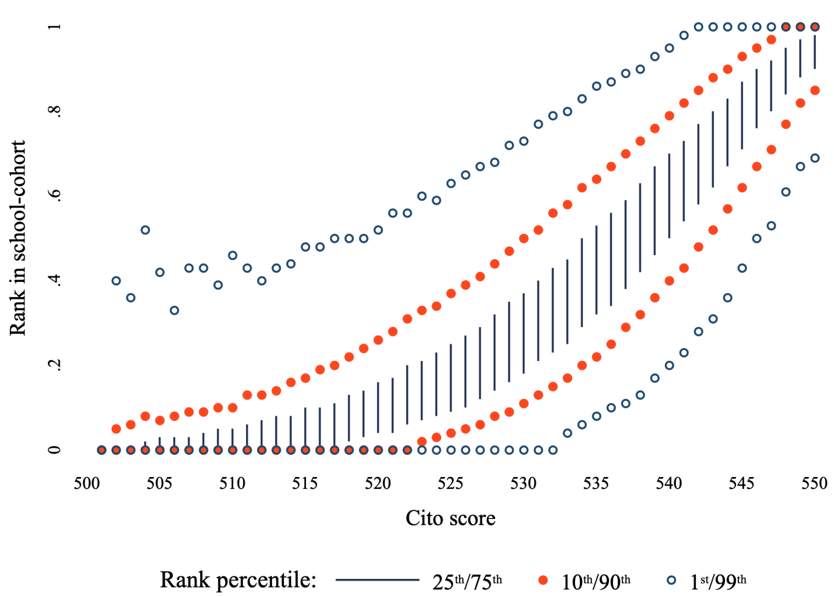
The rank parameter β can be interpreted as causal if child rank is as good as random for their younger sibling and if the relationship between child rank and their younger siblings' outcomes is correctly specified.

Older siblings' cohort-by-school fixed effect are required in the estimation equation to approximate this ideal experiment and to fulfill the two key identification assumptions. In the absence of λ_{s_1}, β could be confounded by unobserved heterogeneity in older siblings' school cohort, such as their peers' or teachers' characteristics that are productive for younger siblings' test scores beyond older siblings' academic achievement, e.g. non-cognitive skills.

This identification strategy for rank effects has been used in several recent studies (see e.g. Elsner and Isphording, 2017, 2018, Murphy and Weinhardt, Forthcoming, Denning, Murphy and Weinhardt, Forthcoming, Elsner, Isphording and Zölitz, Forthcoming). Importantly, while most of the literature on rank effects in education uses the same source of variation that I am using–that is, naturally occurring variation in cohort composition over time within the school–Elsner, Isphording and Zölitz (Forthcoming) uses data from a university where students are randomly allocated to classrooms within courses, and find effects of the same order of magnitude as previous studies. This study reinforces the notion that conditional on a child's school and for a given level of ability, the child's cohort composition is as good as random, and therefore her percentile rank is as good as random.[4].

Figure 3 displays the variation in older sibling rank remaining in my preferred specification. For every older sibling's standardized score–from 501 to 550–, older sibling percentile ordinal rank covers a wide range of values. An older sibling with the national average standardized test score of 535 can have a relative rank ranging from the lower 10% of the cohort to the top 15% of the cohort. This variation is smaller at the tails of the distribution of scores: older siblings with a standardized test scores of 545 have ranks ranging from the lower 45% of the cohort to the top of the class.

Figure 3: *Remaining Variation in the Preferred Specification*



This figure shows the remaining variation in older siblings' rank in the preferred specification. This specification regresses younger siblings' standardized test score on her child's rank, including older sibling cohort-by-school fixed effect, point dummies controlling flexibly for the older sibling's standardized test score, dummies for whether the younger sibling is native Dutch, first- or second-generation migrant. This figure is based on the complete sample of 531,780 dyads. For each standardized test score of the older sibling on the x-axis, a wide range of rank is observed in the y-axis.

4. Several studies in the broader strand of literature on peer effect exploit idiosyncratic variation in cohort composition within school over time, e.g. Hoxby (2000), Hoxby and Weingarth (2005), Lavy and Schlosser (2011), Lavy, Paserman and Schlosser (2012)

4.4 Tests of identifying assumptions

Table 2 presents the results of balancing tests of child rank on ten pre-treatment characteristics of younger siblings. I perform these balancing checks by running my preferred specification with baseline characteristics of younger siblings as outcome variable. Baseline characteristics are determined before the older sibling rank is given, such that any positive and significant regression coefficient indicates some form of selection that could bias my main results. The table shows that older sibling relative rank is effectively as good as random for younger siblings.

Table 2: *Balance of Percentile Rank on Younger Sibling Pre-Determined Characteristics*

	Treatment: percentile rank [std]				
Younger sibling characteristics:	(1) Coef. Est.	(2) Std. Err.	(3) Y Mean	(4) Y S.D.	(5) N dyads
Female	0.002	(0.003)	0.50	0.50	527,448
Date of birth (date)	3.765	(3.114)	14985.25	1003.78	527,448
Years between tests	0.006	(0.008)	2.76	1.48	527,448
Native born	-0.016^{***}	(0.002)	0.81	0.39	527,448
Two-parent household	-0.002	(0.002)	0.83	0.38	527,448
Annual pre-tax income:					
Household	4.629	(372.988)	60,737.33	66,776.37	527,448
Mother	118.273	(170.819)	29,443.59	22,933.39	356,114
Father	554.388	(447.494)	57,796.58	64,357.80	399,040
Annual working days:					
Mother	1.049^{*}	(0.575)	180.73	74.75	356,114
Father	0.608	(0.421)	246.14	56.44	399,040

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's pre-determined characteristics. Each row represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. The estimation sample contains 527,448 sibling pairs, 6,627 older sibling schools, 10,820 younger sibling schools and 53,247 older sibling school-by-cohort fixed effects. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.*

Table 2 largely confirms that older siblings' relative ranks in primary school is as good as random from the perspective of their younger siblings: out of 10 baseline characteristics tested, I find precisely estimated null effects on 8 characteristics, and a small and an economically insignificant imbalance in one characteristic, mother's annual number of working days.

The table indicates a small statistically significant correlation between older sibling rank and younger siblings' migration background. A 1SD increase in older sibling rank is associated with a higher likelihood that younger siblings are born in the Netherlands by 1.6 percentage point (2 percent change for the base rate of 81%).

It is hard to think of ways that this correlation reflects systematic rank-based sorting. Rather, this correlation indicates the presence of school segregation. Children of migrant backgrounds - especially non-Western backgrounds - are more likely to have lower test scores, and to attend schools with more disadvantaged children. Those schools are slightly less diverse in their ability composition compared to less disadvantaged schools, in other words these schools differ in higher order moments of their ability distributions. Thus, for any given test score, migrant children are more likely to have slightly higher ranks compared to native children (see Figure 2 to illustrate this intuition).

Note that these balancing tests are high powered, with a minimum detectable effect (MDE) of 1.61 days; given this precision, I can detect very small correlations that are not economically significant. This test is also high powered, with a minimum detectable effect of 0.56 percentage point change. Nonetheless, I address endogeneity concerns arising from these imbalances by running models including dummies for migration background as balancing controls to account for baseline differences across migration groups. In Section 8 I show that results are robust to various sensitivity analyses. In particular, I further show in Section 8 that there is no evidence of systematic rank-based sorting in this setting, and that results don't change by adding these balancing controls.

5 Main Results

5.1 The Effect of Older Sibling Rank on Younger Siblings Test Scores

Table 3 shows that older siblings' rank has a small negative and significant effect on younger siblings' test scores. A 1SD increase in older sibling rank in his primary school decreases younger siblings' own test score by 1.5%SD, with slightly stronger effects in the core subjects of the test: Dutch language (-2.1%SD) and Mathematics (-2.6%SD).

The effect of older sibling rank on younger sibling test scores may seem large given that it represents the effect of older siblings' percentile rank *net of older sibling absolute ability*. Yet, these effects represent about a third (in magnitude) of the effects of own rank on later test scores in the literature: Murphy and Weinhardt (Forthcoming) find that a 1SD increase in own rank in primary school at age 11 increase test scores in secondary school by 8.4 and 7.1%SD at ages 14 and 16 respectively; Elsner, Isphording and Zölitz (Forthcoming) find that a 1SD increase in own rank in university tutorials increase own later course grades by 7.3%SD.

That I find negative sibling spillovers of rank may seem surprising at first, since previous findings on sibling spillovers of human capital or test scores generally find small positive effects on test scores. However it is difficult to compare my results to estimates of sibling spillover effects found in previous literature because both outcomes and treatment differ greatly. Karbownik and Özek (Forthcoming) estimate the impact of older siblings gaining one extra year of education

Table 3: *The Effects of Older Sibling Rank on Younger Sibling Test Scores*

Outcome:	Younger Sibling Test Scores [std.]				
	(1) Overall	(2) Dutch	(3) Math	(4) Acad. Prep.	(5) Geography
Treatment variable:					
Percentile rank [std.]	-0.015^{***}	-0.021^{***}	-0.026^{***}	-0.014^{**}	-0.013^{**}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
R^2	0.30	0.40	0.35	0.29	0.31
N. schools	6,627	6,627	6,627	6,236	6,062
N. schools younger sibling	10,820	10,820	10,820	6,335	8,141
N. clusters	53,247	53,247	53,247	38,937	43,416
N. observations	527,448	527,448	527,448	353,818	406,187

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores in primary school (overall, Dutch, Math, Academic Preparedness and Geography). Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.*

due to cutoffs in the school starting age, on younger siblings standardized test scores between 3rd grade and 8th grade. They find that 1 additional year of schooling for older siblings leads to up to 6%SD higher test scores for younger siblings. These effects combine the direct impact of additional human capital with the indirect impact of improvement in non-cognitive skills, e.g. as older siblings are exposed to other children in school and to teachers, while my estimates rule out any direct effect of siblings' absolute ability. Nicoletti and Rabe (2019) estimate the effect of older siblings' test scores at age 16 on younger siblings' test scores at age 16, with school-by-cohort-by-subject fixed effects for the younger sibling. They find that a 1SD increase in older sibling test score increase younger sibling test scores by 11%SD to 15.6%SD.

My findings are however consistent with previous evidence of parents shifting resources in the family, and siblings specialization in the household. Black et al. (2021) provide evidence of parental shifts in resources in response to the birth of a disabled child which requires focused parental investments. Using a difference-in-differences strategy, they estimate the difference in test scores between first and second born children in families with v. without a disabled third child. They find that in American families with a disabled third child, second-born siblings have 11.2%SD lower test scores than first-born siblings, versus 6.4%SD in families without a disabled third-child, and effects are similar in Denmark. In a field experiment, Dizon-Ross (2019) further shows direct evidence of parents shifting beliefs after learning about the ability of their children and reallocating resources between siblings. Dahl, Rooth and Stenberg (2020) provide evidence in favor of gender-specific specialization in field of study, and Brenøe (2021) shows similar gender-specific specialization in occupational choice. I explore these mechanisms in Section 6.

5.2 The Effect of Older Sibling Rank on Younger Siblings' Tracking Recommendation

Table 4 shows that older siblings' rank has a small but significant impact on their younger siblings' tracking recommendation, which is crucial for children's educational trajectory and labour market outcomes. Panel A presents the effect of older siblings' rank on teacher recommendations for their younger siblings, without controlling for younger sibling's own test score, and Panel B additional controls for younger sibling test score.

Table 4: *The Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Treatment variable:					
Panel A. Baseline Model with Balancing Control Variables					
Percentile rank [std.]	-0.094	-0.193	0.811***	-0.218	-0.306*
	(0.180)	(0.185)	(0.249)	(0.219)	(0.185)
Panel B. Model with Younger Sibling Test Score controls and Balancing Controls					
Percentile rank [std.]	0.097	-0.117	0.444**	-0.107	-0.317*
	(0.169)	(0.183)	(0.215)	(0.215)	(0.185)
Outcome Mean	12.0	12.0	30.0	18.78	26.50
R^2	0.33	0.16	0.47	0.21	0.54
Nschools	6,627	6,627	6,627	6,627	6,627
Nschools y. sibling	10,820	10,820	10,820	10,820	10,820
Nclusters	53,247	53,247	53,247	53,247	53,247
Nobs	527,448	527,448	527,448	527,448	527,448

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each row and column represents one regression. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Panel A presents the baseline model with controls for dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant; Panel B additionally controls for younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.*

Table 4 shows that older sibling rank has a small positive effect of 0.81 percentage point on the likelihood that teachers recommend younger siblings the Vocational track (+0.44 percentage point after controlling for younger sibling own test scores, from a base rate of 30%). Rank also has a small negative effect on the likelihood of not receiving any recommendation (Col. 5).

Although small, tracking is a crucial outcome for the long-term outcomes of younger siblings since graduates of the Vocational track earn systematically lower wages compared to graduates from the General and Academic tracks.

These findings suggest that teachers' adjust their expectations about younger siblings based on past schooling experiences of older siblings. Since older sibling's rank is as good as random, we may interpret this effect as a form of biased expectation updating. These findings therefore also relate to the literature on teacher bias and the formation of teacher expectations. Papageorge, Gershenson and Kang (2020) find that a 10 percentage point increase in a teacher's subjective probability that a student completes college, increases the likelihood that the student graduates from college by 1.4 percentage points. Lavy and Megalokonomou (2019) find that a 1SD increase in teacher bias in favor of boys during high school increases the probability that boys enrol in university by 2 to 5 percentage points, and decrease this probability for girls by 1 to 5 percentage points. Carlana (2019) estimates that exposure to middle school teachers with a 1SD higher gender bias (in favor of boys) increases the gender gap in math test scores by 3.2%SD. These studies also document potential mechanisms through which teacher bias can affect children's trajectory: Lavy and Megalokonomou (2019) show that teacher bias affects student absenteeism and student effort, and Carlana (2019) shows that teacher bias affects student performance partly through student self-confidence and student sorting into less demanding tracks as a response to teacher biased track recommendations. I further explore the channel of teacher's expectations formation in Section 6.

6 Mechanisms

In this section, I discuss three potential mechanisms for these spillover effects: i) older sibling human capital accumulation, ii) role modeling effects between siblings, iii) parental investment reallocation in the household, and iv) teacher expectations.

6.1 Older Sibling Direct Human Capital Accumulation

In this section, I explore a first potential channel though which older siblings' rank might affect their younger siblings' outcomes: older siblings' own human capital accumulation. Higher ranked older siblings could gain access to better schools or tracks, and experience faster human capital growth in secondary school. It could be that this additional human capital spillovers onto their sibling (Karbownik and Özek, Forthcoming). It could also be that older siblings' increased motivation and non-cognitive skills stemming from a higher rank spillovers onto their siblings (Murphy and Weinhardt, Forthcoming). For example, a higher rank in primary school might increase access to higher value-added secondary schools or enrolment in more academically-oriented tracks.

This exercise replicates in my setting some of the earlier findings on the (direct) effect of own rank on medium to long-run human capital accumulation (Elsner and Isphording, 2017, Murphy and Weinhardt, Forthcoming, Denning, Murphy and Weinhardt, Forthcoming). I find that the effects of own rank in the Dutch setting are well in line with earlier findings in the literature.

Table B.1 shows evidence that older sibling's own rank does affect their own longer-run outcomes. In particular, a 1SD increase in own rank in primary school increases by 2.9 percentage point the likelihood of receiving a higher tracking recommendation than test scores would predict, shifting upward mostly children who are at the margin between the vocational track and the general track. 1SD higher ranked children enter more competitive secondary schools, with 0.3 percentage point higher shares of high-achieving peers in their cohort. Consistent with their higher tracking recommendation in primary school, these children are more likely to start secondary school in the general track, and less likely to start in the vocational track, and this effect persists all the way through to high school graduation. 1SD higher ranked children are 0.5 percentage point more likely to graduate from high school with a STEM specialization. I find however no evidence on the likelihood of graduating on time, enrolling in university or choosing a STEM major.

The hypothesized mechanism–that it is through human capital accumulation that older siblings' rank spillovers to siblings–would however predict that younger sibling achieve *higher* test scores. This suggests that other mechanisms are at hand.

6.2 Sibling Interactions

The effect of older siblings' rank on their younger siblings' test scores might operate through role model effects. I compare the effect of older sibling rank on younger siblings' scores depending on gender match between siblings.

I explore whether effects differ by the gender composition of sibling pairs. Table B.2 shows that the effect of child rank varies with the gender composition of sibling pairs. Although the direction of spillover is stable across all combinations of sibling gender match, I find two patterns. First, same-gender pairs present symmetric responses: 1SD increase in older brother's rank decreases their younger brother's test scores in Math by 3.4%SD (with no effects on Dutch test scores), while 1SD increase in older sister's rank decreases their younger sister's test scores by 2.5%SD (with no effects on Math test scores). Second, mismatched pairs (Brother-Sister, Sister-Brother) have systematically weaker responses in magnitude compared to same-gender pairs (Brother-Brother, Sister-Sister), but still symmetric responses: 1SD increase in older brother's rank decreases their younger sister's test scores in both Dutch by 1.5%SD and Math by 4.0%SD, while it decreases their younger brother's test scores only in Math by 3.4%SD; 1SD in older sister's rank decreases their younger brother's test score in Dutch by 4.2%SD and in Math by 2.3%SD, while it decreases their younger sister s test scores only in Dutch by 2.5%SD.

These findings are consistent with recent evidence on the formation of gender norms, which shows that siblings' gender composition matter for the intensity and direction of sibling spillover effects. Brenøe (2021) shows, for instance, that girls growing up with a brother are more likely to choose more female stereotypical studies compared to girls who grow up with a sister. Qureshi (2018*a*) show that older sisters gaining additional human capital leads to increased educational performance for their younger brothers. Altmejd et al. (2021) show that younger siblings tend to follow in their older sibling's footsteps. Dahl, Rooth and Stenberg (2020) show that sibling spillovers in high school field of study depend on the gender-match of siblings. These findings relate more broadly to a recent body of work showing that exposure to female peers – e.g. in the classroom, in schools – contributes to gender differences in educational outcomes such as scores in Math and language and the likelihood of choosing a STEM major (e.g. Brenøe and Zölitz, 2020, Zölitz and Feld, Forthcoming). My findings on Male-Female pairs could also be explained by younger sisters being discouraged in Math in response to their older brother's success in school, and similarly younger brothers could be discouraged in Dutch in response to their older sister's success in school. These findings are consistent with recent evidence on gender-specific responses to own class rank (Goulas, Griselda and Megalokonomou, 2020, Elsner, Isphording and Zölitz, Forthcoming). In line with another strand of literature on gender attitudes to competition (e.g. Carlana, 2019, Delaney and Devereux, 2019), these recent studies show that girls are more easily discouraged than boys by their relative position in the classroom.

Overall, these findings suggests that siblings interactions are an important channel underlying the transmission of schooling experiences between siblings. Although only suggestive, this evidence is consistent with earlier literature on the formation of gender norms in the family, and represents some of the first exploration of gender match effects in sibling spillover in education.

6.3 Parental Behavioral Responses

One could expect parents to adjust their investments in the family in response to child rank. Recent studies in the literature on sibling spillover effects have argued that parents shifting resources in the family is one of the leading mechanisms for the existence of spillover effects (Qureshi, 2018*b*, Landersø, Skyt-Nielsen and Simonsen, 2019, Black et al., 2021, Karbownik and Özek, Forthcoming). These studies discuss how parents could shift resources, either by focusing resources towards one sibling where higher returns on parental investments are expected (e.g. higher ability siblings) at the expense of other siblings, or on the contrary by channelling resources towards weaker children in the family, e.g. with lower health or lower school performance.

Let us first consider how parents might respond to child rank. Consider two types of parents: *i)* naive parents who view child rank as a true signal of the child's ability, and *ii)* sophisticated parents who realize that child rank reflects not the child's absolute ability but rather the child's ability relative to her peers. For naive parents, a high rank increases their perception about their child's ability and represents "good news". For sophisticated parents, a high rank is a signal of lower public investment in their child, thus representing "bad news". Based on the growing literature on parental investments (Das et al., 2013, Pop-Eleches and Urquiola, 2013, Fredriksson, Öckert and Oosterbeek, 2016, Dizon-Ross, 2019), we would expect naive parents to decrease parental investments, and sophisticated parents to increase their own investment to compensate for the perceived lower public investment, in other words we expect to find substitution effects.

Let us now add siblings as an additional layer of complexity and let us consider how parental responses to child rank might shape sibling spillover effects. Assume a family with two children, and that parents make investments choices under budget constraint. If both children are normal goods for parents, then we would expect naive parents to reinforce investments in the second child, and sophisticated parents to decrease investments in the second child. Our findings that younger siblings experience on average lower test scores could thus be explained by a model of sophisticated parents, making decisions under constraints between children who are normal goods, and whose investments are substitute to public investments.

Following the literature, several types of parental investments might respond to information about the quality of their child's school, for example they could invest more money in tutoring, or invest more time through more engagement with the school or the teacher of their children. In this setting, I do not have access to information about money investments, but I explore two distinct types of parental time investments: first, their choice of primary school for the younger sibling and second, their choice to speak Dutch at home. I also analyze whether the younger child is "on time" relative to her cohort, that is, whether she has repeated any primary school grade. This last outcome only imperfectly captures parental investments because a positive impact on the likelihood of being on time could be the results of additional homework or tutoring. In principle it could also be that younger siblings have higher ability, however this seems unlikely given that I find negative effects on standardized test scores.

I find, overall, little evidence that parents respond to child rank, however with important heterogeneity by parental migration background. Table B.3 presents overall results on parental investments. Column (1) shows that 1SD increase in older sibling rank decreases by 0.6 percentage point the likelihood that younger siblings attend the same primary school; from a base rate of 82%, this is a small effect. Column (2) shows precisely estimated null effect on average standardized test scores in the previous year, as indicator of school quality. Column (3) shows precisely estimated null effects on the number of test taking pupils in the cohort. These effects indicate that younger siblings are less likely to attend the same school as their older sibling, but not likely to attend better or worse schools.

It could be that school choice is not sufficient to pick up parental responses to child rank. For example, younger siblings may be close in age such that they are already enrolled in primary school by the time older sibling rank is observed. In the Netherlands, primary school lasts for six years, so for siblings who are less than six years apart, younger siblings are likely already in primary school by the time their older sibling takes the national standardized test. In that case, parents might be unlikely to change the younger sibling's school in immediate response to the older sibling's rank. I further consider two additional types of parental investments: whether parents speak Dutch at home (Col. 4), and whether the younger sibling is on time relative to her cohort (Cols. 5-7). The measure of parents speaking Dutch at home is available for only a third of my estimation sample, because this measure was collected only until 2012 with standardized test scores. I find that child rank has null effect on the likelihood that parents speak Dutch at home. I also find that 1SD higher ranked older siblings decrease by 0.36 to 0.38 percentage point the likelihood that their younger sibling is early or late for her age.

As discussed earlier, it is likely that my results hide important heterogeneity across family types. I further explore heterogeneity in parental responses by migrant background. Similar to Fredriksson, Öckert and Oosterbeek (2016), we could expect that non-Western migrant families to be more constrained for re-adjusting parental investments to child rank, and Dutch families to be more responsive to child rank. However, it could also be that for migrant families the older child's rank provides information about school quality and how their first child fares in school, potentially closing information gaps that these families may have about the Dutch school system. In this case, we could expect those families to respond most to rank information. Table B.4 shows that parents of Non-Western background choose schools with 1.2%SD higher average test scores for the younger sibling. These families represent most of the migrant sub-population and are more likely to live in disadvantaged neighborhood, and the effects I find suggest that parents prefer better schools for their younger child in response to the older child's rank. This effect is consistent with rank information providing information about school quality for parents from non-Western backgrounds. In addition, a 1SD increase in child rank increases by 3.7 percentage point the likelihood of speaking Dutch at home for parents of non-Western backgrounds, and by 1.1 percentage point for parents of Western backgrounds. In contrast, native Dutch parents divest from Dutch, and are 1.25 percentage point less likely to speak Dutch at home. For migrant families, speaking Dutch at home could represent an important investment in cultural assimilation and their children's future school success, while for Dutch families investing in a foreign language such as English or French could be seen as a form of parental investment, perhaps supporting their children learning those languages at school. Lastly, I also find that 1SD higher ranked older siblings decreases by 1.86 percentage point the likelihood that younger siblings repeat a grade in primary school (Col. 7) for families of non-Western backgrounds, and by 1.16 percentage point for families of Western backgrounds; I find no such effect for native-Dutch families. These findings are consistent with migrant families increasing investments in younger siblings following the signal they receive about the older child's ability relative to her cohort, in other words these parental investments substitute public investments in school quality.

Overall, my findings suggest that parents from migrant backgrounds respond to child rank by increasing parental investments in the younger child, either by speaking Dutch at home, and by choosing a higher quality school for the younger child. These in turn seem to increase the likelihood that younger siblings are on time with respect to their cohort.

6.4 Teacher Expectations

One could expect teachers to also respond to child rank in their interactions with younger sibling. For example, older siblings' rank may provide teachers with incomplete information about younger siblings' ability and unobserved traits, such as parental educational aspirations, genetics, or non-cognitive traits. Thus, older siblings' rank may help teachers in forming expectations about younger siblings when formulating their tracking recommendation. To provide evidence of teacher responses to child rank, I provide three additional sets of results.

First, I study how the spillover effects of children's rank depend on the weight of teacher recommendations for secondary school admissions. I exploit an educational reform, which changed the weight of teacher subjective assessment and standardized test scores for determining the tracks that pupils eventually pursue in secondary school. This reform of secondary school admission procedures was introduced in 2014 and decreased the importance of the standardized test score relatively to the primary school teacher's track recommendation. Before the reform, children would exit primary school with a report card containing their standardized test score and their primary school teacher's track recommendation. Whenever a child had a lower score than expected and corresponding to a lower track compared to their teacher's recommended track, parents could indicate to secondary schools that the test was not representative of the child's ability. From the academic year 2014-2015 onward, the track recommendation became the more important criterion in the report card of children exiting primary school. This reform therefore affects incentives in two ways: 1) it increased incentives for teachers to inflate subjective assessments of children, since assessments became more important for children's admissions to secondary schools; and 2) it also increased incentives for parents to put pressure on teachers to inflate teacher assessments. We should then expect spillover effects of child rank to increase *after* 2014. Table B.5 presents results on teacher tracking recommendation interacting the child rank with a dummy for whether the younger sibling's test took place after the reform. Table B.5 indicates that after the reform a 1SD increase in child rank increases the likelihood that teacher recommend the academic track, not the vocational track. The reform also significantly decreased the likelihood that teachers give mixed or no tracking recommendation, which is consistent with the notion that the reform may have increased teachers incentives to provide recommendations, and perhaps to inflate those recommendations compared to the child's ability.

Second, I compare sibling spillover effects for native-born children and children with a migrant background. Table B.6 presents marginal effects of child rank on their younger sibling's tracking recommendations across ethnic background, and shows that teachers are more likely to recommend the vocational track to native-born younger siblings, but to recommend the academic track to children from Western backgrounds. Although I control for younger sibling test score, these results could be explained by differences in baseline ability in these sub-populations - Western migrant families may be on average more privileged than the median Dutch family. However, Table B.7 shows that the negative spillovers effects on test scores found in Section 5.1 are driven by native-born children, while children from non-Western backgrounds actually experience positive spillovers of similar magnitude and children from Western background experience no spillover on test scores at all. These findings suggest that teachers behave as if their hold biased expectations at baseline about children with different characteristics, and respond to older siblings' rank with asymmetries. These findings are consistent with teachers relying on child rank as a noisy signal of ability to formulate tracking recommendations for younger siblings. They suggest that teachers' responses to child rank may exacerbate gaps in tracking outcomes of native and migrant children.

Third, I explore whether teachers' respond to siblings' rank differently when assessing girls and boys, which could indicate that teachers imperfectly update stereotypical beliefs when exposed to older siblings. Table B.8 presents results on teacher tracking recommendation interacting the child rank with the younger sibling's gender. One could expect that teachers are more likely to recommend the vocational track to boys, and the academic track to girls. Table B.8 shows that there is no heterogeneity in teachers' responses to older sibling rank in their assessment of younger siblings. One could be concerned that younger sibling's gender alone is not sufficient to capture gender stereotypical beliefs. I further show in Table B.9 that there is also no heterogeneity in teacher's responses to child rank by the gender match of siblings, except perhaps that for female-female pairs, 1SD increase in older sister rank does not increase the likelihood that teachers recommend the vocational track to younger sisters.

Thus, these findings suggest that teachers contribute to sibling spillover effects, by learning imperfectly from older siblings' rank about the ability of younger siblings. My findings suggest that teachers may imperfectly update expectations based on past experiences with the family and stereotypical beliefs about disadvantaged children that may be hard to overcome (Carlana, 2019, Papageorge, Gershenson and Kang, 2020, Alesina et al., 2018).

7 Longer-Run Spillover Effects of Child Rank

Table 5 shows that child rank in primary school has small but persistent spillover effects onto their younger siblings' educational trajectory. I present results of models controlling for younger sibling national standardized test score. Panel A presents findings regarding secondary school quality, Panel B present findings on high school track, graduation and specialization profile, and Panel C on university enrolment and major choice.

Table 5: *Spillover Effects of Child Rank on Younger Sibling Long-Term Educational Outcomes*

	Treatment: Own percentile rank [std]					
	(1)	(2)	(3)	(4)	(5)	(6)
Outcomes:	Coef. Est.	Std. Err.	Y Mean	Y S.D.	R^2	N. obs.
Panel A. Secondary school:						
Same school as older sibling (in percentage point)	−0.31	0.270	58.0	49.0	0.24	526,121
School-cohort size	−0.054	0.307	121.15	74.56	0.56	526,530
Percentile rank in cohort [std.]	0.006***	0.001	0.50	0.29	0.53	526,530
Cohort avg. test score [std.]	−0.012***	0.004	0.00	1.00	0.58	526,530
Share high-achieving peers (in percentage point)	0.090	0.090	20.0	18.0	0.47	526,530
Panel B. Tracking, high school graduation and specialization (in percentage point):						
Start Academic track	0.090	0.210	27.0	44.0	0.48	480,600
Start General track	−0.660***	0.250	23.0	42.0	0.16	480,600
Start Vocational track	0.570***	0.210	50.0	50.0	0.60	480,600
Graduate Academic track	−0.140	0.200	23.0	42.0	0.44	480,600
Graduate General track	−0.360	0.260	27.0	45.0	0.17	480,600
Graduate Vocational track	0.500**	0.220	50.0	50.0	0.57	480,600
Graduate STEM profile	0.360*	0.200	11.0	31.0	0.21	411,703
Panel C. Post-secondary education (in percentage point):						
Ever in post-sec. edu.	−0.210	0.250	53.0	50.0	0.51	418,108
STEM major	−0.230	0.370	21.0	40.0	0.14	216,003

*Note: This table shows the results of dyadic regression analyses of the effect of child percentile rank in their primary school-cohort on their younger siblings' longer-run educational outcomes. Each row corresponds to one regression. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score, dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant, and for younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.*

Table 5 first indicates that child rank has small significant spillover effects on younger sibling's secondary school quality. A 1SD increase in percentile rank decreases the quality of secondary school attended by younger sibling by 1.2% SD, as measured by average primary school test scores of younger siblings' cohort peers. This also implies that younger sibling enter secondary school with a relatively better rank in their cohort by 0.6 percentile. I find, however, precisely estimated null effects on other characteristics of secondary schools, such as school size or the share of high-achieving secondary school peers, defined as children who received primary school test score above 545, the official recommended threshold for recommending the Academic track.

Panel B shows that spillover effects of older sibling rank carry over to the track that younger siblings eventually follow. I find that 1SD increase in child rank in primary school increases the likelihood that younger siblings graduate from the Vocational track by 0.5 percentage point, and that they graduate from a STEM profile within their track by 0.4 percentage point. Although small, these effects are important because tracking affects higher education and labour market outcomes.

Overall, child rank generates small persistent spillover effects on their younger siblings' long-run educational outcomes. My findings complement and reinforce recent findings that own rank has lasting effects on a child's educational trajectory (Elsner and Isphording (2017) and Denning, Murphy and Weinhardt (Forthcoming)). Focusing on primary school rank and spillover effects, this study offers new insights into the role of teachers' and parents' behavioral responses to child rank in education. I discuss and provide further evidence on these channels in Section 6.

8 Sensitivity Analyses

In this section I discuss the sensitivity of my findings. I further explore the robustness of the identification strategy and the sensitivity of my results to alternative functional forms and alternative inference methods.

8.1 Robustness of Identification Strategy

I first provide additional evidence of that older siblings' rank is as good as random for younger siblings. Next I present the results under alternative ways to control for older siblings' academic achievement.

8.1.1 Sorting on older siblings' rank

One could be concerned that results could be driven by unobserved traits driving both older sibling rank in class and younger sibling test scores, such as rank-based sorting, in the sense that older siblings could sort into specific schools to achieve higher ranks. This form of sorting would represent a violation of the conditional independence assumption.

To assess this possibility, I first test whether older sibling pre-determined characteristics are correlated with their rank, conditional on test scores and school-by-cohort fixed effects. This form of sorting would be akin to sorting in relatively worse school cohorts, which does seem implausible in light of evidence that parents tend to choose the best possible school for their child Murphy and Weinhardt (Forthcoming). In addition, parents would need to know how their child would be located in the distribution of ability within her school cohort, which seems also implausible - as shown by the myriad of studies exploiting school-by-cohort fixed effects to study the effect of peer cohort composition on outcomes (e.g. Hoxby, 2000, Hoxby and Weingarth, 2005, Brenøe and Zölitz, 2020). Table B.10 confirms that parental pre-determined characteristics do not predict higher ranks for older siblings in any meaningful way.

Next, I also consider sorting on characteristics of parents that are pre-determined when older siblings take their standardized test. This form of sorting could occur in the presence of school segregation, and would not indicate that parents systematically choose school cohorts to maximize rank, but would rather indicate the presence of unobserved characteristics driving both rank and younger sibling outcomes, such as parental resources (other than income). Table B.11 shows that six out of seven pre-determined characteristics of older siblings have precisely estimated zero association with older sibling rank. I do find evidence that native-born older siblings have on average 1%SD higher ranks compared to foreign-born older siblings, which is consistent with earlier findings from Table 2.

8.1.2 Permutation-Based Balancing Tests

Another concern is that older sibling rank is not quite as good as random, such that results could be driven by selection bias or unobserved heterogeneity. In Table B.12, I present the result of permutation tests to assess whether younger sibling characteristics are systematically associated with older sibling rank. This test is usually used in empirical studies on peer with random assignment to peer groups, and compare the actual student group composition to counterfactual peer group compositions simulated under the null of random assignment (see Feld and Zölitz, 2017, Carrell and West, 2010, Lim and Meer, 2017, 2020, de Gendre and Salamanca, 2020).

I adapt this test to my setting, by randomly assigning children to other school-cohorts keeping their test score constant, thereby randomly assigning them a placebo rank. I repeat this random assignment to school-cohort 1,000 times, and in every permutation, I run regressions of the effect of older sibling *placebo* rank (using my preferred specification, that is conditional on test score point and school-by-cohort fixed effects) on younger sibling baseline characteristics. Every time, I store the estimated effect of child rank. Columns (1) to (3) report the number of permutations in which the true regression result (for a given outcome variable) is more extreme than the placebo result. If rank is indeed as good as randomly assigned, then we should expect that the shares in columns (1) to (3) are close to the nominal rejection rates of 0.10, 0.05 and 0.01 in most or all rows.

I then construct empirical p-values as the share of placebo effects that are larger than the effect using the true rank, which I report in Col. (4). Table B.12 largely confirms previous analyses from Tables 2, B.10, and B.11 that older sibling rank is effectively as good as random for all pre-treatment characteristics except migration background. Given that the association between older sibling rank and younger sibling migration status is driven by second-generation migrants, I further explore whether specific ethnic groups have stronger associations. I find that the association is largely driven by Moroccan and Turkish families, which is not surprising since these families are more likely to live in more disadvantaged neighborhoods and attend schools with lower achieving peers.

8.1.3 Conditioning on observable characteristics

One could still be concerned that results are biased by unobserved endogeneity associated with migration background. In my preferred specification, I include dummies for each migration status as balancing control variables (native-born, 1st generation, 2nd generation) to tackle migration background as a potential source of unobserved heterogeneity. I further show that effects on test scores (Table B.13) and on tracking recommendations (Table B.14) are stable to including or removing this set of balancing control variables. In both tables, results are stable when moving from a baseline model in Panel A without balancing control variables to the preferred specification in Panel B which includes balancing controls. For tracking recommendations (Table B.14) results are also stable when moving from a baseline model which additionally controls for younger sibling test scores (Panel C) to a model including balancing controls (Panel D).

I further show in Table B.15 that results on test scores are also robust to controlling for nine additional pre-determined characteristics, which would partially capture sources of unobserved heterogeneity. These additional characteristics are: the older sibling's gender, her exact date of birth, the number of years between the two siblings' standardized test, a dummy taking value 1 if the younger sibling is living with both parents in the year leading to the test, the household's annual pre-tax income two years prior to the younger sibling's test score, both parents' FTE adjusted annual pre-tax income two years prior to the younger sibling's test score, and both parents' annual number of work days two years prior to the younger sibling's test score. Including many controls restricts the estimation sample to almost half of its original size, which reduces the precision of estimates and statistical power in Cols. 1 and 5. Nonetheless, point estimates in Table B.15 are robust to including all these controls (except for Col. 4): our findings on test scores in Dutch and Math are virtually unchanged.

Altogether, the evidence indicates that imbalances related to school segregation are small, and results are robust to including alternative strategies to alleviate endogeneity concerns, such as balancing control variables and controlling for a host of additional pre-determined characteristics.

8.2 Specification error: Alternative Functional Forms

Another cause for concern could be that results are spuriously generated by a misspecified model. Following Elsner and Isphording (2018), Murphy and Weinhardt (Forthcoming), I consider seven alternative specifications with increasing higher order polynomials to account for older sibling national standardized test score. Panels A to E go from linearly controlling for older sibling test score to a quintic polynomial in older sibling test score; Panel F controls for dummies for each exact point of test score from 501 to 550 (our preferred specification used for all results).

Table B.16 presents estimates of the effect of child rank on younger sibling test scores across specifications, and Table B.17 effects on younger sibling tracking recommendation (conditional on younger sibling test score). Both tables show that results are strongly robust to higher order polynomials from a quadratic polynomial onward. In addition, our specification in Panel F using dummies for each point of test score is also highly consistent with results in Panels B to E using quadratic to quintic polynomials. Although results loose a bit of statistical power in Col. 4 and 5 by including dummies for each point of test score compared to the quintic polynomial, this remains my preferred specification because it allows for the most flexible control for older sibling test scores.

8.3 Inference

We have now established the robustness of my point estimates of spillovers of child rank onto younger siblings' outcomes, I further explore the robustness of my results to constructing standard errors using alternative inference procedures, i) constructing standard errors using Young (2019)'s recent randomization inference procedure, and ii) accounting for multiple hypotheses testing in standard error calculations following Romano and Wolf (2005*b*).

8.3.1 Randomization Inference

One could be concerned that uncertainty-based inference in my setting is less relevant, because I use the universe of siblings taking the standardized test between 2006 and 2016, which implies that sampling variability is less of a concern in my setting than statistical uncertainty arising from randomization in rank (Abadie et al., 2020). One could also be concerned that results are potentially driven by a few high-leverage dyads or schools from driving results. To address these concerns, I consider inference on the main results using Young (2019)'s randomization-t procedure.

I construct randomization-t based empirical p-values using a similar simulation procedure to the one used for permutation tests in Section 8.1.2. Importantly, in each simulation, I store the t-statistics of interest – the coefficient of the variable of interest divided by its cluster-robust standard error. I then construct empirical p-values based the share of permutations where simulated t-statistics are more extreme than the actual t-statistic of interest. I use 499 simulations of random assignment to school-cohort to produce randomization-t empirical p-values for the main results. Appendix Table B.18 shows that the main conclusions on the effects of child rank on younger sibling test scores and tracking recommendations hold at least at the 5% significance level, and are thus robust to using randomization-t inference for conducting inference.

8.3.2 Multiple Hypotheses Testing

Another concern could be that I test multiple hypotheses in my main results (five test scores, five tracking outcomes), which increases chance of falsely rejecting a correct null hypothesis. To address this concern, I adjust inference on my main results for multiple hypotheses testing using the Romano-Wolf step-down procedure (Romano and Wolf, 2005*a*,*b*, 2016), implemented using the `rwolf2` Stata command (Clarke, Romano and Wolf, 2019).

Based on resampling methods, this procedure controls the familywise error rate – the probability of rejecting at least one true null hypothesis across a set of hypotheses tested–by ensuring that the familywise error rate does not exceed its predetermined significance. I treat the main results as being all part of the same family of tests. One key advantage of this method is its high power, compared to previous methods which have been criticized for being too conservative (e.g. Bonferroni, 1935, Holm, 1979, Westfall and Young, 1993). Appendix Table B.18 shows that the main conclusions on the effect of child rank on their younger sibling's test scores and tracking recommendation hold at least at the 5% significance level, and are thus robust to multiple hypothesis testing adjustments to standard errors.

9 Conclusions

In this study, I propose a novel identification strategy for sibling spillover effects by exploiting exogenous variation in older siblings' rank within their school cohort in primary school. This identification strategy allows me to estimate sibling spillover effects that are driven not by direct transmission of human capital, but rather by subtle changes in attitudes and perceptions net of the direct effect of siblings' human capital. By studying the effect child rank on their younger sibling's test scores and tracking recommendation, I can explore potentially important yet understudied channels of sibling spillover effects: dynamics in sibling human capital development, parental shifts in educational investments, teacher expectations about younger siblings, and motivational effect of sibling's educational success.

I use Dutch administrative records covering the universe of pupils from primary school until university from 2006 until 2016, a unique setting in which children take a national standardized test upon exiting primary school, and in which primary school teachers spell out their subjective assessment of each pupils' ability, in the form of a tracking recommendation for secondary school. I exploit quasi-experimental variation in the percentile rank of older siblings within their primary school cohort, keeping constant their exact test score on the standardized test. That is, for older siblings with a given test score, I estimate the impact of growing up with a sibling who is top or bottom of her class on younger siblings' test scores and teacher tracking recommendation.

I find that older siblings' rank decreases younger siblings' test scores and increases the likelihood that teachers recommend the vocational track for their younger sibling, even after controlling for younger sibling's own test scores. I further show that child rank has modest but long-lasting effects on their younger siblings' educational trajectories: younger siblings enter secondary schools with higher relative ranks in their cohort, to attend lower-performing schools, and they are more likely to graduate from the Vocational track, and to graduate from high school with a STEM specialization. These results constitute the first evidence that sibling spillovers effects occur net of the direct effect of siblings' ability. This finding implies that qualitative aspects of family environments and how they interact with the school system matter a great deal for child development.

I find evidence of four underlying channels: i) older sibling's own human capital accumulation, ii) parental investments through school choice and speaking Dutch in the home, iii) teachers' expectation formation in formulating their tracking recommendation, and iv) sibling interactions, especially across gender-match. My findings on parental school choice, parental investments in the local language, and on teacher expectations are novel in the literature on sibling spillover effects in education. In addition, my findings on teacher responses are novel and important in the nascent literature on the effect of rank in education.

I further show that through teacher expectations sibling spillover effects can exacerbate social and economic disadvantage and reinforce stereotypes. Teachers are more likely to recommend the prestigious Academic track to children from privileged backgrounds (income and Western migrant background), but not children of less privileged backgrounds (Non-Western migrant backgrounds and lower income), even children from non-Western background are the only group for which child rank has positive spillover effects on younger sibling test scores. This evidence suggests that teachers may reinforce stereotypes by imperfectly learning from experience. I investigate whether this particular form of teacher bias can be neutralized through better-designed institutions. I exploit a policy change in 2014, which increased the weight of teacher assessments over test scores on the report card submitted as secondary school applications. Comparing younger siblings who took the test before v. after the academic year 2014-2015, I find that teachers became more likely to recommend the academic and general track in response to older siblings' rank. This suggests that teachers respond to child rank when incentives are strong to do so, thus demonstrating that policies can in principle exploit sibling spillover effects in order to reduce socioeconomic inequality. In particular, school admission rules moderate the effect of child rank on teacher tracking recommendations, such that school admission rules may be a useful lever to reduce socioeconomic disadvantage through teacher tracking recommendations.

That older siblings' rank affect parental investments and teachers' subjective assessments of younger siblings imply that for some disadvantaged groups, there might be returns to being top of the class in a worse school than bottom of the class in a top school. Yet, being top of the class at fixed ability implies being surrounded by lower ability peers. Thus, the positive effects of higher relative ranks may need to be carefully balanced against possible losses from exposure to fewer high-achieving peers.

Lastly, this study shows that such returns to being a big fish in a small pond are not neutral in the family; negative sibling spillovers of child rank brings a nuance to the positive conclusions of previous findings in the nascent literature on rank effects in education. A direct policy implication of this result is that sibling spillover effects might truly constitute a social multiplier. As sibling spillover effects interact with incentives in the school system, they amplify the impact of policies and their returns.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2020. "Sampling-Based versus Design-Based Uncertainty in Regression Analysis." *Econometrica*, 88(1): 265–296.

Alan, Sule, Seda Ertac, and Ipek Mumcu. 2018. "Gender stereotypes in the classroom and effects on achievement." *Review of Economics and Statistics*, 100(5): 876–890.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. 2018. "Revealing stereotypes: Evidence from immigrants in schools." National Bureau of Economic Research.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood circumstances and adult outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith. 2021. "O brother, where start thou? Sibling spillovers on college and major choice in four countries." *The Quarterly Journal of Economics*.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. "Family disadvantage and the gender gap in behavioral and educational outcomes." *American Economic Journal: Applied Economics*, 11(3): 338–81.

Black, Sandra E, Sanni Breining, David N Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen. 2021. "Sibling spillovers." *The Economic Journal*, 131(633): 101–128.

Bonferroni, Carlo E. 1935. "Il calcolo delle assicurazioni su gruppi di teste." *Studi in onore del professore salvatore ortu carboni*, 13–60.

Brenøe, Anne Ardila. 2021. "Brothers increase women's gender conformity." *Journal of Population Economics*, 1–38.

Brenøe, Anne Ardila, and Ulf Zölitz. 2020. "Exposure to more female peers widens the gender gap in stem participation." *Journal of Labor Economics*, 38(4): 000–000.

Carlana, Michela. 2019. "Implicit stereotypes: Evidence from teachers' gender bias." *The Quarterly Journal of Economics*, 134(3): 1163–1224.

Carrell, Scott E, and James E West. 2010. "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy*, 118(3): 409–432.

Clarke, Damian, Joseph P Romano, and Michael Wolf. 2019. "The Romano-Wolf Multiple Hypothesis Correction in Stata. IZA Discussion Paper Series, No. 12845." Institute for the Study of Labor (IZA).

Currie, Janet, and Douglas Almond. 2011. "Human capital development before age five." In *Handbook of labor economics*. Vol. 4, 1315–1486. Elsevier.

Dahl, Gordon B, Dan-Olof Rooth, and Anders Stenberg. 2020. "Family Spillovers in Field of Study." National Bureau of Economic Research.

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School inputs, household substitution, and test scores." *American Economic Journal: Applied Economics*, 5(2): 29–57.

de Gendre, Alexandra, and Nicolás Salamanca. 2020. "On the Mechanisms of Ability Peer Effects." IZA Discussion Paper No. 13938.

Delaney, Judith M, and Paul J Devereux. 2019. "Understanding gender differences in STEM: Evidence from college applications." *Economics of Education Review*, 72: 219–238.

Denning, Jeffrey T, Richard Murphy, and Felix Weinhardt. Forthcoming. "Class rank and long-run outcomes." *Review of Economics and Statistics*.

Dizon-Ross, Rebecca. 2019. "Parents' beliefs about their children's academic ability: Implications for educational investments." *American Economic Review*, 109(8): 2728–65.

Elsner, Benjamin, and Ingo E Isphording. 2017. "A big fish in a small pond: Ability rank and human capital investment." *Journal of Labor Economics*, 35(3): 787–828.

Elsner, Benjamin, and Ingo E Isphording. 2018. "Rank, sex, drugs, and crime." *Journal of Human Resources*, 53(2): 356–381.

Elsner, Benjamin, Ingo E Isphording, and Ulf Zölitz. Forthcoming. "Achievement rank affects performance and major choices in college." *The Economic Journal*.

Feld, Jan, and Ulf Zölitz. 2017. "Understanding peer effects: On the nature, estimation, and channels of peer effects." *Journal of Labor Economics*, 35(2): 387–428.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2016. "Parental responses to public investments in children: Evidence from a maximum class size rule." *Journal of Human Resources*, 51(4): 832–868.

Goulas, Sofoklis, Silvia Griselda, and Rigissa Megalokonomou. 2020. "Comparative Advantage and Gender Gap in STEM." IZA Discussion Paper No. 13313.

Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science*, 312(5782): 1900–1902.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics*, 24(3): 411–482.

Holm, Sture. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian journal of statistics*, 65–70.

Hoxby, Caroline. 2000. "Peer effects in the classroom: Learning from gender and race variation." National Bureau of Economic Research.

Hoxby, Caroline M, and Gretchen Weingarth. 2005. "Taking race out of the equation: School reassignment and the structure of peer effects." Citeseer.

Joensen, Juanna Schrøter, and Helena Skyt-Nielsen. 2018. "Spillovers in education choice." *Journal of Public Economics*, 157: 158–183.

Karbownik, Krzysztof, and Umut Özek. Forthcoming. "Setting a Good Example? Examining Sibling Spillovers in Educational Achievement Using a Regression Discontinuity Design." *Journal of Human Resources*.

Kiessling, Lukas, and Jonathan Norris. 2020. "The long-run effects of peers on mental health." *MPI Collective Goods Discussion Paper*, , (2020/12).

Landersø, Rasmus Kløve, Helena Skyt-Nielsen, and Marianne Simonsen. 2019. "Effects of school starting age on the family." *Journal of Human Resources*, 1117–9174R1.

Lavy, Victor. 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of public Economics*, 92(10-11): 2083–2105.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and impacts of gender peer effects at school." *American Economic Journal: Applied Economics*, 3(2): 1–33.

Lavy, Victor, and Rigissa Megalokonomou. 2019. "Persistency in teachers' grading bias and effects on longer-term outcomes: University admissions exams and choice of field of study." National Bureau of Economic Research.

Lavy, Victor, M Daniele Paserman, and Analia Schlosser. 2012. "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom." *The Economic Journal*, 122(559): 208–237.

Lim, Jaegeum, and Jonathan Meer. 2017. "The impact of teacher–student gender matches random assignment evidence from South Korea." *Journal of Human Resources*, 52(4): 979–997.

Lim, Jaegeum, and Jonathan Meer. 2020. "Persistent Effects of Teacher–Student Gender Matches." *Journal of Human Resources*, 55(3): 809–835.

Manski, Charles F. 1993*a*. "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60(3): 531–542.

Manski, Charles F. 1993*b*. "Identification problems in the social sciences." *Sociological Methodology*, 1–56.

Manski, Charles F. 2000. "Economic analysis of social interactions." *Journal of economic perspectives*, 14(3): 115–136.

Murphy, Richard, and Felix Weinhardt. Forthcoming. "Top of the Class: The Importance of Ordinal Rank." *The Review of Economic Studies*.

Nicoletti, Cheti, and Birgitta Rabe. 2019. "Sibling spillover effects in school achievement." *Journal of Applied Econometrics*, 34(4): 482–501.

Papageorge, Nicholas W, Seth Gershenson, and Kyung Min Kang. 2020. "Teacher expectations matter." *Review of Economics and Statistics*, 102(2): 234–251.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. "Family Spillover Effects of Marginal Diagnoses: The Case of ADHD." National Bureau of Economic Research.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a better school: Effects and behavioral responses." *American Economic Review*, 103(4): 1289–1324.

Qureshi, Javaeria A. 2018*a*. "Additional returns to investing in girls' education: Impact on younger sibling human capital." *The Economic Journal*, 128(616): 3285–3319.

Qureshi, Javaeria A. 2018*b*. "Siblings, teachers, and spillovers on academic achievement." *Journal of Human Resources*, 53(1): 272–297.

Romano, Joseph P, and Michael Wolf. 2005*a*. "Exact and approximate stepdown methods for multiple hypothesis testing." *Journal of the American Statistical Association*, 100(469): 94–108.

Romano, Joseph P, and Michael Wolf. 2005*b*. “Stepwise multiple testing as formalized data snooping.” *Econometrica*, 73(4): 1237–1282.

Romano, Joseph P, and Michael Wolf. 2016. “Efficient computation of adjusted p-values for resampling-based stepdown multiple testing.” *Statistics & Probability Letters*, 113: 38–40.

Westfall, Peter H, and S Stanley Young. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment.* Vol. 279, John Wiley & Sons.

Young, Alwyn. 2019. “Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results.” *The Quarterly Journal of Economics*, 134(2): 557–598.

Zölitz, Ulf, and Jan Feld. Forthcoming. “The Effect of Peer Gender on Major Choice in Business School.” *Management Science*.

Appendix A Data Preparation

In this section, I complement Section 3 to explain how I construct my estimation sample, going from registry data to a dyadic sibling pair dataset. For confidentiality reasons, I coarsely round all numbers presented here. The unit of observation in my estimation sample is a dyad, linking every child to her closest older sibling. I first construct a dataset containing for each pupil her rank among her primary school classmates, along with information at the school-cohort level such as cohort size for each subject and for the complete standardized test. I start by appending yearly registries into a dataset of repeated cross-sections of pupils taking the standardized test. Few pupils repeat the test; I reshape the data into wide format and focus only on the first test, as I will later exclude test repeaters from my estimation sample. I keep these pupils for now, as I am interested in children's rank in their cohort, including classmates who are retaking the test. I set this dataset aside for later merging.

To construct a dataset of sibling which I later reshape into dyadic sibling pair format, I start from the national population registry, which records every person residing in the Netherlands. I keep only persons born since 1980, either born in the Netherlands or abroad; at this stage I have roughly 16.2 million persons. I merge in parental identifiers using family records, which yields 7.5 million matches. I construct a mother identifier, to mark all children recorded with the same mother; only 30,000 children have no recorded mother.

I then merge education test taking registry information, and the dataset containing rank and school-cohort information. This match yields 1.6 million matched observations; non-matched observations are mostly children residing in the Netherlands with no test taking information (5.7 million) and few are children with test taking information but not recorded as residents of the Netherlands (13,000 children). As records for the standardized test start in 2006, most non-matched children are simply born too early.

At this stage, I make the following sample selection decisions that will later allow me to reshape the data into a dyadic sibling pair dataset: I drop all children in 1) families where no child ever took the test (4.7 million children), 2) families with at least once two children taking the test the same year (77,000 children), and 3) families with at least one test repeater child (12,000 children). This amounts to dropping 4.8 million unique child observations, which leaves me with a remaining sample of 2.6 million children, including 955,846 children with no test taking information.

I then construct sibling pair identifiers. I follow Karbownik and Özek (Forthcoming) and Black et al. (2021) and define a sibling as a child born to the same mother. For 87% of children in the Netherlands, this also means living in the same residence. Importantly, unlike Karbownik and Özek (Forthcoming) and Black et al. (2021) I do not construct sibling pairs based on birth order, but rather on the timing of test taking. This is to prevent cases such as older siblings repeating a class and taking their test after their adjacent younger sibling, or having to choose between twin siblings which would be the treatment sibling for their next younger sibling. Instead, I assign to every child a pair identifier that is only shared by the oldest sibling who last took the standardized test, and who has not taken his first attempt on the standardized test at the same time as another sibling.

A child can be a younger child for one pair with pair identifier 1, and also be an older child for another pair with pair identifier 2. To allow this, I reshape the data into dyadic format, and I expand observations for children simultaneously belonging to two pairs. I then reshape the data to obtain one observation per sibling pair, in which the older sibling is always sibling 1 and the younger sibling is always sibling 2. In this process, I am able to construct 531,780 unique pairs, for 975,256 unique children (1,063,560 observations after including expanded observations). These pairs constitute the baseline estimation sample.

Before reshaping the data into a dyadic format, I merge additional information in wide format for each child corresponding to the year immediately preceding the primary school exit test: 1) maternal and paternal employment status, working hours, earned income and taxable income, 2) maternal and paternal residence, 3) later life education records such as secondary school attendance, track and major, graduation, time to graduation, and post-secondary education attendance, and major. After merging this information to every single child in the sibling pairs dataset, I reshape the data into dyadic format and construct final variables which required information for both siblings of a pair to be constructed, such as age difference, same gender, and likelihood of having had the same primary school teacher.

Appendix B Additional Tables and Figures

Table B.1: *The Effects of Own-Rank on Human Capital Accumulation in the Long-Run*

	Treatment: Own percentile rank [std]				
	(1)	(2)	(3)	(4)	(5)
Outcomes:	Coef. Est.	Std. Err.	Y Mean	R^2	N. observations
Panel A. Primary school tracking recommendation (in percentage point):					
Academic	−0.314	0.199	12.24	0.48	527,448
General	1.012***	0.187	11.19	0.25	527,448
Vocational	−2.403***	0.243	27.28	0.57	527,448
Over-recommendation	2.877***	0.242	51.97	0.54	527,448
Under-recommendation	−2.234***	0.212	20.93	0.38	527,448
Panel B. Secondary school (in percentage point unless otherwise stated):					
Cohort avg. test score [std.]	−0.015	0.035	0.00	0.61	526,910
Share high-achieving peers	0.269**	0.112	21.43	0.51	526,910
Start Academic track	0.385*	0.219	31.31	0.62	350,926
Start General track	1.041***	0.282	24.39	0.29	350,926
Start Vocational track	−1.426***	0.233	44.30	0.66	350,926
Graduate Academic track	−0.170	0.226	25.94	0.54	350,926
Graduate General track	1.192***	0.307	31.50	0.27	350,926
Graduate Vocational track	−1.022***	0.244	42.56	0.59	350,926
Graduate STEM profile	0.550**	0.237	15.30	0.25	350,926
Graduate on time	−0.369	0.378	60.61	0.19	288,758
Panel C. Post-secondary education (in percentage point):					
Ever in post-sec. edu.	−0.216	0.172	94.36	0.26	288,758
STEM major	0.293	0.514	18.34	0.16	112,338

*This table shows the results of regression analyses of the effect of older sibling own percentile rank on their own later outcomes. Each row represents one regression. Regressions include school-by-cohort fixed effects and control non-parametrically for own standardized test score, and dummies for whether the child is native-born, 1st generation or 2nd generation migrant. The estimation sample contains 527,448 dyads, 6,627 schools, and 53,247 school-by-cohort fixed effects. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 6.1.*

Table B.2: *Marginal Effects of Older Sibling Rank on on Younger Sibling Test Scores by Sibling Gender Match*

Outcome:		Younger Sibling Test Scores [std.]				
		(1)	(2)	(3)	(4)	(5)
		Overall	Dutch	Math	Acad. Prep.	Geography
Marginal Effects of Percentile Rank [std.] by:						
Old sibling:	*Young sibling:*					
Male	Male	−0.010*	−0.003	−0.034***	−0.010	−0.010
		(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Male	Female	−0.019***	−0.015***	−0.040***	−0.019***	−0.012*
		(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Female	Male	−0.028***	−0.042***	−0.023***	−0.023***	−0.026***
		(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Female	Female	−0.003	−0.025***	−0.005	−0.000	0.001
		(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
R^2		0.30	0.41	0.36	0.29	0.34
N. schools		6,627	6,627	6,627	6,627	6,062
N. schools younger sibling		10,820	10,820	10,820	10,820	8,141
N. clusters		53,247	53,247	53,247	53,247	43,416
N. observations		527,448	527,448	527,448	527,448	406,187

*This table shows marginal effects by sibling gender match from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores. Each column represents one regression, and rows represent sibling gender match. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score, dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant and younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.2.*

Table B.3: *The Effects of Older Sibling Rank on Parental Investments*

Outcome:	Choice of Primary School			Parental Investments			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Same school (in p.p.)	School mean [std.]	Cohort size	Speak Dutch (in p.p.)	Early	On time (in p.p.)	Late
Treatment variable:							
Percentile rank [std.]	-0.565^{***}	-0.001	-0.002	-0.354	-0.356^{**}	0.733^{***}	-0.377^{*}
	(0.197)	(0.005)	(0.044)	(0.228)	(0.149)	(0.240)	(0.201)
Outcome Mean	82.0	0.04	25.76	92.0	7.54	78.93	13.52
Outcome S.D.	39.0	0.98	15.92	27.0	26.41	40.78	34.20
R^2	0.35	0.63	0.78	0.58	0.16	0.13	0.15
N. schools	6,627	6,343	6,627	5,929	6,627	6,627	6,627
N. schools y. sibling	10,820	6,227	10,820	5,963	10,820	10,820	10,820
N. clusters	53,247	47,921	53,247	22,153	53,247	53,247	53,247
N. observations	527,448	455,224	527,448	165,456	527,448	527,448	527,448

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on parental investments for the younger sibling. Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. School mean [std.] (Col.2) is the average test score in the primary school in the previous cohort. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.3.*

Table B.4: *Marginal Effects of Older Sibling Rank on Parental Investments by Migrant Background*

Outcome:	Choice of Primary School			Parental Investments			
	(1) Same school (in p.p.)	(2) School mean [std.]	(3) Cohort size	(4) Speak Dutch (in p.p.)	(5) Early	(6) On time (in p.p.)	(7) Late
Marginal Effects of Percentile Rank [std.] by:							
Native-born	−0.813***	−0.004	−0.019	−1.248***	−0.394***	0.349	0.045
	(0.198)	(0.005)	(0.044)	(0.217)	(0.152)	(0.243)	(0.202)
Non-Western ethnicity	0.292	0.012**	0.083	3.727***	−0.249	2.107***	−1.857***
	(0.246)	(0.006)	(0.053)	(0.393)	(0.170)	(0.286)	(0.245)
Western ethnicity	−0.125	−0.003	−0.030	1.066**	−0.116	1.272***	−1.156***
	(0.302)	(0.007)	(0.069)	(0.486)	(0.227)	(0.227)	(0.300)
Outcome Mean	82.0	0.04	25.76	92.0	7.54	78.93	13.52
Outcome S.D.	39.0	0.98	15.92	27.0	26.41	40.78	34.20
R^2	0.35	0.63	0.78	0.60	0.16	0.13	0.15
N. schools	6,627	6,343	6,627	5,929	6,627	6,627	6,627
N. schools younger sibling	10,820	6,227	10,820	5,963	10,820	10,820	10,820
N. clusters	53,247	47,921	53,247	22,153	53,247	53,247	53,247
N. observations	527,448	455,224	527,448	165,456	527,448	527,448	527,448

*This table shows marginal effects by migrant background from dyadic regressions of the effect of percentile rank in primary school cohort on parental investments for the younger sibling. Each column represents one regression, and rows corresponds to younger sibling's ethnic background (native, Non-Western or Western ethnicity). Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. School mean [std.] (Col.2) is the average test score in the primary school in the previous cohort. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.3.*

Table B.5: *Marginal Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations by Pre- and Post-2014 Reform*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Marginal Effects of Percentile Rank [std.] by:					
Pre reform	−0.115	−0.284	0.517**	0.114	−0.232
	(0.173)	(0.187)	(0.221)	(0.221)	(0.191)
Post reform	1.411***	0.234	0.027	−1.169***	−0.504**
	(0.191)	(0.207)	(0.245)	(0.238)	(0.212)
Outcome Mean	11.80	12.11	30.43	18.89	26.78
R^2	0.32	0.16	0.47	0.21	0.53
N. schools	6,416	6,416	6,416	6,416	6,416
N. schools y. sibling	10,582	10,582	10,582	10,582	10,582
N. clusters	47,473	47,473	47,473	47,473	47,473
N. observations	497,801	497,801	497,801	497,801	497,801

*This table shows marginal effects by pre- and post-2014 reform from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each column represents one regression, and rows corresponds to the timing of younger sibling's test score - before v. after the 2014 reform. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score, dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant, and younger sibling's own test score. The sample is restricted to pairs where older siblings took the CITO test before 2014. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.4.*

Table B.6: *Marginal Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations by Migrant Background*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Marginal Effects of Percentile Rank [std.] by:					
Native-born	0.008	−0.130	0.547**	−0.196	−0.229
	(0.172)	(0.186)	(0.217)	(0.218)	(0.187)
Non-Western ethnicity	0.187	0.129	0.037	0.243	−0.596***
	(0.188)	(0.208)	(0.259)	(0.246)	(0.218)
Western ethnicity	1.218***	−0.871***	0.337	−0.191	−0.494*
	(0.258)	(0.269)	(0.307)	(0.317)	(0.279)
Outcome Mean	12.17	12.25	30.30	18.78	26.50
R^2	0.33	0.17	0.47	0.21	0.54
N. schools	6,627	6,627	6,627	6,627	6,627
N. schools younger sibling	10,820	10,820	10,820	10,820	10,820
N. clusters	53,247	53,247	53,247	53,247	53,247
N. observations	527,448	527,448	527,448	527,448	527,448

*This table shows marginal effects by migrant background from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each column represents one regression, and rows corresponds to younger sibling's ethnic background (native, Non-Western or Western ethnicity). Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.4.*

Table B.7: *Marginal Effects of Older Sibling Rank on Younger Sibling Test Scores by Migrant Background*

Outcome:	Younger Sibling Test Scores [std.]				
	(1) Overall	(2) Dutch	(3) Math	(4) Acad. Prep.	(5) Geography
Marginal Effects of Percentile Rank [std.] by:					
Native-born	-0.027^{***}	-0.03^{***}	-0.034^{***}	-0.024^{***}	-0.022^{***}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Non-Western ethnicity	0.031^{***}	0.015^{**}	0.004	0.031^{***}	0.034^{***}
	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)
Western ethnicity	-0.001	-0.011	-0.014^{*}	-0.004	0.015^{*}
	(0.008)	(0.007)	(0.007)	(0.009)	(0.009)
R^2	0.30	0.41	0.35	0.29	0.32
N. schools	6,627	6,627	6,627	6,236	6,062
N. schools younger sibling	10,820	10,820	10,820	6,335	8,141
N. clusters	53,247	53,247	53,247	38,937	43,416
N. observations	527,448	527,448	527,448	353,818	406,187

*This table shows marginal effects by migrant background from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores in primary school (overall, Dutch, Math, Academic Preparedness and Geography). Each column represents one regression, and rows corresponds to younger sibling's ethnic background (native, Non-Western or Western ethnicity). Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.4.*

Table B.8: *Marginal Effects of Older Sibling Rank on on Younger Sibling Tracking Recommendations by Gender*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Marginal Effects of Percentile Rank [std.] by:					
Male	0.163	−0.151	0.497**	−0.170	−0.339*
	(0.174)	(0.189)	(0.221)	(0.221)	(0.190)
Female	0.033	−0.084	0.390*	−0.044	−0.295
	(0.174)	(0.188)	(0.221)	(0.221)	(0.191)
Outcome Mean	12.17	12.25	30.30	18.78	26.50
R^2	0.33	0.17	0.47	0.21	0.54
N. schools	6,627	6,627	6,627	6,627	6,627
N. schools younger sibling	10,820	10,820	10,820	10,820	10,820
N. clusters	53,247	53,247	53,247	53,247	53,247
N. observations	527,448	527,448	527,448	527,448	527,448

*This table shows marginal effects by younger sibling gender from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each column represents one regression, and rows represent younger sibling gender. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score, dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant and younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 6.4.*

Table B.9: *Marginal Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations by Sibling Gender Match*

Outcome:		Younger Sibling Tracking Recommendation (in percentage point)				
		(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Marginal Effects of Percentile Rank [std.] by:						
Old sibling:	*Young sibling:*					
Male	Male	0.242	−0.143	0.432*	−0.177	−0.355*
		(0.184)	(0.199)	(0.233)	(0.234)	(0.201)
Male	Female	0.088	−0.160	0.557**	−0.169	−0.316
		(0.184)	(0.199)	(0.233)	(0.233)	(0.202)
Female	Male	−0.040	0.042	0.505**	−0.158	−0.349*
		(0.183)	(0.198)	(0.233)	(0.233)	(0.203)
Female	Female	0.108	−0.211	0.277	0.073	−0.246
		(0.184)	(0.199)	(0.233)	(0.233)	(0.201)
Outcome Mean		12.17	12.25	30.30	18.78	26.50
R^2		0.33	0.17	0.47	0.21	0.54
N. schools		6,627	6,627	6,627	6,627	6,627
N. schools y. sibling		10,820	10,820	10,820	10,820	10,820
N. clusters		53,247	53,247	53,247	53,247	53,247
N. observations		527,448	527,448	527,448	527,448	527,448

*This table shows marginal effects by sibling gender match from dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each column represents one regression, and rows represent sibling gender match. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score, dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant and younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 6.4.*

Table B.10: *Rank-Based Sorting I: The Effects of Parental Characteristics on Older Sibling Rank*

	(1)	(2)	(3)	(4)	(5)
	Outcome: Percentile rank [std.]				
Parental pre-treatment characteristics:					
Mother's annual pre-tax income	0.0004*				
(per 10,000 Euros)	(0.0002)				
Father's annual pre-tax income		−0.0000			
(per 10,000 Euros)		(0.0001)			
Mother's annual working days			0.0018***		
(per 100 days)			(0.0006)		
Father's annual working days				−0.0005	
(per 100 days)				(0.0008)	
Household annual pre-tax income					−0.0000
(per 10,000 Euros)					(0.0001)
R^2	0.95	0.95	0.95	0.95	0.94
Nschools	6,452	6,545	6,452	6,545	6,627
Nclusters	47,736	49,774	47,736	49,774	53,247
Nobs	346,841	408,265	346,841	408,265	527,448

*This table shows the results of regression analyses of the effect of parental characteristics on (older sibling) percentile rank in primary school. Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 8.1.1.*

Table B.11: *Rank-Based Sorting II: The Effects of Child Characteristics on Own Rank*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Outcome: Percentile rank [std.]						
Parental pre-treatment characteristics:							
Female	−0.001						
	(0.001)						
Date of birth		−0.000					
		(0.000)					
Native-born			-0.010^{***}				
			(0.001)				
Living with father				−0.001			
				(0.001)			
Living with mother					0.002		
					(0.003)		
Number siblings						0.001	
						(0.001)	
Living in 4 largest cities							0.008
							(0.006)
R^2	0.94	0.94	0.94	0.94	0.94	0.94	0.94
Nschools	6,627	6,627	6,627	6,627	6,627	6,627	6,618
Nclusters	53,247	53,247	53,247	53,247	53,247	53,247	53,156
Nobs	527,448	527,448	527,448	527,448	527,448	527,448	526,680

*This table shows the results of regression analyses of the effect of older sibling characteristics on own percentile rank in primary school. Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 8.1.1.*

Table B.12: *Permutation-Based Sorting Tests in Estimation Sample*

	Share of permutations with empirical p-val. under			Avg.
	0.10	0.05	0.01	p-value
	(1)	(2)	(3)	(4)
Younger siblings pre-treatment characteristics:				
Female	0.099	0.053	0.011	0.513
Date of birth	0.066	0.041	0.004	0.207
Difference in years between tests	0.072	0.021	0.002	0.458
Two-parent household	0.093	0.047	0.007	0.284
Household annual pre-tax income	0.088	0.038	0.003	0.995
Mother's annual pre-tax income	0.083	0.038	0.005	0.515
Father's annual pre-tax income	0.072	0.029	0.004	0.216
Mother's annual working days	0.081	0.022	0.003	0.075
Father's annual working days	0.068	0.030	0.005	0.133
Migration background):				
Native born	0.079	0.031	0.005	0.000
First-generation migrant	0.071	0.031	0.009	0.917
Second-generation migrant	0.087	0.035	0.004	0.000

Note: This table shows the results of permutation-based sorting tests, in our estimation sample containing 6,627 schools, 53,247 older siblings school-cohorts (clusters) and 527,448 younger siblings. For these tests, we simulate 1,000 reassignments to older sibling rank under the null of random assignment, keeping older sibling test score constant. We then regress younger sibling pre-treatment characteristics on older sibling synthetic rank with fixed effects for older sibling synthetic school-cohort, and construct empirical p-values as the share of times that the correlation of synthetic rank with pre-treatment characteristics were more extreme than the correlation of actual rank with pre-treatment characteristics. Each row presents empirical p-values for a different pre-assignment characteristic. The last column shows the average p-value for all school cohorts. *Back to Section 8.1.2.*

Table B.13: *The Effects of Older Sibling Rank on Younger Sibling Test Scores - Balancing Controls*

Outcome:	Younger Sibling Test Scores [std.]				
	(1) Overall	(2) Dutch	(3) Math	(4) Acad. Prep.	(5) Geography
Treatment variable:					
Panel A. Model without Balancing Control Variables					
Percentile rank [std.]	-0.016^{***}	-0.021^{***}	-0.026^{***}	-0.016^{**}	-0.015^{**}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Panel B. Model with Balancing Control Variables (Preferred Specification)					
Percentile rank [std.]	-0.015^{***}	-0.021^{***}	-0.026^{***}	-0.014^{**}	-0.013^{**}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
R^2	0.30	0.40	0.35	0.29	0.32
Nschools	6,627	6,627	6,627	6,236	6,062
Nschools younger sibling	10,820	10,820	10,820	6,335	8,141
Nclusters	53,247	53,247	53,247	38,937	43,416
Nobs	527,448	527,448	527,448	353,818	406,187

This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores in primary school (overall, Dutch, Math, Academic Preparedness and Geography). Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Panel B adds dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant as balancing controls. Panel B (in ***bold****) contains the preferred specification presented in Table 3. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 8.1.3.*

Table B.14: *The Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations - Balancing Controls*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1)	(2)	(3)	(4)	(5)
	Academic	General	Vocational	Mixed	Missing
Treatment variable:					
Panel A. Baseline Model without Balancing Control Variables					
Percentile rank [std.]	−0.108	−0.201	0.849***	−0.226	−0.313*
	(0.180)	(0.185)	(0.249)	(0.170)	(0.185)
Panel B. Baseline Model with Balancing Control Variables					
Percentile rank [std.]	−0.090	−0.190	0.810***	−0.218	−0.306*
	(0.180)	(0.190)	(0.250)	(0.219)	(0.185)
Panel C. Model with Younger Sibling Test Score Controls, No Balancing Controls					
Percentile rank [std.]	0.100	−0.120	0.450**	−0.107	−0.324*
	(0.170)	(0.180)	(0.210)	(0.215)	(0.185)
Panel D. Model with Younger Sibling Test Score controls and Balancing Controls					
Percentile rank [std.]	0.100	−0.120	0.440**	−0.107	−0.317*
	(0.170)	(0.180)	(0.210)	(0.215)	(0.185)
Outcome Mean	12.17	12.25	30.30	18.78	26.50
R^2	0.33	0.16	0.47	0.21	0.54
Nschools	6,627	6,627	6,627	6,627	6,627
Nschools younger sibling	10,820	10,820	10,820	10,820	10,820
Nclusters	53,247	53,247	53,247	53,247	53,247
Nobs	527,448	527,448	527,448	527,448	527,448

This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school. Each row and column represents one regression. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Panel A presents the baseline model which do not control for younger sibling test scores, nor balancing controls. Panel B presents models including balancing controls for dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant; Panel C presents results for models controlling for younger sibling test score but not balancing controls, and Panel D presents models including both younger sibling test score controls and balancing controls. Panels B and D (in ***bold****) are the preferred specification presented in Table 4. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 8.1.3.*

Table B.15: *The Effects of Older Sibling Rank on Younger Sibling Test Scores - Balancing Controls*

Outcome:	Younger Sibling Test Scores [std.]				
	(1) Overall	(2) Dutch	(3) Math	(4) Acad. Prep.	(5) Geography
Treatment variable:					
Panel A. Model with Balancing Control Variables (Preferred Specification)					
Percentile rank [std.]	-0.015^{***}	-0.021^{***}	-0.026^{***}	-0.014^{**}	-0.013^{**}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
R2	0.30	0.40	0.35	0.29	0.32
Nschools	6,627	6,627	6,627	6,236	6,062
Nschools younger sibling	10,820	10,820	10,820	6,335	8,141
Nclusters	53,247	53,247	53,247	38,937	43,416
Nobs	527,448	527,448	527,448	353,818	406,187
Panel B. Model with Many Control Variables					
Percentile rank [std.]	-0.014^{*}	-0.022^{***}	-0.028^{***}	-0.006	-0.013
	(0.007)	(0.007)	(0.007)	(0.009)	(0.008)
R^2	0.37	0.49	0.44	0.34	0.38
Nschools	6,360	6,360	6,360	5,970	5,565
Nschools younger sibling	10,167	10,167	10,167	6,078	7,685
Nclusters	44,538	44,538	44,538	32,071	35,725
Nobs	291,392	291,392	291,392	196,931	228,918

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores in primary school (overall, Dutch, Math, Academic Preparedness and Geography). Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. Panel B adds dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant as balancing controls. Panel B (in **bold**) contains the preferred specification presented in Table 3. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 8.1.3.*

Table B.16: *The Effects of Older Sibling Rank on Younger Sibling Test Scores Across Ability Specification*

Outcome:	Younger Sibling Test Scores [std.]				
	(1) Overall	(2) Dutch	(3) Math	(4) Acad. Prep.	(5) Geography
Treatment variable:					
Panel A. Linear control					
Percentile rank [std.]	0.096***	0.059***	0.056***	0.075***	0.081***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.005)
Panel B. Quadratic polynomial control					
Percentile rank [std.]	−0.024***	−0.029***	−0.031***	−0.020***	−0.023***
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Panel C. Cubic polynomial control					
Percentile rank [std.]	−0.020***	−0.026***	−0.029***	−0.017**	−0.019***
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Panel D. Quartic polynomial control					
Percentile rank [std.]	−0.020***	−0.026***	−0.029***	−0.017**	−0.019***
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Panel E. Quintic polynomial control					
Percentile rank [std.]	−0.020***	−0.026***	−0.029***	−0.017**	−0.019***
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
Panel F. Test Score Point fixed effects (Preferred Specification)					
Percentile rank [std.]	−0.015***	−0.021***	−0.026***	−0.014*	−0.013*
	(0.006)	(0.005)	(0.005)	(0.007)	(0.006)
R^2	0.31	0.41	0.36	0.30	0.33
Nschools	6,813	6,813	6,813	6,397	6,520
Nschools younger sibling	10,891	10,891	10,891	6,376	8,223
Nclusters	57,516	57,516	57,516	42,713	49,427
Nobs	531,717	531,717	531,717	357,594	412,198

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's standardized test scores in primary school (overall, Dutch, Math, Academic Preparedness and Geography) for seven different ways to control for older sibling ability, captured by test scores. Each row and column represents one regression. Regressions include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score and dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively. Back to Section 8.2.*

Table B.17: *The Effects of Older Sibling Rank on Younger Sibling Tracking Recommendations Across Ability Specification*

Outcome:	Younger Sibling Tracking Recommendation (in percentage point)				
	(1) Academic	(2) General	(3) Vocational	(4) Mixed	(5) Missing
Treatment variable:					
Panel A. Linear control					
Percentile rank [std.]	4.916***	−2.043***	−0.779***	−2.181***	0.087
	(0.141)	(0.149)	(0.181)	(0.176)	(0.156)
Panel B. Quadratic polynomial control					
Percentile rank [std.]	−0.573***	0.405**	0.045	0.424*	−0.301
	(0.179)	(0.193)	(0.226)	(0.226)	(0.194)
Panel C. Cubic polynomial control					
Percentile rank [std.]	−0.222	0.143	0.252	0.150	−0.323*
	(0.177)	(0.193)	(0.226)	(0.226)	(0.195)
Panel D. Quartic polynomial control					
Percentile rank [std.]	−0.222	0.143	0.252	0.150	−0.323*
	(0.177)	(0.193)	(0.226)	(0.226)	(0.195)
Panel E. Quintic polynomial control					
Percentile rank [std.]	−0.222	0.143	0.252	0.150	−0.323*
	(0.177)	(0.193)	(0.226)	(0.226)	(0.195)
Panel F. Test Score Point fixed effects (Preferred Specification)					
Percentile rank [std.]	0.097	−0.117	0.444*	−0.107	−0.317
	(0.179)	(0.194)	(0.228)	(0.228)	(0.196)
Outcome Mean	12.17	12.25	30.34	18.78	26.46
R^2	0.33	0.17	0.48	0.21	0.54
Nschools	6,813	6,813	6,813	6,813	6,813
Nschools younger sibling	10,891	10,891	10,891	10,891	10,891
Nclusters	57,516	57,516	57,516	57,516	57,516
Nobs	531,717	531,717	531,717	531,717	531,717

*This table shows the results of dyadic regression analyses of the effect of percentile rank in primary school cohort on younger sibling's tracking recommendation upon leaving primary school, for seven different ways to control for older sibling ability, measured by test scores. Each row and column represents one regression. All models include older sibling school-by-cohort fixed effects and control non-parametrically for (older sibling) standardized test score. All also control for dummies for whether the younger sibling is native, 1st generation or 2nd generation migrant, and for younger sibling test score. Standard errors are clustered at the level of the older sibling's school-by-cohort. *, ** and *** denote significance levels at the 10%, 5%, and 1% respectively.* *Back to Section 8.2.*

Table B.18: *Corrected P-values for the Effect of Class Rank on Younger Sibling Outcomes using Young (2019)'s Randomization Inference, and Romano and Wolf (2005b)'s Step-Down Familywise Error Rate Adjustment Procedures*

	Corrected p-values for the effect of older sibling rank [std] using:		
	Young (2019)	Romano and Wolf (2005*b*)	Holm (1979)
	Randomization-t inference	Step-down procedure	
Outcomes:	(1)	(2)	(3)
Test scores:			
Overall	**0.004**	**0.003**	**0.016**
Dutch	**0.000**	**0.001**	**0.009**
Math	**0.000**	**0.001**	**0.010**
Academic Preparedness	**0.025**	**0.025**	**0.028**
Geography	*0.035*	**0.025**	**0.030**
Tracking recommendation:			
Academic	0.584	0.717	0.831
General	0.305	0.717	1.000
Vocational	**0.000**	**0.025**	**0.025**
Mixed	0.286	0.717	0.463
Missing	0.101	*0.066*	*0.068*
Number of replications	1,000	1,000	1,000

This table presents corrected p-values for the main results using i) Young (2019)'s randomization-t inference procedure to account for high-leverage and finite sample properties of the model error term (Col. (1) based on 1,000 permutations), ii) Romano-Wolf step-down procedure for controlling for familywise error rate in multiple hypotheses testing (Romano and Wolf, 2005a,b, 2016) (Col. 2), and iii) Holm (1979)'s correction for multiple hypothesis testing (Col. 3). Both Romano and Wolf (2005a,b, 2016) and Holm (1979) procedures are implemented using Clarke, Romano and Wolf (2019) `rwolf2` Stata package, based on 1,000 replications. p-values smaller than 0.10 are shown in italics and smaller than 0.05 in bold. Back to Section 8.3.