A Work Project, presented as part of the requirements for the Award of a Master's degree in Economics from the NOVA School of Business and Economics.

SIBLING SPILLOVERS IN EDUCATIONAL ACHIEVEMENT: EVIDENCE FROM TANZANIA

JOÃO PEDRO RAMOS FERREIRA

#39188

Work project carried out under the supervision of:

Professor Wayne Sandholtz

13-01-2023

Sibling Spillovers in Educational Achievement: Evidence from Tanzania*

João Ferreira

January 13, 2023

Abstract

Sibling spillovers have been increasingly documented in several domains, but the effect of older siblings' educational achievement on one's academic success has received less attention in applied research. Evidence from developing countries is particularly scarce. I use multiyear data on national exam scores from Tanzania to estimate sibling spillover effects from passing the PSLE, a high-stakes test that determines eligibility to enroll in government-run secondary schools. Exploiting this *discontinuity* in exposure to schooling opportunities around the passing score, I find no conclusive evidence of sibling spillovers. This result is robust to the consideration of possible heterogeneity by sibling gender match.

JEL: D10, I20, J13
Keywords: Sibling spillovers, sibling-dependency in education, peer effects, educational achievement

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

*I am deeply thankful to Professor Wayne Sandholtz for his patience and guidance throughout the project. I am also deeply grateful for the positive (and undoubtedly significant) spillovers from friends and family throughout the last four and a half years.

1 Introduction

Parenting decisions have long been of interest for social scientists, due to their consequential effects on one's offspring.[1] Hence, parental inputs are featured in theoretical models of skill formation and human capital accumulation (e.g., Cunha and Heckman 2007), while empirical analyses grounded on those frameworks have shed light on how parental behavior affects child development, later-life outcomes, and even social mobility (Heckman and Mosso 2014). However, sibling interactions and their effects were generally understudied until the last decade. Considering the extent to which the presence of siblings in a child's household may affect their personal development, any evaluation of educational policy that fails to account for sibling spillovers may not fully capture the social benefits and costs arising from the intervention.

Sibling spillovers have been identified in multiple settings.[2] The literature on education has mostly focused on choice spillovers, or the causal impact of older siblings on one's educational choices. Joensen and Nielsen (2018), for instance, found that older siblings' social influence can affect younger siblings' course choices in high school, particularly among closely-spaced brothers. Moreover, a growing body of research on sibling spillovers with respect to college or major choice has recently emerged (see, for instance, Goodman et al. 2015; Aguirre and Matta 2021; Altmejd et al. 2021). However, sibling spillovers in educational achievement have seldom been analysed;[3] studies that use evidence from developing countries to answer this research question are especially scarce. Correlated unobservables, the reflection problem (Manski 1993), and simultaneity in a household's schooling decisions[4] have been noted as persistent obstacles to further advancements in this literature (Lindskog 2013).

[1] See, for a review, Doepke et al. (2019).

[2] Sibling spillover effects have been observed in parental leave taking (Dahl et al. 2014), teenage substance use (Altonji et al. 2017), and early-life medical interventions (Daysal et al. 2022), for instance.

[3] The existing, mostly recent studies that address this particular issue will be summarized in Section 2.

[4] If children are involved in household or farm work, parents' schooling decisions regarding each child cannot be separated from each other: within the family, allowing a child to dedicate more time to studying might imply a redistribution of household duties and work to the other siblings (Barrera-Osorio et al. 2008).

I follow Sandholtz (2022) in using multiyear, national-level data from Tanzania to investigate how siblings' educational achievement affects their younger counterparts' progression in school. While there is no publicly-available information regarding students' relatives or household size, presumable siblings will be linked according to the most conservative procedure for string matching, as described in Section 4. In particular, last names—and their uniqueness within a cohort and school—will be used to identify different pupils as "siblings." Similar methods have been widely used across the social sciences, including in economics.[5]

I focus on individuals that were matched to a younger "sibling" and that scored at or near the passing threshold on the Primary School Leaving Examination—the high-stakes national test that determines Tanzanians' post-primary schooling opportunities. While there is no information about students' numerical scores, their letter grades in each exam subject are publicly available. I use them to establish which older siblings are likely to have scored *closer* to the passing grade. These individuals should be similar in most respects, other than the fact that some of them barely passed the exam while others barely failed—and were thereby excluded from enrolling in government-run secondary schools.[6] I find that, when school information is omitted from the estimated model, there are positive and statistically significant spillovers from an older sibling's academic achievement on a younger child's attainment. However, the results become statistically insignificant when school effects are considered in virtually any specification. These findings do not change when I test the possibility of heterogeneous effects due to sibling gender match; include individuals with different numbers of presumable siblings in the sample; and perform a series of other robustness analyses.

[5]For instance, Clark and Cummins (2015) use a panel of individuals with rare surnames to study intergenerational wealth mobility in England, Lehne et al. (2018) use Indian surnames—well-known identifiers of caste and religion—in their analysis of political corruption in India, and Cruz et al. (2017) use Philippine naming conventions to identify politicians' family networks and assess their impact on electoral outcomes.

[6]Kippenberg (2014) notes that failing this exam is associated with a higher likelihood of moving from primary school to "mining, grazing, and family activities" instead of secondary instruction.

The remainder of this article is structured as follows: Section 2 reviews the existing literature on sibling-dependency in education and sibling spillovers in educational attainment. Section 3 provides context to the data used in the analysis. Section 4 describes the empirical strategy, including the procedure for sibling matching and the estimated reduced-form models. The main results are presented in Section 5. Section 6 contains an extensive list of robustness checks. The limitations of the proposed approach are discussed in Section 7. Section 8 concludes.

2 Literature Review

Sibling-dependency in education can be seen as a deviation from human capital theory (Becker 1962): in the standard model, an individual's schooling should not depend on "sibship size" (i.e., the number of siblings sharing a biological mother), sibling characteristics (e.g., their gender), or siblings' education. Empirical research has mostly focused on the effects of the number[7] and gender[8] of siblings, as well as the role of birth order,[9] in shaping academic or later-life outcomes—on the contrary, the effect of siblings' educational attainment on one's schooling was, until recently, rarely discussed. At any rate, several mechanisms have been proposed in order to explain both positive and negative sibling-dependency in education.

A possible mechanism is parents' desire to diversify human capital investment across their

[7] The tradeoff between quantity and "quality" of children has long been noted (Becker and Lewis 1973), and an extensive—mostly sociological—literature on sibship size effects ensued. Overall, the suggestion of a causal interpretation for the relationship between the quantity and "quality" of one's offspring has been disputed (Guo and VanWey 1999). Moreover, it has been noted that the negative relationship between sibship size and educational attainment, which is conventionally found in industrialized nations, is much less consistent in the developing world (Lu and Treiman 2008).

[8] The presence of a sister in the household has been found to reduce the education of daughters (Butcher and Case 1994). Moreover, oldest sisters' schooling has been shown to positively affect the education of younger brothers in Pakistan, arguably through the higher quality of child care provided by them to their younger siblings (Qureshi 2018a). Using data from Taiwan, Parish and Willis (1993) identified a positive effect of older sisters on their younger siblings (of both sexes), especially for earlier cohorts and poorer families. Overall, their results suggest that the effect of gender composition of one's sibship is especially relevant in the developing world.

[9] Baland et al. (2016) show that, in Western Cameroon, informal transfers (or "reciprocal credit") within the family result in younger siblings being more educated than their older counterparts. Opposite results have been documented for France, where older siblings have an educational advantage (Mechoulan and Wolff 2015).

children. Parents may want their offspring to assume different professional or family roles in the future, which might require different levels of formal instruction. Furthermore, uncertainty with respect to the returns from schooling should also motivate diversification, leading to negative sibling-dependency in education (Lilleør 2008). Under this hypothesis, children of parents with lower levels of education would be particularly affected. Similarly, parental preferences regarding the distribution of family resources among their offspring (e.g., due to inequality aversion, parents may desire to equalize the resources spent on each child or the expected future income of each sibling) constitute another potential mechanism linking siblings' achievement (Lindskog 2013).

Credit constraints and labor market imperfections have received a greater level of interest in the literature. The human capital investment of credit-constrained households is limited even if, in expectation, future returns to schooling outweigh present costs (Jacoby 1994). Morduch (2000) noted that this induces negative sibling-dependency in education, especially among those closer in age, as siblings would have to compete for scarce household resources. Sometimes, older siblings contribute to alleviate these constraints (Emerson and Souza 2008).

The underdevelopment of formal labor markets can also promote child labor, undermining educational achievement. Indeed, Bhalotra and Heady (2003) show that land-rich households in rural Pakistan and Ghana often have a higher prevalence of child labor than poorer families, since the marginal product of their children's farm work (and, consequently, the opportunity cost of their schooling) should be higher. Similarly, within-sibship differences in ability to perform domestic or farm work can also determine which siblings manage to achieve higher levels of formal instruction, thus creating sibling-dependency in the schooling decision.

Finally, economies of scale in schooling[10] and changes in parents' perceptions of educa-

[10] Aguirre and Matta (2021) show, using evidence from Chile, that positive sibling spillovers in university choice arise due to benefits from siblings attending college together. On a somewhat related note, older individuals may also help their younger siblings with their studies, contributing to their academic achievement (Brody 2004).

tional costs and benefits as a result of siblings' schooling experience are additional mechanisms worth mentioning. Attanasio and Kaufmann (2009) and Jensen (2010) illustrate how perceived returns to education influence the schooling decision; the former notes how youths play an important role in household decisions with respect to human capital investments "on them," while the latter shows how a simple informational treatment can have significant effects on pupils' school attainment. In general, *mis*perceptions seem to significantly influence educational choices, with mothers' being particularly relevant.[11]

Given these proposed mechanisms, results from the empirical literature on sibling spillovers in education can be contextualized. Lindskog (2013) measured how a child's annual primary school entry probability was affected by the education of their brothers and sisters, exploiting within-household variation. She used data from several waves of an Ethiopian household survey, collected in a region where rain-fed subsistence farming was still predominant and the primary-school enrollment rate was slightly above 50% at the time. Her results are consistent with the "credit constraints" explanation of sibling-dependency in education.[12]

More evidence has been presented using data from Europe and the United States. Nicoletti and Rabe (2019), using English data that allowed them to control for child and school-by-cohort-by-subject fixed effects, found a spillover effect from academic achievement (measured by an older sibling's performance in English, Mathematics, and Science tests at age 16) of 11% of a standard deviation on a younger sibling's test scores.[13] Using Dutch data, de Gendre (2022)

[11] See, for instance, Cunha et al. (2020). Further evidence can be found in Attanasio et al. (2022), which concluded that English parents' perceptions regarding the returns to different types of human capital investment (namely, parental time and material investments, as well as school quality) are highly correlated with their actual choices in that regard.

[12] Indeed, empirical results in accordance with a "sibling rivalry" mechanism (motivated by resource constraints) can be frequently found in the literature, and not just with respect to the spillover effects of siblings' education: see, e.g., Shrestha and Palaniswamy (2017) for an analysis focused on the gender gap within the sibship.

[13] The authors defend a causal interpretation for their estimates (against the possibility of reverse causality) due to the fact that younger siblings' outcomes are measured in the future with respect to older siblings' results, and are therefore not expected to affect their main explanatory variables—the same logic applies to my analysis. Other papers, such as Qureshi (2018b), prefer to separately estimate the effect of younger siblings' exposure to a certain treatment on their older counterparts.

identified small negative spillovers from older siblings' (higher) class rank on their younger siblings, driven in part by changes in parental investments and expectations.

Karbownik and Ozek (2019), using data from a Florida school district, followed a different methodology—closer to what is attempted in the analysis below—in order to identify causal effects. They exploit the quasi-experimental variation created by a school starting age policy (whereby students born within days of each other enter primary school in different years due to a school-entry cutoff, being either the youngest or oldest in their class) to estimate sibling spillovers on test scores. Their regression discontinuity design yielded positive and statistically significant estimates of older-to-younger sibling effects; they found no evidence of younger-to-older sibling spillovers. Importantly, they found heterogeneity in sibling effects by family socioeconomic background (e.g., the above-mentioned positive estimates are entirely driven by children in less affluent households).

In conclusion, evidence of sibling spillovers in educational achievement is still relatively scarce, particularly with respect to developing countries. Moreover, studies that explore natural experiments or quasi-experimental variation (instead of employing the more commonly used fixed-effects models) are especially rare. Furthermore, the existing empirical analyses do not focus on the effects of older siblings' outcomes on high-stakes exams of the same nature as those found in Tanzania: i.e., tests that create substantial, plausibly-exogenous variation in school access among similar individuals that scored around the passing threshold. Considering that analyses of peer effects have yielded highly context-specific results even in experimental studies (Sacerdote 2014), casting doubt on the extrapolation of findings to different settings, one may conclude that the literature on sibling spillovers in school achievement is still lacking in extensive, policy-relevant evidence with respect to the developing world.

3 Setting and Data

In Tanzania, primary school—which is compulsory from the age of 7—lasts for seven years ("Standard 1" through "Standard 7"), while there are six years of secondary instruction ("Form 1" through "Form 6"). However, in practice, the Tanzanian education system is structured around a series of national examinations that determine whether and how students may continue their studies.[14] The National Examinations Council of Tanzania is responsible for these exams; student results for several assessments and years are available online.[15] Crucially, there is no unique identifier for each pupil across different exams. Furthermore, these tests are not graded by each school's own teachers, thereby preventing systematic manipulation in grading.

The sources of outcome data for my analysis are the Primary School Leaving Examination (PSLE) and the Form Two National Assessment (FTNA), as detailed below. Summary statistics can be found in Tables A.1 and A.2.

3.1 The Primary School Leaving Examination (PSLE)

At the end of the seventh and final year of primary school, in September, pupils must sit and pass the PSLE in order to enroll in secondary (i.e., non-vocational) instruction at a government-run school; while private schooling can be an alternative for some families, PSLE scores may still be considered for admission. Hence, this test can be thought of as one of the "highest-stakes" national examinations in the country. Parents' choices regarding their children's primary schooling have been shown to reflect this.[16]

[14]Perhaps due to the role of national examinations in selecting which pupils may continue their studies at government-run schools even before they reach secondary schooling, only about one-third of Tanzanian teens completed lower secondary instruction as of 2020; Tanzania trails the average completion rate in Sub-Saharan Africa by approximately 10 percentage points (Source: World Bank).

[15]TETEA, a non-profit organization, maintains an archive (Maktaba Online Resources) of national exam results.

[16]Solomon and Zeitlin (2019) document Tanzanian parents' large *willingness to walk* (i.e., the additional distance from home to school they would accept) in order to have their children enroll in schools with higher average PSLE scores.

Five subjects are evaluated in the PSLE: English, Mathematics, Swahili, Social Studies, and Science, with each subject accounting for one-fifth of the possible exam marks (i.e., 50 out of 250 points). However, NECTA does not share students' marks (either by subject or the overall score), but simply the associated letter grades.[17] The passing threshold is 100 marks (i.e., a C average).

My analysis covers 7 PSLE cohorts (2013-2019) totalling 6,068,930 pupils, with the smallest cohort (around 775,000 students) taking the exam in 2015 and the largest (almost 958,000) in 2018. The pass rate increased every year over the above-mentioned period, going from almost 50% (2013) to 80% (2019).

3.2 Transition to Secondary School and the FTNA

Two years after enrolling in secondary education, pupils in good-standing should pass the FTNA in order to proceed to Form 3. In the absence of complete student records, sitting the FTNA two years following the PSLE can be interpreted as a (conservative) proxy for having transitioned from primary to secondary school—the main outcome of interest in this analysis.

Students are linked according to their full names between the PSLE_t and FTNA_{t+2} national databases. Only unique perfect string matches (i.e., identical, non-repeated full names in both databases) are coded as representing a successful transition from primary to secondary school.

Under this procedure, the fraction of PSLE takers in year t for which there are unique full-name matches sitting the FTNA in year $t + 2$ should be lower than the actual proportion of students that enroll in secondary school following year t's PSLE. First, only 97.9% of PSLE takers had unique full names in their cohort. Moreover, pupils may drop out of secondary school before sitting the FTNA or take the exam at a different time (i.e., some may not advance

[17] The marks obtained in each subject are converted into letter grades according to the following scale: 0-9 marks correspond to an E, 10-19 to a D, 20-29 to a C, 30-39 to a B, and 40-50 to an A. Regarding the overall score: 200 marks is the threshold for an A and 150, 100, and 50 are the cutoffs for a B, a C or a D, respectively.

through secondary school grades at the expected pace). Therefore, the "transition rate" between primary and secondary education is probably underestimated. Figure 1 shows how the estimated transition rate evolved between 2013 and 2019; notice that the probability of transition is virtually zero for a *normal* student that fails the PSLE.

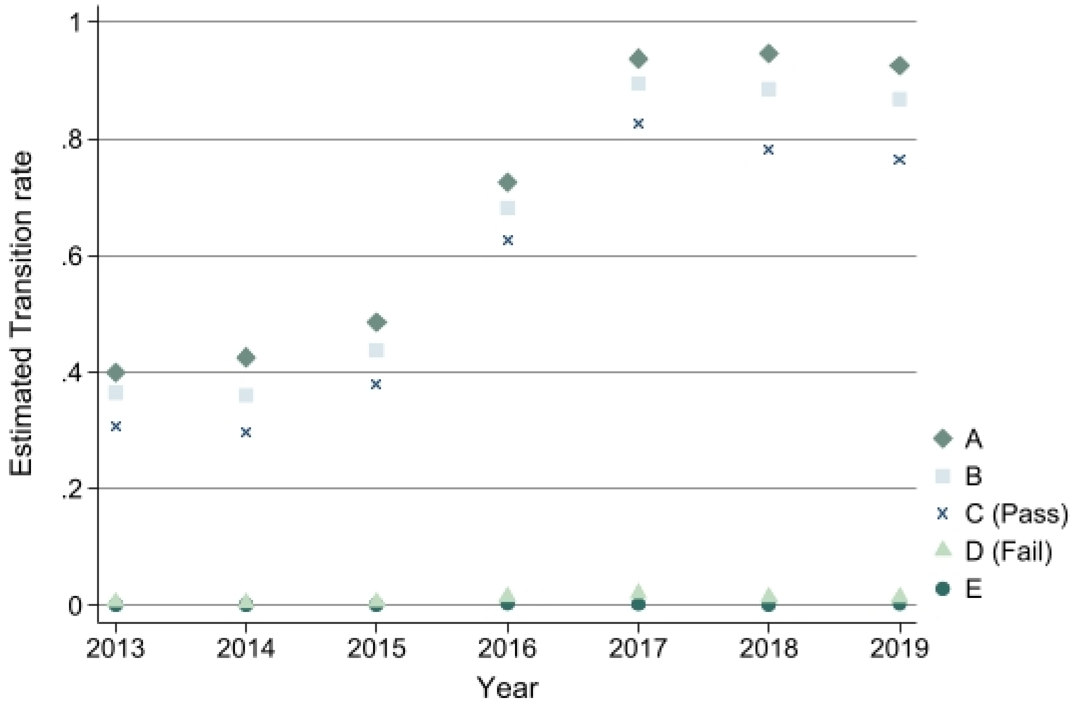

Figure 1: **Primary-to-secondary-school transition rate by PSLE grade and cohort**

I use data from the 2015-2021 FTNA cohorts to obtain the proxy variable for transitioning to secondary school. The proportion of PSLE takers that eventually sit the FTNA two years later increased every year over the aforementioned period, from 16% (2013 PSLE/2015 FTNA cohort) to 65% (2019 PSLE/2021 FTNA cohort). Most of the growth can be attributed to the 2016 and 2017 PSLE cohorts, which is consistent with (and plausibly a consequence of) the introduction of free secondary education in late 2015, with effect from 2016 onwards.[18]

[18] Asim et al. (2019) predicted that Tanzania's Fee-Free Basic Education Policy would lead to a transition rate of 80% by 2025; the planned implementation of automatic promotion from primary to secondary instruction could further increase the enrollment rate by 9 percentage points.

4 Empirical Strategy

4.1 Sibling Matching

Presumable siblings are identified in two stages: first, students whose last names are not unique within their school's PSLE cohort are removed from the sample; second, students from the same school and different cohorts are linked provided they share a last name.[19] Unmatched observations are eliminated. This procedure leaves 1,579,428 pupils in the sample,[20] approximately 26% of the total number of PSLE takers in the original database. Then, I focus most of my analysis on the subset of sibling *pairs* (i.e., those pupils that are matched with a single presumed sibling): in so doing, the sample is reduced to 964,392 individuals (or 482,196 sibling pairs), accounting for almost 16% of the observations in the initial database.[21] The advantage of focusing on sibling pairs is that the relationship between siblings' educational achievement becomes independent (in-sample) from both sibship size and parental preferences regarding the "children quantity-quality tradeoff."

In Subsection 6.1 I show that ward-level[22] (instead of school-level) sibling matching does not change the conclusions of this analysis. Results for sibling *trios* will be presented in Section 5; the estimates of interest are remarkably similar to those obtained when focusing on sibling pairs. Then, in Subsection 6.3, the sample is expanded to include all families (instead of focusing on sibling pairs or trios). That robustness check, more than introducing variation in sibship

[19] In Tanzania, last names are usually shared by all siblings. Surname diversity across the population is also quite high, which increases the likelihood of matched individuals actually being siblings.

[20] Fewer than 600 of these students had someone with the same full name take the PSLE in the previous year in the same school (fewer than 1000 at the ward level). This suggests that the number of PSLE retakers should be negligible.

[21] In the sibling pairs subsample, the Pearson correlation coefficient between siblings' PSLE scores (from A to E)—using, for computation purposes, the midpoint mark of each letter grade—is 0.327. This is lower than the raw correlation between siblings' standardized test scores (0.476) found by Nicoletti and Rabe (2019), for instance, but not implausibly low so as to create obvious doubts regarding the relationship between the individuals I call "siblings."

[22] The "local ward" is the lowest-level administrative division of Tanzania that is common to both urban and rural areas. There are over 2,300 wards in Tanzania; these are grouped into 184 districts and 31 regions.

size to the model, provides suggestive evidence that restricting the analysis to individuals with low *presumed* sibship sizes is not significantly affecting the results.

4.2 "Passing Probability"

Passing the PSLE dramatically influences a pupil's further schooling opportunities: barely passing or failing the exam leads to different choice sets for students who may be fundamentally similar. Therefore, this setting is apparently adequate for a regression discontinuity design. However, in the absence of numerical PSLE scores, there is no clear—let alone continuous—assignment variable for an RDD implementation. Estimating a "local" treatment effect (i.e., the spillover effect from an academically-marginal student's achievement on a younger sibling) requires a different approach.

With the information on subject letter grades for older siblings, I define *passing probability* as the conditional probability of passing the PSLE given a certain combination of subject grades. Students whose five subject grades are all equal to or higher than C have a passing probability of 1; pupils for whom all subject grades are either E or D have a passing probability of 0. There are, however, grade combinations for which the passing probability lies between 0 and 1.[23]

Intuitively, as the passing probability associated with a grade combination approaches 0.5, one would expect the center of the interval of PSLE scores consistent with that combination to approach 100 marks, the passing threshold. Therefore, older siblings whose "passing probabilities" are *close* to 0.5 should have PSLE scores that are, on average, much closer to the cutoff than individuals whose passing probability was not far from either 0 or 1.[24]

[23]For instance, a pair of students who had the same letter grades in all subjects, with C in three parts of the exam and D in two, may end up on different sides of the passing threshold.

[24]The models below will be estimated for different passing probability intervals of the form [0.5-x; 0.5+x], as well as]0; 1[. While the choice of "bandwidth" x is discretionary—and not optimized according to an established procedure— the results are highly consistent across different intervals.

4.3 Econometric Models

Under the assumption that, for students whose grade combinations are associated with passing probabilities approaching 0.5, passing or failing the PSLE happens as if *by chance* (i.e., otherwise similar students are exposed to dissimilar opportunities on account of a small difference in their exam scores), I use the following linear probability model to obtain a first estimate of the "local" spillover effect of an older sibling's academic achievement on a younger brother or sister:

$$P(\theta_i^y = 1|\mathbf{X}_i) = \alpha + \beta\phi_i^o + \gamma X_i^o + u_i, \tag{1}$$

where θ is a binary variable that is equal to 1 if the youngest sibling of family *i* transitioned to secondary school, and 0 otherwise; ϕ is a dummy variable that takes value 1 if the oldest sibling of sibship *i* passed the PSLE, and 0 otherwise; *X* are "grade combination fixed effects"; and *u* is an error term. The superscripts *o* and *y* are included to highlight whether the variables refer to older- or younger-sibling outcomes, respectively.

The baseline equation above does not include year dummies. Moreover, in spite of Tanzania's high regional standing in terms of gender equality (African Development Bank 2015), the empirical literature on sibling effects has consistently found that the gender composition of a sibship may lead to heterogeneous outcomes for younger brothers and sisters. Hence, in order to explicitly account for possible variations in educational achievement due to time effects or gender norms, I expand the baseline equation as follows:

$$P(\theta_i^y = 1|\mathbf{X}_i) = \alpha + \beta\phi_i^o + \gamma X_i^o + \rho t_i + \lambda g_i + u_i, \tag{2}$$

where λ is a vector of coefficients attached to binary indicators for the gender of each sibling, while t_i dummies are added to capture year fixed effects (i.e., related to the time in which each sibling took the PSLE). The consideration of time effects may be particularly important due to the significant changes in Tanzania's educational system throughout the 2013-2019 period, namely the introduction of free secondary education and the large nationwide increases in both the PSLE pass rate and secondary school enrollments.

The third and final specification builds on equation (2), adding the term s_i to account for school effects. In principle, provided that there is covariate balance across the threshold, the estimates for β should be robust to the inclusion of controls—i.e., similar for specifications (1), (2), and (3).

$$P(\theta_i^y = 1|\mathbf{X}_i) = \alpha + \beta\phi_i^o + \gamma X_i^o + \rho t_i + \lambda g_i + \tau s_i + u_i \tag{3}$$

School dummies were not used. Instead, schools were grouped into decile bins according to their overall (2013-2019) PSLE pass rate.[25] The estimates below are robust to the alternative division of schools into quartiles, quintiles, or percentiles.

It is possible that, even for a certain grade combination, students from *better* schools are still disproportionately more likely to fall on the right-hand side of the threshold than pupils from worse schools, which threatens the identification of causal effects using models (1), (2), and (3). Evidence of this behavior can be found in Table A.3. Therefore, I further estimate the following equations:

$$P(\theta_i^y = 1|\mathbf{X}_i) = \alpha + \beta\phi_i^o + \gamma X_i^o s_i + u_i, \tag{4}$$

and

$$P(\theta_i^y = 1|\mathbf{X}_i) = \alpha + \beta\phi_i^o + \gamma X_i^o s_i + \rho t_i + \lambda g_i + u_i, \tag{5}$$

[25] Grouping schools into decile bins based on their 2013 PSLE pass rate (dropping pupils from schools that were only established afterwards, as well as older siblings that belonged to the 2013 PSLE cohort) would probably provide a better measure of school quality (i.e., *separated* from peer effects). Besides, it would mitigate any concern with reverse causality (particularly when the dependent variable is younger siblings' PSLE result). However, it would entail removing observations from over 1,000 of the most recent primary schools in the country, distorting what is meant to be a nationally-representative sample.

where the term associated with the vector of parameters γ is composed of grade-combination-by-school-decile controls.

The use of linear probability models, while convenient for ease of coefficient interpretation, has well-known drawbacks: notably, predicted probabilities may fall below 0 or above 1. Re-estimating these equations through logit regressions would lead to average partial effects that are extremely close to the LPM coefficients presented below, particularly for narrower passing probability intervals. Overall, in this context, linear probability models' computational and interpretative simplicity does not seem to imply a high cost in terms of precision.

5 Results

Results for the narrowest passing probability interval are shown in Table 1. Considering only sibling pairs, I find a statistically significant but economically small older-sibling effect of 2.1 percentage points on the outcome of interest (i.e., younger siblings' primary-to-secondary-school transition probability). The estimate of interest is reduced to 1.5 percentage points when gender and time dummies are added in model (2); results are still significant but no longer at the 5% level. However, the inclusion of a measure of school effects leads to a change of sign and loss of statistical significance for the estimated coefficient, as can be seen in column (3).[26] Results are mostly unchanged in columns (4) and (5), where it is assumed that, controlling for grade-combination-adjusted-by-school (decile), passing the PSLE is essentially "as good as random."

Table A.4 contains the results for three wider passing probability intervals, as well as for the case in which all older siblings with D or C averages are considered, which is equivalent to using a bandwidth of 50 points on each side of the passing threshold (100/250). The sign and

[26]Using school-level dummies would lead to the same conclusion with respect to statistical significance, albeit at a cost in terms of precision.

Table 1: **LPM estimates for the [0.4; 0.6] passing probability window**

	(1) Transition	(2) Transition	(3) Transition	(4) Transition	(5) Transition
Older sibling's result	0.0211**	0.0150*	-0.00960	-0.00263	-0.00886
	(2.356)	(1.739)	(-1.139)	(-0.295)	(-1.031)
Constant	0.545***	0.548***	0.559***	0.556***	0.559***
	(87.59)	(91.86)	(97.77)	(91.91)	(96.82)
Controls:					
Grade Combination	✓	✓	✓		
Gender and Year		✓	✓		✓
School (deciles)			✓		
GC-by-School (deciles)				✓	✓
R^2	0.012	0.088	0.136	0.085	0.160
N	12,518	12,518	12,518	12,518	12,518

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. 'Transition' is a binary variable that takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school), and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE, and 0 otherwise.
*** p<0.01, ** p<0.05, * p<0.1

significance of the estimated coefficients is mostly in line with the results in Table 1, suggesting that the discretionary choice of interval bounds does not unduly affect the main conclusions drawn above.[27] The coefficients from models (1) and (2) tend to be slightly higher and more precisely estimated for the four intervals in Table A.4 than for the narrower window presented in Table 1. However, the magnitude of the point estimates is highly consistent across those four intervals, even as increasingly different older siblings are included in the sample. Overall, the results obtained appear to be both qualitatively and quantitatively consistent across passing probability intervals.[28]

[27] Table A.5 shows logit average partial effect estimates for the four passing probability intervals; these are indeed similar to the LPM estimates presented in Tables 1 and A.4.

[28] Adding geographic controls (e.g., ward dummies) to model (2) would reduce the magnitude of the point estimates of interest but leave the conclusions regarding statistical significance usually unchanged; including ward dummies in equation (5) would lead to quantitatively similar results. Furthermore, replacing younger siblings' cohort fixed effects (t_i^y) by a variable defined as $t_i^y - t_i^o$, i.e., the difference in years between older and younger siblings' PSLE cohorts (a proxy for how close in age these siblings are), would not substantially affect the results.

Estimating the same models for a sample composed of sibling trios (specifically, the middle child and their younger sibling) yielded roughly the same results, as shown in Table A.6. Besides, these results are robust to the inclusion, in any specification, of a binary variable for whether the oldest sibling passed or failed the PSLE.

Positive, statistically significant, but modest effects were identified using specifications (1) and (2). Then, the inclusion of school effects—or the adjustment of "grade combination fixed effects" in models (4) and (5)—eliminated the statistical significance of the estimates for any interval, regardless of the sibship size considered. However, these general conclusions were drawn from regressions that "pooled" siblings of both sexes: in fact, social norms and gender roles could lead to heterogeneous effects from older siblings' educational achievement, depending on the gender of each sibling.

5.1 Heterogeneity by Sibling Gender Match

In order to address the question above, I re-estimate equations (1), (4), and a slightly modified version of equation (5)[29] on two separate samples, depending on the sex of the older sibling. Results can be found in Tables A.7 (for male older siblings) and A.8. The estimated coefficients for equation (1), with the exception of the narrowest passing probability interval, suggest that the effect of an older sister's achievement on her younger sibling may be slightly higher than that of an older brother (one finds differences of around 0.4-0.6 percentage points between the estimated coefficients in each column). However, in the other specifications, estimates generally approach zero and lose statistical significance. With the exception of the interval that includes all older siblings with D or C averages, I do not find evidence of differential effects (on account of gender) from either a male or female older sibling's academic achievement on a younger

[29]This specification includes a dummy variable for the gender of the younger sibling (and naturally not for the older's), as well as an interaction between that variable and the older sibling's PSLE result (Pass or Fail).

sibling.

The analyses in Tables A.7 and A.8 are repeated in Table A.9 for a sample composed entirely of same-sex sibling pairs. The estimated coefficients in the basic specification are higher than the results obtained when estimating the same model for the four intervals in Table A.4. This may suggest that there is a slightly stronger older-sibling effect for siblings of the same gender. In light of the potential mechanisms underlying sibling-dependency in education, as discussed in Section 2, and bearing in mind that these are reduced-form estimates, these findings could be interpreted as suggestive evidence in favor of the role of changes in perceptions of school returns—due to older siblings' educational attainment—in shaping one's outcomes; and against explanations that emphasize the substitutability of (particularly, same-sex) siblings' academic achievement, such as the "household labor need" and diversification hypotheses. However, once more, the estimates generally approach zero when school effects are included, suggesting a negligible older-sibling achievement effect even when possible heterogeneity by sibling gender match is considered.

6 Robustness

The main conclusions regarding the impact of an older sibling's education on their younger brother or sister are robust to (i) the use of different passing probability windows, (ii) a focus on sibships with two or three elements, and (iii) restricting the sample to siblings of the same sex, individuals with a male older sibling, or sibling pairs for which the older sibling is female, as described in Section 5. A series of robustness checks is presented below. Subsection 6.1 explores how changing the fundamental assumption behind sibling matching affects the results. Subsection 6.2 focuses on a policy reform so as to address possible endogeneity arising from omitted variable bias, particularly due to unobserved socioeconomic background. Variation in *presumed* sibship size is introduced to the model in Subsection 6.3. Finally, Subsection 6.4 explores a different variable as a measure of the outcome of interest (i.e., of younger siblings' educational achievement).

6.1 Ward-level Sibling Matching

In order to assess whether the estimates are robust to an even stricter sibling-matching procedure, I repeated the main analysis considering only students with unique last names in their local ward (instead of school) and cohort.

Results can be found in Table A.10. A comparison with the effects obtained under the school-level sibling matching procedure, presented in Tables 1 and A.4, suggests a quantitatively similar estimated impact of older siblings' achievement on their younger counterparts across procedures. The only notable difference is that, for the two largest windows, the estimated coefficients of interest from equation (4) are now statistically significant at the 5% level (although the estimates are small, pointing to a 0.9 percentage point increase in younger siblings' probability of primary-to-secondary-school transition). Therefore, the choice of matching method does not seem to be unduly influencing the conclusions of this analysis.

6.2 Introduction of Free Secondary Education

In November 2015, the Tanzanian government issued "Circular 5," a policy whereby tuition and examination fees (as well as some school equipment charges) would be abolished for the first four years of secondary instruction. Tanzanians became able to attend public schools free of charge for 12 years, starting with one-year of compulsory pre-primary education, and provided that they remained in good standing (namely, as measured by their performance in national exams). While tuition fees were not especially high—even for boarding-school pupils—the total cost of schooling was oftentimes prohibitive for lower-income groups (Oxford Business Group 2019).

Therefore, by lowering the direct costs of secondary schooling, "Circular 5" might have led to higher educational attainment for a significant number of children that were previously too poor to continue their formal instruction, even if their PSLE scores would have allowed them to enroll in government-run secondary schools. Hence, as a robustness check, I re-estimate equations (1), (4), and (5) using two separate samples: first, the group of sibling pairs for which the older sibling took the PSLE in 2013, 2014, or 2015 (i.e., before the announcement of the new policy); second, the group of sibling pairs for which the older sibling sat for the PSLE afterwards (i.e., from 2016 to 2018 inclusive).[30]

Results can be found in Table A.11. Overall, the estimated coefficients with respect to model (1) tend to be substantially higher when one focuses on older siblings from earlier cohorts (panel "A. Pre-FSE"). The fact that there is a stronger relationship between siblings' educational achievement before this policy was implemented can be rationalized (e.g., unobserved family background or household resources could lead to upward bias in the estimates of interest *before* the new policy drastically reduced the number of resource-constrained households). However, it is somewhat surprising: while the empirical literature on sibling spillovers generally suggests that effects tend to be stronger among closely-spaced siblings, sibling pairs in panel A. should, on average, be *less* closely-spaced than their counterparts in panel B. Overall, the main conclusions discussed in Section 5 are largely unaffected: there are positive and mostly significant effects when the baseline equation is estimated; these estimates uniformly become insignificant when any school information is considered.

[30]Estimates are similar if one defines the 2014 PSLE cohort as the last one not to be "treated" by the free secondary education policy. Such a choice would not be unreasonable, given that the 2015 PSLE cohort should have entered secondary school in 2016, i.e., *after* "Circular 5" was issued, even if that cohort sat the PSLE *before* the policy was officially announced.

6.3 Sibship Size

The primary analyses presented above focused on the spillover effect of the oldest sibling in a sibship of size 2; the estimated coefficients were similar, particularly for the largest windows, to those found using the subsample of sibling trios (Table A.6). Focusing on presumed sibling pairs had the advantage of simplifying the structure of the estimated equations: first, one could measure directly the spillover effect of an older sibling on a younger child, without having to consider the outcomes of any other sibling in the estimated model; second, given that presumed sibship size was equal for all individuals in the sample, the functional form of the three specifications above did not have to account for variation in sibship size.

Table A.12 presents the results from an enlarged sample, in which spillovers from academically-marginal *penultimate* siblings on the younger sibling of each presumed sibship are estimated. Overall, the results are in line with the estimates shown in Tables 1, A.4, and A.6, when the sample was restricted to sibships of sizes 2 or 3. Estimates are also robust to the inclusion of the number of older siblings (and its square) as additional independent variables in specification (5); the omission of the squared term would not meaningfully change the estimates of interest.[31] Hence, the preferred use of the subsample of sibling *pairs*, which allowed me to focus most of the analysis on streamlined models that did not have to account for variation in sibship size, does not seem to be driving my conclusions.

6.4 The PSLE Result as a Measure of Younger Siblings' Outcomes

Younger siblings' enrollment in secondary school has thus far been used as the dependent variable in all regressions, or as the measure of younger siblings' academic achievement. However,

[31] Considering the sibling matching procedure followed here and the fact that only 7 years of PSLE data are available, it would be unadvised to treat any sibship size "effect" obtained below as a valid estimate. The purpose of this analysis is not to identify a causal effect of sibship size on educational achievement; the inclusion of PSLE takers with different *presumable* sibship sizes in the sample is merely performed as part of a robustness check.

as explained in Section 3, that variable is measured with error: in practice, the information therein should lead to an underestimation of the primary-to-secondary-school transition rate. A different metric for younger siblings' outcomes—their PSLE result (Pass or Fail)—can be used to perform a simple robustness check of the general conclusions, with two advantages. First, I have no reason to suspect that this variable is measured with error; second, I am able to include in the sample the relatively few individuals who did not have a unique full name in their cohort's national FTNA database, and for whom their secondary school enrollment status was therefore undetermined.

Table A.13 shows the results for the case in which the younger sibling's PSLE result (Pass or Fail) is used as the dependent variable, with the older sibling's result as the explanatory variable of greater interest. Once more, the estimated coefficients are in accordance with baseline spillover estimates, suggesting that the choice of dependent variable is not driving the conclusions with respect to sibling spillovers; additionally, they are also quantitatively similar in both panels, minimizing concerns regarding the exclusion of individuals with repeated names in their cohort from the preferred sample.

7 Discussion

Adding to the findings related to heterogeneity due to the older sibling's gender, the extensive robustness analyses presented in Section 6 have reinforced the fundamental conclusion of this study: namely, that any evidence of positive and small sibling spillovers—which would be in line with Nicoletti and Rabe (2019)—is invariably annulled by the inclusion of any form of school information in the estimated equations. Without prejudice for the apparent robustness of its findings, this analysis was subject to a series of limitations that are worth discussing.

First, while the use of national-level data in this analysis is distinctive (previous studies on the same topic generally use regional- or state-level datasets), there is a costly tradeoff with respect to the granularity and richness of the available information. In particular, the lack of data on pupils' PSLE numerical scores precludes the use of a regression discontinuity design,[32] while the modest amount of available covariates diminishes my ability to present extensive balance checks. In particular, longitudinal data on pupils' educational achievement, as well as information on age and sibship size, would be especially useful.

Perhaps more relevant are the difficulties related to the measurement of pupils' primary-to-secondary-school transition, in addition to the procedure used to match presumable siblings. While the former does not seem to significantly affect the results (compare, for instance, both panels of Table A.13), the latter is especially concerning. In the previous section I provided evidence that linking observations at the ward (instead of school) level did not meaningfully affect the estimates of greater interest. However, the matching procedure itself—i.e., linking individuals with unique surnames in their cohort, at any level—cannot easily be tested against an alternative. By definition, twins in the same school cohort will be dropped from the sample, while it is virtually assured that a certain percentage of unrelated pupils will be matched as presumable siblings; these are two types of error whose prevalence in the sample cannot be tested.

Nevertheless, there are advantages to this setting and data. First, it exploits the results from nationwide, standardized examinations in a developing country; given the context-specific nature of findings in the peer effects literature (Sacerdote 2014), contributions based on often overlooked sources can be useful. Moreover, Tanzania is a highly diverse country, both linguistically and ethnically; crucially, this contributes to a high degree of surname variability in the

[32]Moreover, I am unable to test for bunching on the right-hand side of the cutoff using, for instance, a McCrary test (McCrary 2008). However, recall that exams are not graded at the local or school level; hence, grade manipulation is highly unlikely.

country.[33] Even if the sibling-matching procedure followed here could *bias* the sample toward individuals with unusual last names (which might be indicators of specific ethnic, religious, or social backgrounds), the fact that Tanzanian naming conventions are highly flexible even within each of these groups (Mlahleki 2019) mitigates such concerns.

8 Conclusions

This study sought to identify a causal relationship between siblings' educational achievement using pooled cross-sectional data from one of the "highest-stakes" national exams in Tanzania, under the assumption that, for certain combinations of subject letter grades, passing or failing that exam was "as if randomized." The spillover effect of an academically-marginal student's passing grade on the PSLE on a younger sibling, measured as either the increase in probability of primary-to-secondary-school transition or of passing the same national examination, has been estimated to be essentially zero, which is particularly clear when one focuses on estimates for equation (4).

Naturally, it could be the case that spillover effects are heterogeneous along the distribution of older siblings' achievement (e.g., having a high score on the exam could lead to a shift of household resources in favor of the talented child and away from the other siblings, or create a greater incentive for younger siblings to excel academically); the methodology followed herein focuses on quasi-experimental variation around the passing threshold, thereby excluding the *best* and *worst* older siblings, preempting such an heterogeneity analysis. Further research on this issue could provide suggestive evidence regarding the mechanisms underlying sibling-dependency in education.

[33] Out of approximately 6 million PSLE takers across a 7-year period, I find over 300,000 different last names, of which over 170,000 appear more than once. In the subsample of sibling pairs, over 78,300 different last names can be found.

Moreover, one could argue that older siblings' higher levels of schooling, instead of greater achievement as measured by the PSLE, could be a better choice of independent variable for the models above. However, transition to secondary school is not entirely determined by academic merit; socioeconomic background and insufficient school supply are two additional predictors of primary-to-secondary-school transition that I cannot account for with available data. Therefore, if included as a regressor in any specification, a binary variable for older siblings' transition to secondary school would most likely be endogenous. Using older siblings' PSLE result (Pass or Fail) as an Instrumental Variable would probably imply a violation of the exclusion restriction.[34] Subject to data availability, future studies could show how different measures of older siblings' achievement (i.e., of both *quantity* and *quality* of schooling), in the same context, lead to either identical or dissimilar sibling spillover estimates.

In conclusion, the growing empirical evidence of the importance of role models and peer effects on individual outcomes justifies further research with a focus on sibling interactions, which have long been considered relevant but were—at least until the last decade—mostly overlooked due to identification challenges. In particular, novel studies of educational achievement spillovers that use data from developing countries are especially relevant, given the current lack of extensive evidence from such settings. Furthermore, both the policy and scientific relevance of further research are enhanced by the importance given to expanding educational access in the developing world, particularly in regions where there is still underprovision of secondary instruction, as household dynamics and other social interactions could affect the outcome of related policy interventions.

[34]The exclusion restriction, in this context, would require that passing the PSLE only affected the outcome of interest through further schooling (i.e., through the endogenous variable). However, estimating equation (1) for the subsample of students that did not enroll in secondary school (even if they passed the PSLE) yielded "older sibling effect" estimates that were significantly different from zero, suggesting that having an older sibling that passed the PSLE may affect younger siblings' outcomes through channels other than further schooling.

References

African Development Bank (2015). Empowering African women: An agenda for action. Technical report. Retrieved from https://www.afdb.org/en/topics-and-sectors/topics/quality-assurance-results/gender-equality-index.

Aguirre, J. and J. Matta (2021). Walking in your footsteps: Sibling spillovers in higher education choices. *Economics of Education Review 80*, 102062.

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (2021). O brother, where start thou? Sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics 136*(3), 1831–1886.

Altonji, J. G., S. Cattan, and I. Ware (2017). Identifying sibling influence on teenage substance use. *Journal of Human Resources 52*(1), 1–47.

Asim, S., D. Chugunov, and R. Gera (2019). Fiscal implications of free education. Policy note.

Attanasio, O., T. Boneva, and C. Rauh (2022). Parental beliefs about returns to different types of investments in school children. *Journal of Human Resources 57*(6), 1789–1825.

Attanasio, O. and K. Kaufmann (2009). Educational choices, subjective expectations, and credit constraints. *NBER Working Paper no. 15087*.

Baland, J.-M., I. Bonjean, C. Guirkinger, and R. Ziparo (2016). The economic consequences of mutual help in extended families. *Journal of Development Economics 123*, 38–56.

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2008). Conditional cash transfers in education: design features, peer and sibling effects Evidence from a randomized experiment in Colombia. *NBER Working Paper no. 13890*.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy 70*(5, Part 2), 9–49.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2, Part 2), S279–S288.

Bhalotra, S. and C. Heady (2003). Child farm labor: The wealth paradox. *The World Bank Economic Review 17*(2), 197–227.

Brody, G. H. (2004). Siblings' direct and indirect contributions to child development. *Current Directions in Psychological Science 13*(3), 124–126.

Butcher, K. F. and A. Case (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics 109*(3), 531–563.

Clark, G. and N. Cummins (2015). Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility. *The Economic Journal 125*(582), 61–85.

Cruz, C., J. Labonne, and P. Querubin (2017). Politician family networks and electoral outcomes: Evidence from the Philippines. *American Economic Review 107*(10), 3006–37.

Cunha, F., I. Elo, and J. Culhane (2020). Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *Journal of Econometrics*.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review 104*(7), 2049–74.

Daysal, N. M., M. Simonsen, M. Trandafir, and S. Breining (2022). Spillover effects of early-life medical interventions. *The Review of Economics and Statistics 104*(1), 1–16.

de Gendre, A. (2022). Class rank and sibling spillover effects. Working paper, University of Melbourne.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The economics of parenting. *Annual Review of Economics 11*, 55–84.

Emerson, P. M. and A. P. Souza (2008). Birth order, child labor, and school attendance in Brazil. *World Development 36*(9), 1647–1664.

Goodman, J., M. Hurwitz, J. Smith, and J. Fox (2015). The relationship between siblings' college choices: Evidence from one million SAT-taking families. *Economics of Education Review 48*, 75–85.

Guo, G. and L. K. VanWey (1999). Sibship size and intellectual development: Is the relationship causal? *American Sociological Review*, 169–187.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*, 689–733.

Jacoby, H. G. (1994). Borrowing constraints and progress through school: Evidence from Peru. *The Review of Economics and Statistics*, 151–160.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Joensen, J. S. and H. S. Nielsen (2018). Spillovers in education choice. *Journal of Public Economics 157*, 158–183.

Karbownik, K. and U. Ozek (2019). Setting a good example? Examining sibling spillovers in educational achievement using a regression discontinuity design. *NBER Working Paper no. 26411*.

Kippenberg, J. (2014). Tanzania: Let's tackle the primary school leaving exam. *Daily News (Tanzania)*. May 13. Last accessed on December 14, 2022. Retrieved from https://www.hrw.org/news/2014/05/13/tanzania-lets-tackle-primary-school-leaving-exam.

Lehne, J., J. N. Shapiro, and O. V. Eynde (2018). Building connections: Political corruption and road construction in India. *Journal of Development Economics 131*, 62–78.

Lilleør, H. B. (2008). Sibling dependence, uncertainty and education. Findings from Tanzania. Working paper, University of Copenhagen.

Lindskog, A. (2013). The effect of siblings' education on school-entry in the Ethiopian highlands. *Economics of Education Review 34*, 45–68.

Lu, Y. and D. J. Treiman (2008). The effect of sibship size on educational attainment in China: Period variations. *American Sociological Review 73*(5), 813–834.

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

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Mechoulan, S. and F.-C. Wolff (2015). Intra-household allocation of family resources and birth order: evidence from France using siblings data. *Journal of Population Economics 28*(4), 937–964.

Mlahleki, G. (2019). Is Tanzania losing its identity? *The Citizen*. November 26. Updated on March 31, 2021. Last accessed on December 14, 2022. Retrieved from https://www.thecitizen.co.tz/tanzania/magazines/sound-living/is-tanzania-losing-its-identity--2697964.

Morduch, J. (2000). Sibling rivalry in Africa. *American Economic Review 90*(2), 405–409.

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

Oxford Business Group (2019). Enrolment rates rise in Tanzania with free secondary education. Last accessed on December 4, 2022. Retrieved from https://oxfordbusinessgroup.com/overview/smart-adaptations-introduction-free-secondary-education-has-boosted-enrolment-rates-while-private.

Parish, W. L. and R. J. Willis (1993). Daughters, education, and family budgets: Taiwan experiences. *Journal of Human Resources*, 863–898.

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

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

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annual Review of Economics 6*(1), 253–272.

Sandholtz, W. A. (2022). Secondary school access raises primary school achievement in Tanzania. Working paper, Nova School of Business and Economics.

Shrestha, S. A. and N. Palaniswamy (2017). Sibling rivalry and gender gap: intrahousehold substitution of male and female educational investments from male migration prospects. *Journal of Population Economics 30*(4), 1355–1380.

Solomon, S. and A. Zeitlin (2019). What do Tanzanian parents want from primary schools—and what can be done about it? Technical report. Last accessed December 11, 2022. Retrieved from https://riseprogramme.org/publications/what-do-tanzanian-parents-want-primary-schools-and-what-can-be-done-about-it.

A Tables

Table A.1: **Descriptive statistics**

	2013	2014	2015	2016	2017	2018	2019
Primary Schools							
Number of schools	15,656	15,867	16,096	16,350	16,575	16,826	17,047
Government-run	0.9623	0.9639	0.9607	0.9553	0.9516	0.9453	0.9395
PSLE							
Exam takers (#)	867,983	808,085	775,273	795,740	916,885	957,893	947,071
Female	0.5252	0.5317	0.5340	0.5314	0.5281	0.5249	0.5236
Pass rate	0.4927	0.5586	0.6681	0.6980	0.7221	0.7662	0.8032
Female pass rate	0.4568	0.5272	0.6380	0.6712	0.7043	0.7622	0.7990
FTNA							
Transition rate (est.)	0.1613	0.1774	0.2700	0.4533	0.6197	0.6398	0.6535
Female transition rate	0.1560	0.1752	0.2684	0.4469	0.6165	0.6494	0.6664

Notes: the proxy for primary-to-secondary-school transition (obtained using FTNA data) is explained in Subsection 3.2.

Table A.2: **Summary statistics**

	N	Prop. of Total
All PSLE takers	6,068,930	1
Full information	5,903,062	0.973
Unique name in school cohort	3,873,297	0.638
Presumed sibship size > 1:		
2	964,392	0.159
3	422,331	0.070
4	148,396	0.024
5	38,025	0.006
6	5,808	0.001
7	476	
Total (sibship size > 1)	1,579,428	0.260

Notes: "Full information" means that the student's PSLE grade, subject letter grades, and school location are available.

Table A.3: **Covariate balance check: *t-tests* around the cutoff**

	[0.4; 0.6]	[0.25; 0.75]	D or C Avg.
Female	0.0101	-0.0026	0.057***
	(1.161)	(-.531)	(34.377)
School pass rate:			
1st decile	0.0354***	0.0404***	0.1345***
	(6.496)	(13.568)	(131.993)
2nd decile	0.0285***	0.0284***	0.0816***
	(4.937)	(9.036)	(79.191)
3rd decile	0.0154***	0.0165***	0.0524***
	(2.656)	(5.102)	(50.475)
4th decile	0.0053	0.0114***	0.0266***
	(.935)	(3.599)	(25.774)
5th decile	-0.0019	-0.004	0.0013
	(-.335)	(-1.265)	(1.278)
6th decile	-0.0137**	-0.0057*	-0.0181***
	(-2.374)	(-1.792)	(-17.745)
7th decile	0.0017	-0.0144***	-0.042***
	(.32)	(-4.715)	(-41.68)
8th decile	-0.0115**	-0.0186***	-0.0652***
	(-2.15)	(-6.361)	(-65.533)
9th decile	-0.0284***	-0.0264***	-0.0876***
	(-5.934)	(-10.216)	(-92.026)
10th decile	-0.03086***	-0.0275***	-0.0835***
	(-9.079)	(-15.954)	(-106.314)

Notes: t-statistics in parentheses; negative values indicate that the sample mean on the left-hand side of the threshold is higher than to the right. The fact that substantial differences arise even for the narrower passing probability intervals suggest that they are not merely due to a *mechanical* effect related to the way "school deciles" are defined (i.e., it is not surprising that students who pass are more likely to attend schools with higher overall pass rates, but the fact that this is noticeable *right around the threshold* could be unexpected).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.4: **LPM estimates for wider intervals around the threshold**

	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)				
Older sibling's result	0.0266***	0.0273***	0.0298***	0.0298***
	(5.392)	(6.468)	(7.652)	(7.645)
Constant	0.532***	0.540***	0.540***	0.532***
R^2	0.008	0.008	0.012	0.039
Equation (2)				
Older sibling's result	0.0204***	0.0204***	0.0233***	0.0235***
	(4.344)	(5.067)	(6.287)	(6.322)
Constant	0.534***	0.544***	0.543***	0.536***
R^2	0.094	0.103	0.103	0.120
Equation (3)				
Older sibling's result	-0.00548	-0.00568	-0.00432	-0.00463
	(-1.189)	(-1.443)	(-1.197)	(-1.284)
Constant	0.544***	0.559***	0.560***	0.552***
R^2	0.139	0.148	0.149	0.166
Equation (4)				
Older sibling's result	0.00154	0.00170	0.00272	0.00272
	(0.314)	(0.407)	(0.706)	(0.702)
Constant	0.541***	0.555***	0.555***	0.548***
R^2	0.068	0.066	0.070	0.105
Equation (5)				
Older sibling's result	-0.00478	-0.00515	-0.00354	-0.00343
	(-1.026)	(-1.296)	(-0.971)	(-0.935)
Constant	0.544***	0.559***	0.559***	0.551***
R^2	0.152	0.158	0.158	0.182
N	43,263	74,218	165,626	368,069

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. Each column presents estimates from a restricted sample: in the first 3 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. See Subsection 4.3 and Table 1 for details regarding the regressors.
*** p<0.01

Table A.5: **Logit estimates for sibling pairs (average partial effects)**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[
Same controls as model (1)				
Older sibling's result	0.0211**	0.0266***	0.0272***	0.0296***
	(2.362)	(5.398)	(6.482)	(7.665)
Same controls as model (2)				
Older sibling's result	0.0149*	0.0204***	0.0203***	0.0231***
	(1.741)	(4.336)	(5.060)	(6.269)
Same controls as model (3)				
Older sibling's result	-0.00939	-0.00557	-0.00580	-0.00455
	(-1.118)	(-1.210)	(-1.483)	(-1.274)
N	12,505	43,250	74,200	165,608
Same controls as model (4)				
Older sibling's result	-0.00265	0.00154	0.00170	0.00269
	(-0.299)	(0.316)	(0.409)	(0.710)
Same controls as model (5)				
Older sibling's result	-0.00891	-0.00501	-0.00541	-0.00385
	(-1.049)	(-1.081)	(-1.376)	(-1.071)
N	12,310	42,792	73,628	164,728

Notes: standard errors clustered by school; z-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. Each column presents estimates from subsamples determined by the passing probability intervals shown in the header. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. The estimated APE closely resemble the estimates in Tables 1 and A.4.
*** p<0.01, ** p<0.05, * p<0.1

Table A.6: **LPM estimates using a sample of sibling *trios***

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Middle sibling's result	0.0411***	0.0207**	0.0218***	0.0268***	0.0268***
	(2.641)	(2.173)	(2.665)	(3.571)	(3.563)
R^2	0.018	0.011	0.010	0.015	0.039
Equation (2)					
Middle sibling's result	0.0352**	0.0178*	0.0187**	0.0245***	0.0246***
	(2.283)	(1.910)	(2.331)	(3.331)	(3.342)
R^2	0.046	0.047	0.045	0.047	0.066
Equation (3)					
Middle sibling's result	0.0113	-0.00802	-0.00681	-0.00180	-0.00217
	(0.749)	(-0.873)	(-0.867)	(-0.251)	(-0.303)
R^2	0.103	0.098	0.097	0.098	0.118
Equation (4)					
Middle sibling's result	0.0155	-0.00788	-0.00612	-0.00205	-0.00205
	(0.971)	(-0.816)	(-0.744)	(-0.274)	(-0.271)
R^2	0.138	0.100	0.092	0.090	0.132
Equation (5)					
Middle sibling's result	0.00994	-0.0103	-0.00894	-0.00416	-0.00410
	(0.630)	(-1.088)	(-1.108)	(-0.565)	(-0.550)
R^2	0.166	0.133	0.126	0.121	0.157
N	3,961	11,028	19,787	42,690	102,118

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to *two* presumable siblings. N is the number of such sibling *trios*. The binary dependent variable in all regressions takes value 1 if the youngest sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Middle sibling's result' is a dummy that takes value 1 if the middle child in the sibship passed the PSLE and 0 otherwise. Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all "middle" siblings who had a D or C average (and their younger counterparts) are included. All regressions include an intercept. See Subsection 4.3 and Table 1 for details regarding controls.
*** p<0.01, ** p<0.05, * p<0.1

Table A.7: **LPM estimates for sibling pairs — male older sibling**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Older sib. PSLE result	0.0292**	0.0248***	0.0254***	0.0268***	0.0268***
	(2.024)	(3.264)	(3.911)	(4.467)	(4.461)
Constant	0.543***	0.525***	0.532***	0.534***	0.531***
	(54.72)	(107.3)	(107.4)	(124.3)	(131.1)
R^2	0.019	0.010	0.009	0.012	0.038
Equation (4)					
Older sibling's result	0.0104	0.00114	0.000534	-0.000484	-0.000484
	(0.707)	(0.149)	(0.0826)	(-0.0811)	(-0.0805)
Constant	0.552***	0.535***	0.547***	0.551***	0.547***
	(56.29)	(113.3)	(115.0)	(133.7)	(139.9)
R^2	0.125	0.083	0.074	0.079	0.116
With additional controls					
Older sibling's result	-0.00943	-0.00319	-0.00326	-0.00251	0.000606
	(-0.461)	(-0.303)	(-0.390)	(-0.367)	(0.0973)
Gender (Female = 1)	-0.0232	0.00343	0.00172	0.0109*	0.0141***
	(-1.202)	(0.380)	(0.206)	(1.922)	(3.620)
Sibling's result x Gender	0.0268	0.000560	-0.00299	-0.00483	-0.0109**
	(0.959)	(0.0387)	(-0.281)	(-0.673)	(-2.252)
Constant	0.567***	0.534***	0.549***	0.548***	0.543***
	(41.28)	(81.20)	(86.09)	(110.3)	(128.4)
R^2	0.197	0.165	0.168	0.168	0.197
N	4,815	18,090	32,294	71,788	166,034

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort, who were matched to a single presumable sibling. N is the number of such sibling pairs for which there is a male older sibling. Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. Equation (1) includes grade combination controls, while model (4) includes "grade-combination-by-school-pass-rate-decile fixed effects." For details on the last panel, see Subsection 5.1.
*** p<0.01, ** p<0.05, * p<0.1

Table A.8: **LPM estimates for sibling pairs — female older sibling**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Older sib. PSLE result	0.0171	0.0285***	0.0295***	0.0326***	0.0326***
	(1.491)	(4.379)	(5.286)	(6.325)	(6.317)
Constant	0.546***	0.536***	0.546***	0.544***	0.534***
	(69.63)	(124.8)	(128.1)	(151.3)	(162.0)
R^2	0.015	0.010	0.011	0.015	0.045
Equation (4)					
Older sibling's result	-0.00951	0.00194	0.00305	0.00507	0.00507
	(-0.823)	(0.298)	(0.547)	(0.985)	(0.978)
Constant	0.558***	0.546***	0.561***	0.559***	0.549***
	(72.31)	(133.1)	(137.2)	(163.8)	(173.8)
R^2	0.094	0.074	0.072	0.078	0.117
With additional controls					
Older sibling's result	-0.00534	-0.00990	-0.00331	-0.00422	-0.0162***
	(-0.323)	(-1.075)	(-0.444)	(-0.704)	(-2.958)
Gender (Female = 1)	0.0429***	0.0116	0.0146**	0.00999**	-0.00281
	(2.862)	(1.497)	(2.018)	(2.114)	(-0.860)
Sibling's result x Gender	-0.0200	0.00688	-0.00347	0.00248	0.0245***
	(-0.912)	(0.562)	(-0.374)	(0.398)	(5.748)
Constant	0.537***	0.543***	0.558***	0.558***	0.555***
	(47.95)	(92.37)	(98.79)	(133.4)	(157.4)
R^2	0.173	0.160	0.162	0.163	0.191
N	7,703	25,173	41,924	93,838	202,035

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort, who were matched to a single presumable sibling. N is the number of such sibling pairs for which there is a female older sibling. Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. Equation (1) includes grade combination controls, while model (4) includes "grade-combination-by-school-pass-rate-decile fixed effects." For details on the last panel, see Subsection 5.1.

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

Table A.9: **LPM estimates using same-sex sibling pairs**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Older sibling's result	0.0147	0.0322***	0.0298***	0.0345***	0.0345***
	(1.196)	(4.719)	(5.128)	(6.406)	(6.398)
Constant	0.561***	0.532***	0.542***	0.538***	0.530***
	(66.32)	(118.8)	(120.5)	(141.6)	(151.1)
R^2	0.015	0.011	0.010	0.014	0.045
Equation (4)					
Older sibling's result	-0.00799	0.00719	0.00490	0.00762	0.00762
	(-0.639)	(1.051)	(0.843)	(1.418)	(1.408)
Constant	0.571***	0.542***	0.556***	0.553***	0.545***
	(68.30)	(126.3)	(129.0)	(152.6)	(161.1)
R^2	0.100	0.075	0.069	0.075	0.118
With additional controls					
Older sibling's result	-0.00953	-0.00158	-0.00344	-0.00313	-0.0107*
	(-0.469)	(-0.150)	(-0.415)	(-0.478)	(-1.829)
Gender (Female = 1)	0.0154	0.00732	0.00685	0.00595	0.000146
	(0.892)	(0.871)	(0.879)	(1.137)	(0.0414)
Sibling's result x Gender	-0.0165	-0.00150	-0.00275	0.00283	0.0157***
	(-0.656)	(-0.113)	(-0.276)	(0.422)	(3.480)
Constant	0.567***	0.541***	0.558***	0.555***	0.550***
	(40.72)	(81.75)	(87.81)	(117.0)	(141.5)
R^2	0.175	0.163	0.161	0.163	0.194
N	6,643	22,765	38,736	86,387	190,893

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort, who were matched to a single presumable sibling, *and* who have the same gender. N is the number of such sibling pairs. Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. Equation (1) includes grade combination controls, while model (4) includes "grade-combination-by-school-pass-rate-decile fixed effects." For details on the last panel, see Subsection 5.1.
*** p<0.01, ** p<0.05, * p<0.1

Table A.10: **LPM estimates for sibling pairs following ward-level sibling matching**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Older sibling's result	0.0265***	0.0254***	0.0245***	0.0304***	0.0304***
	(2.617)	(4.646)	(5.173)	(6.983)	(6.975)
Constant	0.553***	0.548***	0.560***	0.554***	0.548***
	(79.17)	(144.3)	(146.8)	(166.7)	(177.4)
R^2	0.009	0.008	0.008	0.011	0.034
Equation (4)					
Older sibling's result	0.00655	0.00578	0.00463	0.00905**	0.00905**
	(0.640)	(1.045)	(0.968)	(2.066)	(2.054)
Constant	0.563***	0.555***	0.571***	0.566***	0.560***
	(82.59)	(155.2)	(158.4)	(182.6)	(193.2)
R^2	0.076	0.054	0.053	0.056	0.087
Equation (5)					
Older sibling's result	0.000837	-0.00156	-0.00288	0.000974	0.00106
	(0.0852)	(-0.295)	(-0.633)	(0.234)	(0.252)
Constant	0.565***	0.558***	0.576***	0.571***	0.565***
	(86.75)	(164.8)	(168.6)	(194.8)	(204.9)
R^2	0.157	0.148	0.151	0.150	0.172
N	9,758	34,321	56,752	130,191	290,279

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their *ward*'s cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. Equation (1) includes grade combination controls, while models (4) and (5) includes "grade-combination-by-school-pass-rate-decile fixed effects." The last equation also includes gender and time controls.

*** $p<0.01$, ** $p<0.05$

Table A.11: **LPM estimates for sibling pairs before and after the introduction of FSE**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
A. Pre-FSE (older sib.)					
Equation (1)					
Older sibling's result	0.0258**	0.0286***	0.0307***	0.0323***	0.0323***
	(2.201)	(4.892)	(6.057)	(6.863)	(6.856)
R^2	0.009	0.007	0.006	0.010	0.034
Equation (4)					
Older sibling's result	-0.000825	0.00319	0.00470	0.00453	0.00453
	(-0.0700)	(0.548)	(0.937)	(0.972)	(0.967)
R^2	0.094	0.068	0.065	0.070	0.103
Equation (5)					
Older sibling's result	-0.00902	-0.00411	-0.00303	-0.00219	-0.00203
	(-0.816)	(-0.751)	(-0.642)	(-0.503)	(-0.463)
R^2	0.205	0.175	0.183	0.184	0.206
Mean (dep. var.)	0.5156	0.5137	0.5243	0.5231	0.5083
N	7,424	30,428	51,251	112,081	243,611
B. Post-FSE					
Equation (1)					
Older sibling's result	0.0127	0.0190**	0.0152**	0.0196***	0.0196***
	(0.922)	(2.146)	(2.028)	(2.851)	(2.847)
R^2	0.017	0.014	0.014	0.015	0.039
Equation (4)					
Older sibling's result	-0.00517	-0.00602	-0.0106	-0.00808	-0.00808
	(-0.371)	(-0.675)	(-1.410)	(-1.182)	(-1.171)
R^2	0.121	0.104	0.093	0.094	0.130
Equation (5)					
Older sibling's result	-0.00584	-0.00585	-0.00991	-0.00771	-0.00779
	(-0.420)	(-0.658)	(-1.326)	(-1.131)	(-1.132)
R^2	0.124	0.108	0.097	0.098	0.133
Mean (dep. var.)	0.6123	0.6090	0.6264	0.6280	0.6301
N	5,094	12,835	22,967	53,545	124,458

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. Each column presents estimates from a restricted sample (according to the criteria in the header). The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. 'Older sibling's result' is a dummy that takes value 1 if the older sibling passed the PSLE and 0 otherwise. Panel A (B) includes older siblings that took the PSLE in 2013-2015 (2016-2018).

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

Table A.12: **LPM estimates for the full sample (any sibship size)**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
Equation (1)					
Older sibling's result	0.0211***	0.0245***	0.0269***	0.0295***	0.0295***
	(2.855)	(5.776)	(7.276)	(8.823)	(8.816)
Constant	0.557***	0.547***	0.550***	0.555***	0.549***
R^2	0.012	0.007	0.007	0.012	0.038
Equation (4)					
Older sibling's result	-0.00310	-0.000495	0.00136	0.00267	0.00267
	(-0.422)	(-0.118)	(0.373)	(0.810)	(0.807)
Constant	0.568***	0.556***	0.565***	0.571***	0.565***
R^2	0.085	0.066	0.065	0.069	0.100
Equation (5)					
Older sibling's result	-0.00825	-0.00652	-0.00544	-0.00336	-0.00326
	(-1.158)	(-1.616)	(-1.557)	(-1.067)	(-1.032)
Constant	0.571***	0.559***	0.569***	0.575***	0.569***
R^2	0.146	0.140	0.145	0.144	0.166
Extended equation (5)					
Older sibling's result	-0.00820	-0.00655	-0.00546	-0.00340	-0.00330
	(-1.152)	(-1.625)	(-1.565)	(-1.083)	(-1.047)
# Older siblings	0.0210	0.0274*	0.0178	0.0243***	0.0251***
	(0.858)	(1.851)	(1.528)	(3.463)	(5.465)
# Older siblings (sq.)	-0.00264	-0.00383	-0.00186	-0.00284*	-0.00320***
	(-0.446)	(-1.047)	(-0.638)	(-1.657)	(-2.888)
Constant	0.548***	0.531***	0.550***	0.549***	0.542***
R^2	0.146	0.140	0.145	0.144	0.167
N	17,969	57,553	93,609	227,773	502,315

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs. Each column presents estimates from a restricted sample: in the first 3 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling took the FTNA two years after sitting the PSLE (i.e., transitioned to secondary school) and 0 otherwise. See Subsections 4.3 and 6.3, as well as Table 1, for details regarding the regressors; all regressions include an intercept (for which the t-statistics are omitted).
*** p<0.01, ** p<0.05, * p<0.1

Table A.13: **LPM estimates for sibling pairs. Dependent variable: PSLE result (Pass = 1)**

	[0.4; 0.6]	[0.25; 0.75]	[0.1; 0.9]	]0; 1[	D or C Avg.
A. Same sample (1/A.4)					
Equation (1)					
Older sibling's result	0.0229***	0.0303***	0.0309***	0.0327***	0.0327***
	(2.920)	(7.024)	(8.323)	(9.562)	(9.554)
R^2	0.007	0.006	0.006	0.011	0.046
Equation (4)					
Older sibling's result	-0.00234	0.000572	0.000761	0.000745	0.000745
	(-0.304)	(0.136)	(0.211)	(0.225)	(0.224)
R^2	0.114	0.100	0.100	0.105	0.146
Equation (5)					
Older sibling's result	-0.00371	-0.00183	-0.00176	-0.00157	-0.00157
	(-0.486)	(-0.440)	(-0.493)	(-0.480)	(-0.477)
R^2	0.133	0.121	0.121	0.126	0.165
N	12,518	43,263	74,218	165,626	368,069
B. All sibling pairs					
Equation (1)					
Older sibling's result	0.0242***	0.0308***	0.0309***	0.0327***	0.0327***
	(3.113)	(7.218)	(8.410)	(9.625)	(9.616)
R^2	0.007	0.006	0.006	0.011	0.045
Equation (4)					
Older sibling's result	-0.00176	0.000566	0.000286	0.000269	0.000269
	(-0.231)	(0.136)	(0.0802)	(0.0819)	(0.0815)
R^2	0.115	0.100	0.100	0.105	0.146
Equation (5)					
Older sibling's result	-0.00332	-0.00193	-0.00230	-0.00212	-0.00212
	(-0.440)	(-0.468)	(-0.651)	(-0.653)	(-0.652)
R^2	0.133	0.121	0.121	0.126	0.165
N	12,724	43,963	75,451	168,359	373,923

Notes: standard errors clustered by school; t-statistics in parentheses. Sample limited to 2013-2019 PSLE takers who had a unique last name in their school cohort and who were matched to a single presumable sibling. N is the number of such sibling pairs (the within-column differences in N between panels A and B are explained in Subsection 6.4). Each column presents estimates from a restricted sample: in the first 4 columns, passing probability intervals are used to determine the observations in the sample; in the last column, all older siblings who had a D or C average (and their younger counterparts) are included. The binary dependent variable in all regressions takes value 1 if the younger sibling passed the PSLE and 0 otherwise. All regressions include an intercept; for details regarding the regressors, see Subsection 4.3 and Table 1.

*** p<0.01