Peer Effects in Math and Science*

Juanna Schrøter Joensen
Stockholm School of Economics

Helena Skyt Nielsen
Aarhus University

October 2013

Abstract:

Peer effects may seriously dampen or amplify the effectiveness of policies aimed at increasing the quantity and quality of core skills. In this paper, we investigate the importance of peer effects in the decision to pursue advanced math and science in high school. We exploit quasi-experimental variation stemming from a pilot scheme inducing some older siblings to choose advanced math and science at a lower cost, while not directly affecting the course choices of younger siblings. Therefore, any influences of this pilot scheme on the younger siblings may be attributed to the peer influence of the older sibling. Our results suggest that peer effects among siblings are strongest among closely spaced siblings and their significance depends on the gender composition of the sibling pair. We find the strongest social interaction effects between closely spaced brothers.

JEL Classification: I21, J24.

Keywords: Social interaction, siblings, STEM fields, high school curriculum.

* Contact details: Joensen, Department of Economics, Stockholm School of Economics, email: Juanna.Joensen@hhs.se. Nielsen: Department of Economics and Business, Aarhus University, email: HNielsen@econ.au.dk. We appreciate comments from Maria Knoth Humlum, Eric Maurin, participants at the IWAEE and seminar participants at Oslo University and Sussex University. We thank Kasper Jørgensen and Pernille Hansen for competent research assistance. The usual disclaimers apply.

"[I]f the United States is to maintain its historic pre-eminence in the STEM fields - science, technology, engineering, and mathematics - and gain the social, economic, and national-security benefits that come with such pre-eminence, then we must produce approximately one million more workers in those fields over the next decade than we are on track now to turn out."

S. James Gates Jr., Toll Physics Professor, University of Maryland, and Chad Mirkin, Rathmann Professor of Chemistry, Northwestern University. *Chronicle of Higher Education*, June 25, 2012

1. Introduction

Increasing globalization means that the accumulation of high-quality skills – such as math and science skills – is on the top of the policy agenda in most high-wage countries seeking to be on the forefront of technological progress and sustain economic growth. The demand for college graduates in science, technology, engineering, and mathematics – the STEM fields – far exceeds the supply (President's Council of Advisors on Science and Technology, 2011). A range of reasons for the low supply of such skills have been suggested; such as uninspiring courses, difficulty with the required math, and an academic culture that is unwelcoming (President's Council of Advisors on Science and Technology, 2012). A prerequisite for increasing the supply of STEM graduates is sufficient investment in advanced math and science skills prior to college. Many policies have been suggested to strengthen the college preparedness of high school graduates; including increased course requirements in core subjects like math and science (A Nation at Risk, Gardner et al., 1983). Any policy aiming to increase the investment in math and science skills may be seriously dampened or amplified by social interaction effects, which may be extremely important during the teenage years when decisions on more advanced coursework are taken (Card and Giuliano, 2013; Akerlof, 1997; Akerlof and Kranton, 2002). How important are these peer effects? In this paper, we investigate the importance of such peer effects among sibling pairs based on quasi-experimental variation stemming from a pilot scheme which induced some older siblings to pursue advanced math and science by lowering their cost.

Estimating the causal effect of social interactions is challenging due to simultaneity, correlated unobservables, and endogenous peer group membership (Manski, 1993). We study naturally occurring peer groups and exploit exogenous variation in the cost of taking up advanced math and science in high school among a partial population (Moffitt, 2001).[1] We exploit the fact that some older siblings in 1984-1987 were unexpectedly exposed to a pilot scheme after entering high school

[1] Our study is in this regard similar in spirit to the study of social interaction effects in program participation by Dahl, Løken and Mogstad (forthcoming) and Avvisati, Gurgand, Guyon and Maurin (forthcoming).

.

and investigate whether they influenced the course choices of their younger siblings. In our previous work we employed this pilot scheme to investigate the impact on the individuals themselves. Here we are after a completely different issue, namely the spillover effects on younger siblings who were unexposed themselves. Any influence of this pilot scheme on younger siblings' course choices can be interpreted as a causal peer effect, since the pilot scheme only reduced the cost of choosing advanced math and science for older siblings directly. Younger siblings are 3.5 percentage points more likely to choose math and science if their older sibling was exposed to the pilot scheme. Since the first-stage estimate is 7 percentage points, this implies a peer influence of older siblings on younger siblings of about 50 percentage points.

Siblings are the first peers one closely interacts with and for most they entail a lifelong relationship. Therefore, peer effects from close social interaction between siblings may be extremely important (Buhrmester, 1992). Rigorous economics research on social interactions among siblings is scarce. Butcher and Case (1994) find that the education of females decreases with the presence of any sister in the sib ship, and that this effect gradually vanishes for more recently born cohorts. They argue that the presence of a second daughter in the household changes the reference group of the first daughter. Qureshi (2011) studies education of pairs of siblings in Pakistan. She finds that the education of older sisters improves the education of younger brothers, and she argues that the result reflects improved quality of child care since the older sister takes care of younger siblings. This is evidence in favor of the productivity spillover. Dahl, Mogstand and Løken (forthcoming) document spillover effects in parental leave taking while Monstad, Propper and Salvanes (2011) find spillover effects in teenage pregnancy. However, Adermon (2013) finds no spillover effects among siblings from extending compulsory schooling laws.

The importance of peer effects depends on how individuals make educational choices. In our context of choice of high school course work, we postulate that social interaction effects between siblings may work through four broadly defined channels. We here formulate these mechanisms as explanations for positive spillover effects, though they may as well give rise to the opposite effects. One mechanism may be *information sharing*. A student without a network with peers who previously pursued advanced math and science may face more uncertainty about the difficulty and joy of this course package and about the future prospects of students who complete these courses. On the other hand, having an older sibling who pursued this course package resolves some of this uncertainty. A second mechanism is *productivity spillover effects*. A student with a peer, who *also* studies math and science, may be able to perform better in school in math and science due to assistance with homework. A third mechanism is *conformity* or *norms* where a student gains utility from behaving similarly (or opposite) to specific peers. In some instances, an older sibling or a friend may be a role model and inspiration for academic behavior and aspirations. A fourth mechanism may be *joint leisure*. A student with a peer who *also* studies math and science, may be able to share some public goods with this peer (through a common interest for technological development or nerdy jokes and movies) and may appreciate more spending joint leisure time with this peer. An unwelcoming academic culture may also be more welcoming with a peer sharing ones passion for math and science. In order to shed light on which mechanism is more important, we draw upon psychological literature on the relationship between sibling interactions and sib ship characteristics.

We find strong positive correlations between math and science choices of siblings. Our results suggest that causal peer effects persist among closely spaced siblings, and that their significance depends on the gender composition of the sibling pair. We find the largest and most significant peer effects for relatively closely spaced brothers.

The remainder of the paper unfolds as follows: Section 2 discusses identification of social interaction effects and presents the institutional background which our empirical strategy relies on. Section 3 describes the data, while section 4 presents the empirical analysis of social interaction effects in the choice of math and science in high school. Section 5 investigates heterogeneity in peer effects. Section 6 concludes the paper.

2. Identification of Peer Effects Using a High School Pilot Scheme

This section describes our identification strategy and the educational environment of the Danish high school. In the first subsection, we briefly explain the empirical difficulty of identifying peer effects and how we exploit the unique institutional setup to identify social interaction effects from older to younger siblings. Then we describe the two relevant high school regimes, which form the basis for our identification strategy. The second and third subsections, concern the high school regime and the pilot scheme that provides us with exogenous variation in the cost of acquiring advanced math and science courses for the older siblings. The fourth subsection, concerns the high school regime forming the basis for the math and science choices of their younger siblings.

2.1. *Identifying Peer Effects*

Peer (or social interaction) effects occur when the choice of one individual affects the choices of other individuals in the same peer (or social) group. In this paper, we are interested in how math and science choices of an older sibling affect whether his or her younger sibling pursues advanced math and science courses. The general difficulty of identifying peer effects lies in the empirical issues of: (i) endogenous group membership, (ii) simultaneity (the *reflection* problem), and (iii) correlated unobservables in the peer group.[2] These identification issues can be illustrated in a model which is linear in the peer effect. We assume, without loss of generality, that there are only two individuals in each peer group - an older sibling and a younger sibling.[3]

$$MathScience_{old} = \pi_0 + \pi_1 MathScience_{young} + \pi_2 X_{old} + \pi_3 X_{young} + \pi_4 a_f + \varepsilon_{old,f} \qquad (1)$$

$$MathScience_{young} = \beta_0 + \beta_1 MathScience_{old} + \beta_2 X_{old} + \beta_3 X_{young} + \beta_4 a_f + \varepsilon_{young,f} \qquad (2)$$

where $MathScience_i$ denotes whether sibling i chose advanced math with an advanced science (chemistry or physics) course in high school, X_i denotes observable characteristics of sibling i, a_f denotes sibling pair specific characteristics like family background, gender, and age difference. Finally, $\varepsilon_{i,f}$ denotes other unobserved factors affecting the $MathScience$ choice of individual i in sibling pair *f*.

Our objective is to estimate a causal effect of the older sibling's $MathScience$ choice on the younger sibling's $MathScience$ choice. To be able to give a causal interpretation of the parameter estimate of β_1 in (2) we need to address the empirical issues (i)-(iii) mentioned above. The third issue of correlated unobservables is naturally a big concern in our setting, since siblings share many common social and genetic influences; including common genes, family background, neighborhood, and schools. All these common influences shape both siblings' preferences and abilities and could lead them to making similar high school course choices. An omitted variables bias due to contextual effects arises if we are not able to observe all these relevant sibling pair specific (a_f) and individual variables (X_i). The first and the second issues are presumably minor in our setting: (i) siblings are born into the same family thus do not choose each other based on each other's characteristics and

[2] Manski (1993, 1995) provides a more complete and general analysis of the identification of peer effects (or more generally endogenous effects), while Moffit (2001) introduces the conceptual framework we adopt here.

[3] It is straightforward to generalize this setting to larger peer groups. Brock and Durlauf (2001) discuss identification in nonlinear peer effects models.

choices, and (ii) given the timing of high school course choices it seems plausible that the older sibling's course choice is independent of the younger sibling's choice ($\pi_1 = 0$) since the older sibling makes this choice years before the younger sibling. This exclusion restriction overcomes the reflection problem, as we postulate that the direction of the sibling effect goes from the older sibling to the younger sibling.[4] Nevertheless, this is not a necessary exclusion restriction as our empirical strategy addresses all these three empirical concerns, since the exogenous variation in the cost of acquiring advanced math and science for the older sibling is independent of both sibling pair specific factors and individual sibling characteristics.

More specifically, our identification strategy exploits exogenous variation in the cost of acquiring advanced math and science stemming from a pilot scheme, where some older siblings unexpectedly got the option of a more flexible course combination. Let $PilotIntro_{old} = 0$ for older siblings in a traditional high school, where advanced math and science could *only* be achieved in a package ofadvanced math, advanced physics and intermediate chemistry. Let $PilotIntro_{old} = 1$ for older siblings in a pilot high school, where advanced math and science could *also* be achieved in a package of advanced math, advanced chemistry and intermediate physics. This additional course package option was introduced unexpectedly just before the older sibling made the choice of advanced high school courses. The pilot scheme thus provides us with exogenous variation in the cost of acquiring advanced math and science for the older sibling (captured by $PilotIntro_{old}$) that does not directly influence the younger sibling *and* is independent of any sibling pair specific (a_f) and individual variables (X_i). Substituting this into (1) and (2) we get:

$$MathScience_{old} = \pi_0 + \pi_2 X_{old} + \pi_3 X_{young} + \pi_4 a_f + \gamma PilotIntro_{old} + \varepsilon_{old,f} \quad (3)$$

$$MathScience_{young} = \beta_0 + \beta_1 MathScience_{old} + \beta_2 X_{old} + \beta_3 X_{young} + \beta_4 a_f + \varepsilon_{young,f} \quad (4)$$

Younger siblings attend high school in a regime, where they have an even more flexible curriculum as advanced math and science courses can be combined as they like - the main requirement is that they choose at least two (and at most three) optional advanced courses. This particular institutional setting thus provides us with a unique quasi-experiment for identifying peer effects in math and science - going from the older sibling's course choice to the younger sibling's course choice, as well as enables the possibility of identifying spillover effects. We can thus interpret the IV estimate of β_1

[4] The developmental psychology literature supports that the direction of behavioral influence goes from the older sibling to the younger sibling (Buhrmester, 1992). Altonji et. al (2010) also corroborate this assumption and impose it as an identifying assumption to estimate causal sibling influences on adolescence substance use.

in the structural equation (4) as capturing this causal peer effect when the first-stage equation (3) includes $PilotIntro_{old}$ as an instrument for $MathScience_{old}$ which endogenously affects $MathScience_{young}$, because the instrument only affects the older sibling directly and the younger one merely through endogenous social interaction.[5] The identifying assumptions are corroborated in Joensen and Nielsen (2009) showing that $PilotIntro_{old}$ is independent of predetermined individual, family, and school characteristics for the students entering high school in 1984-87. This implies that older siblings are as good as randomly assigned to high schools who unexpectedly introduce the pilot scheme when they are enrolled in their second high school year. Furthermore, the instrument has a strong influence on the choice of math and science courses for the older sibling. We return to these empirical issues in Section 4. The following subsections describe the educational environment of the two relevant high school regimes: The Pre-1988 High School with restrictive course packages that the older siblings attended and the Post-1988 High School with much more flexible course choices for their younger siblings.

2.2. The Pre-1988 High School

In the period 1961-1988, the Danish high school system was a "branch-based" high school regime in which courses were grouped into restrictive course packages.[6] We focus on the cohorts entering high school in 1984-87. The main reason to focus on this period is that the supply of course packages provides us with relevant exogenous variation in the cost of acquiring advanced math and science for the older siblings.

This regime implied that students upon high school graduation would have achieved one of three math levels available: advanced, intermediate, or basic level. The difference between the three levels is reflected in the number of lessons per week, as well as in the content of the courses. For instance, the extent of geometry and algebra increases as the level becomes more advanced. In the empirical analysis, we focus on whether students choose advanced math and science, meaning that the intermediate and basic level courses are lumped together. The decision about which package to opt for is taken at the end of the first year in high school. The only way to obtain advanced math and science was the package consisting of advanced math, advanced physics and intermediate chemistry, unless the student was enrolled at a pilot school, where the package could be adjusted to include

[5] Moffitt (2001) dubs this type of identifications strategy a partial-population policy intervention.

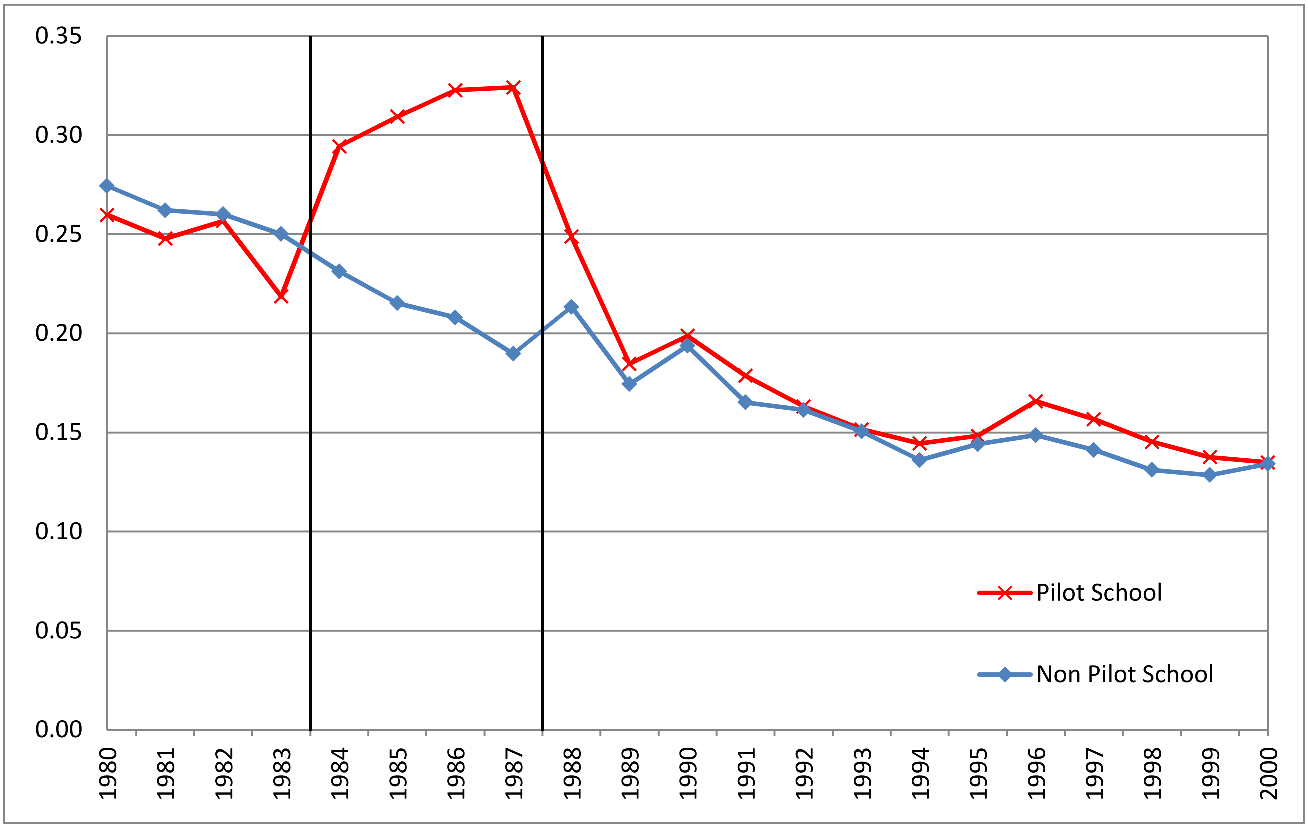
[6] Available course packages were labelled: Social Science and Languages, Music and Languages, Modern Languages, Classical Languages, Math-Social Science, Math-Natural Science, Math-Music, Math-Physics and Math-Chemistry.

advanced chemistry and intermediate physics instead. It is exactly this increased course flexibility which some students were unexpectedly exposed to that constitutes the quasi-experiment we exploit in this paper.

2.3. The Pilot Scheme

The pilot scheme was implemented as an experimental curriculum at about half of the high schools prior to the 1988-reform. The purpose of the pilot scheme was to test the impact of increased flexibility prior to the 1988-reform. Figure 1 illustrates the consequences of the pilot scheme on the course packages of the high school youth. Prior to the pilot scheme, the faction choosing advanced math and science declined and went below 25% in 1983. The pilot scheme counteracted this declining tendency by attracting youth to do the alternative course package with a higher weight on chemistry and a lower weight on physics.

Figure 1. Fraction of High School Cohorts Choosing Math-Science across School Types



Note: Pilot schools include all schools with pilot status at any point in time during 1984-87 (64 schools in total).

Table 1 gives an overview of the gradual implementation of the pilot scheme from 1984-87. The table is divided by types of high schools: schools with no pilot scheme (*PilotSchool=0*), schools where the pilot scheme was introduced after enrollment of the relevant cohort (*PilotSchool=1, PilotIntro=1*), and schools where the pilot scheme was implemented prior to enrollment of the relevant cohort (*PilotSchool=1, PilotIntro=0*).

Table 1. Introduction of the Pilot Scheme

High School	Pilot School = 0			Pilot School = 1 Pilot Intro = 1			Pilot School = 1 Pilot Intro = 0			All		
Cohort	N	MathScience	Schools	N	MathScience	Schools	N	MathScience	Schools	N	MathScience	Schools
1984	10,964	0.2395	123	2,718	0.3282	22	0	0	0	13,682	0.2571	145
1985	9,249	0.2306	109	1,558	0.3434	15	2,663	0.3308	22	13,470	0.2635	146
1986	7,842	0.2132	92	1,526	0.3244	15	4,342	0.3544	37	13,710	0.2703	144
1987	7,223	0.1940	81	1,353	0.2860	12	6,591	0.3333	52	15,167	0.2627	145
Total	35,278	0.2220		7,155	0.3227		13,596	0.3395		56,029	0.2634	

The table displays the fraction of students choosing advanced Math with advanced Physics or Chemistry. The numbers are displayed by entry cohort and type of high school attended.

Schools were not randomly assigned to become pilot schools. Instead, from 1984-86, they could apply to the Ministry of Education for permission to adopt the experimental curriculum, whereas in 1987 the high school principals could make this decision without approval from the ministry.[7] It is not possible to directly test whether the pilot schools represent a sample of schools which is essentially random with respect to math ability, but we corroborate that this is a reasonable approximation.

It is clear, however, that students with a particular preference for chemistry may self-select into schools that are known to offer the pilot program before entrance. This is why we distinguish between students at pilot schools where the pilot scheme was unexpectedly introduced after they had enrolled in the high school (*PilotSchool=1, PilotIntro=1*), and those who knew that the school was a pilot school before they applied for entering the school (*PilotSchool=1, PilotIntro=0*).

The instrumental variable strategy exploits the fact that the pilot scheme reduces the psychological cost of choosing advanced math and science since the students exposed to the scheme are free to choose either advanced physics and intermediate chemistry or advanced chemistry and intermediate

[7] The schools which introduced the program in 1987 tend to be slightly negatively selected in terms of the students' math abilities, while no similar concerns are raised regarding the other cohorts. However, to maintain a large number of sibling pairs, we include the 1987 cohort of older siblings in the study, while checking the sensitivity of our results to leaving out this cohort.

physics.[8] Hence, first-year high school students enrolled at a school when it decided to introduce the pilot scheme were exposed to an unexpected exogenous cost shock, which induced more students to choose advanced math and science compared to students at non-pilot schools. If the selection of newly participating schools is exogenous with respect to student ability, the pilot scheme provides exogenous variation in students' math and science qualifications without influencing the outcomes of interest except through the effect on math and science qualifications.

The instrumental variable, $PilotIntro$, is equal to one if the individual enrolled in a high school which then introduces the experimental curriculum for the first time, and it takes the value zero otherwise. This instrument is valid if the pilot scheme is randomly assigned to schools and if individuals are randomly distributed across schools that have not yet decided to introduce the experimental curriculum. This assumption is violated only if the school decides to participate in the program based on the math abilities of local students. In Section 3 below, we test for similarities of the student and parent bodies across school status, and we find almost no significant differences in characteristics determined pre high school (table to be added).

The instrument is strong if the unexpected introduction of the pilot scheme induces students to choose advanced math and science, which is directly tested and validated in Section 4. The instrument satisfies the monotonicity (or uniformity) condition if individuals who chose advanced math and science when he or she was required to do advanced physics and intermediate chemistry would also have chosen advanced math and science if they had unexpectedly had the option of replacing advanced physics with advanced chemistry and replacing intermediate chemistry with intermediate physics. We are confident that the monotonicity assumption is reasonable in our application, since all the options available at non-pilot schools were also available at schools that introduced the pilot scheme.

Our instrument exploits the exogenous variation in the exposure of students to the option of switching the levels of physics and chemistry. Hence, the "treatment" of the older sibling that we investigate is the combined treatment of advanced math with advanced chemistry and intermediate

[8] Traditionally, the opportunity cost of attending high school is interpreted as earnings forgone from unskilled work. We use a broader interpretation associated with time allocation across courses as well as between studies, leisure, and unskilled work. If students choose course combinations optimally given their preferences and abilities, then a more flexible course choice set reduces the cost of taking a given course as there is a higher probability of a good match between feasible course combinations and the students' preferences and abilities.

physics. We cannot separate the effect of the separate math and science courses from the potential synergy between them.

2.4. The Post-1988 High School

In 1988 there was an extensive structural reform of the Danish High School, which was the most fundamental high school reform since 1903. The reform abolishes the "branch based" regime and substitutes it with a "choice based" regime, where the main distinction is between mathematical and linguistic track students. The reform implied an extended choice set in the form of more flexible opportunities to combine optional courses.[9] In particular, the mathematical students have the option of combining advanced math with any other advanced course; for example physics, chemistry, biology, social science, or a language course. This is the regime within which the younger siblings in our sample make their educational choices. We focus on the younger siblings' choice of advanced math with advanced physics and/or advanced chemistry, since these are comparable to the relevant course combinations for the older sibling attending high school in the pre-1988 regime. As a robustness check, we also study the younger sibling's choice of advanced math with other advanced courses.

The structure of the post-1988 high school regime is as follows: Students choose either the mathematical or the linguistic track upon entry. Each course is either common to all students on the chosen track (compulsory courses), compulsory for some and optional for others, or exclusively optional. The optional courses can be obtained at either advanced or intermediate level reflecting the complexity of the content, the number of lessons per week and the intensity of exams (written and/or oral). Furthermore, students write an elaborate term paper in one of the advanced courses in the third and final year.

All students are required to follow at least two (and at most three) optional advanced courses, and for the mathematical students there was a minimum required amount of math-science content, while for the linguistic students there was a minimum required amount of language content. The first year of high school consists only of compulsory courses (common as well as track-specific courses) taught in classes of at most 28 students. The second and third year of high school added at least three

[9] The reform also implied more weight on the high schools' role of preparing students for college, more required readings, more written assignments, more stringent non-attendance regulation, more grading, and more hours of instruction allocated to the compulsory courses.

and at most four optional courses.[10]. In addition to the requirements of at least two advanced optional courses, there were some bonds between some courses in order to preserve the possibility for the courses to complement each other.

We follow younger siblings in this high school regime until the entry cohort of 1997 and focus on the younger siblings' choice of advanced math with either advanced physics or advanced chemistry, since these are comparable to the relevant course combinations for the older sibling attending high school in the pre-1988 regime.[11] Thus $MathScience_{young}$ in equation (4) is an indicator for whether the younger sibling chooses to combine advanced math with either advanced physics or advanced chemistry.

[10] The compulsory courses common to all students are advanced Danish and history, intermediate English and basic physical education, biology, geography, religion, music, (visual) art, and ancient history. Track-specific compulsory courses for mathematical students comprise intermediate math and physics, basic chemistry, and a second foreign language. For the linguistic students the track-specific compulsory courses are basic natural sciences (including math) and Latin, as well as two other foreign languages. Commonly available optional intermediate courses comprise: biology, geography, chemistry, technical science, business and economics, drama, sports, and movie science, while optional advanced courses include all feasible continuations of the intermediate courses.

[11] Some curriculum changes are introduced with the reform, e.g. a historical dimension was incorporated into the math course while some advances in the experimental direction were incorporated into the physics course.

3. Data Description

3.1. Sample Selection

For our empirical analysis we use a panel data set comprising the population of individuals starting high school from 1984 and onwards. The data are administered by Statistics Denmark, which has gathered the data from administrative registers. The data set includes basic demographic information such as date of birth, place of residence, and gender. What is crucial for this study is that we observe which institutions offered the pilot scheme when, and we can identify which institution the individual attended as well as the chosen course package. Furthermore, we have information about the dates for entering and exiting a high school education, along with an indication of whether the individual completed the education successfully, dropped out, or is still enrolled as a student. We augment this data set with background information about the parents including educational achievement and gross income. This information is recorded when the individual was 15-years old, which is prior to enrolling in high school.

The core sample consists of individuals who are directly influenced by the quasi-experimental variation due to the gradual introduction of the pilot scheme for cohorts entering high school 1984-1987. From this sample, only high school graduates who finished in three years[12] and who have a younger sibling who entered high school after 1987 are selected. An overview of the sample selection procedure is given in Table 2.

We construct two estimation samples each of which imposes a minimum amount of homogeneity on the sample by setting a maximum limit on the age gap between siblings. One sample includes only closely spaced sibling pairs (cohorts 1988-91, age gap $\leq$ 4 years) and one includes also widely spaced sibling pairs (cohorts 1988-97, age gap $\leq$ 10 years). The closely spaced sample consists of 18,846 sibling pairs (involving 16,592 older siblings), while the closely spaced sample consists of 8,259 sibling pairs (involving 8,097 older siblings).

[12] This restriction is still to be implemented. About 40% of a birth cohort attended the academic high school track at this point in time; hereof 10% do not complete in three years. The main part of drop out takes place before the choice of advanced math and science course packages. Dropout is uncorrelated with pilot school status.

Table 2. Overview of the Sample Selection.

	N
All high school cohorts 1984-87	68,408
with younger siblings	
sibling pairs	57,798
accounting for older siblings once	40,176
with younger siblings in high school	
sibling pairs	26,518
accounting for older siblings once	21,648
with younger siblings in high school cohorts 1988-97	
sibling pairs	20,006
accounting for older siblings once	17,355
and age difference < 10 years	
sibling pairs	**18,846**
accounting for older siblings once	16,592
with younger siblings in high school cohorts 1988-91	
sibling pairs	12,883
accounting for older siblings once	12,305
and age difference < 4 years	
sibling pairs	**8,259**
accounting for older siblings once	8,097

In Table 3, we describe the distribution of sibling pairs across the older siblings' exposure to the pilot scheme for each high school cohort of younger siblings. As expected, older siblings who were exposed - expectedly or unexpectedly - to the pilot program were much more likely to choose advanced math and science. However, it is evident from the table that the probability that the older sibling choose this course package goes down as the age distance between siblings increases; among cohorts 1988-91, more than 30% choose this course package, while the number is down to around 27% for cohorts 1996-97.

Table 3. Summary of Older Siblings' Course Choice by their Exposure to the Pilot Scheme and by High School Cohort of Younger Sibling

Younger Sib High School Cohort	**Older Sibling** Pilot School = 0 Pilot Intro = 0		**Older Sibling** Pilot School = 1 Pilot Intro = 1		**Older Sibling** Pilot School = 1 Pilot Intro = 0		All	
	N	MathScience old	N	MathScience old	N	MathScience old	N	MathScience old
1988	2,282	0.2875	457	0.3260	723	0.3582	3,462	0.3073
1989	2,313	0.2763	454	0.3370	1,005	0.4199	3,772	0.3218
1990	1,921	0.2801	344	0.3983	863	0.4322	3,128	0.3350
1991	1,480	0.2662	276	0.2899	765	0.3895	2,521	0.3062
1992	1,094	0.2386	230	0.3087	534	0.3558	1,858	0.2809
1993	938	0.2633	174	0.3678	435	0.3655	1,547	0.3038
1994	712	0.2303	133	0.2632	351	0.3875	1,196	0.2801
1995	663	0.2594	124	0.3306	274	0.4015	1,061	0.3044
1996	489	0.2249	96	0.2917	235	0.3660	820	0.2732
1997	408	0.2574	81	0.2840	152	0.2829	641	0.2668
Total	12,300	0.2672	2,369	0.3297	5,337	0.3890	20,006	0.3071
1988-1997	11,452	0.2676	2,220	0.3329	5,174	0.3902	18,846	0.3090
1988-1991	4,759	0.2746	900	0.3433	2,600	0.4042	8,259	0.3229

The table displays the number of younger siblings and the fraction of their older siblings choosing advanced Math with advanced Physics or Chemistry. The numbers are displayed by younger siblings' high school entry cohort and type of high school attended by the older sibling. The two rows at the bottom summarize information for cohorts 1988-97 (age gap ≤ 10 years) and for cohorts 1988-91 (age gap ≤ 4 years).

3.2. Outcome and Control Variables

The outcome of interest is whether the peers in the post-reform era choose the course package consisting of advanced math and science or not.

At the top of Table 4, we see that there is a strong correlation in the choice of this course package across siblings: 27 % (13 %) of younger siblings chose this course package when the older sibling did (did not) choose this package for the widely spaced sample, and the correlation varies across gender composition of the sib ship. Also when it comes to parental background, there is a significant difference. Parents of sibling pairs where the older sibling chose advanced math and science have a higher education and fathers have a higher annual income (to be added to table).

At the top of Table 5, we see that there is some variation in the choice of advanced math and science when we distinguish between whether the older sibling was exposed to the pilot scheme or not. The proportion of younger siblings who chose this course package is 16.8 % when the older sibling was not exposed and 17.7% when the older sibling was unexpectedly exposed in the widely spaced sample, and 17.9 % and 21.4 % in the closely spaced sample. The relationship appears to me very strong among pairs of brothers in the closely spaced sample. We note from the table that the parental background is similar across pilot school status (to be added to table).

As control variables we include parental background as well as entry cohort fixed effects and high school fixed effect. Parental background includes a set of mutually exclusive indicator variables for the level of highest completed education of the mother and father, respectively, and their income as observed at the end of the year before the individual started high school. We leave out post-graduation control variables and thus estimate the total effect of advanced math.

Table 4. Descriptive Statistics by Course Choice of Older Sibling

PANEL A: Cohorts 1988-97 (age gap ≤ 10 years)

Younger Sibling		Older Sibling MathScience = 0			Older Sibling MathScience = 1			All		
Gender	**Course Choice**	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
					All					
All	MathScience	13,023	0.1300	0.3363	5,823	0.2679	0.4429	18,846	0.1726	0.3779
					Older Brother					
Brother	MathScience	2,253	0.1806	0.3848	1,904	0.4070	0.4914	4,157	0.2843	0.4512
Sister	MathScience	2,592	0.0436	0.2042	2,190	0.1297	0.3360	4,782	0.0830	0.2759
					Older Sister					
Brother	MathScience	3,478	0.2461	0.4308	782	0.4233	0.4944	4,260	0.2786	0.4484
Sister	MathScience	4,700	0.0674	0.2508	947	0.1795	0.3840	5,647	0.0862	0.2807

PANEL B: Cohorts 1988-91 (age gap ≤ 4 years)

Younger Sibling		Older Sibling MathScience = 0			Older Sibling MathScience = 1			All		
Gender	**Course Choice**	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
					All					
All	MathScience	5,592	0.1425	0.3496	2,667	0.2820	0.4500	8,259	0.1876	0.3904
					Older Brother					
Brother	MathScience	932	0.1985	0.3991	882	0.4172	0.4934	1,814	0.3049	0.4605
Sister	MathScience	1,114	0.0422	0.2011	971	0.1390	0.3462	2,085	0.0873	0.2823
					Older Sister					
Brother	MathScience	1,517	0.2808	0.4495	371	0.4259	0.4951	1,888	0.3093	0.4623
Sister	MathScience	2,029	0.0685	0.2527	443	0.2054	0.4045	2,472	0.0930	0.2906

Note: Bold and italics indicate that the mean for MathScience=0 is significantly different from the mean for MathScience=1 at the 5 % and the 10% level, respectively.

Table 5. Descriptive Statistics by Older Sibling's Exposure to the Pilot Scheme

PANEL A: Cohorts 1988-97 (age gap ≤ 10 years)

Younger Sibling		**Older Sibling** Pilot School = 0 Pilot Intro = 0			**Older Sibling** Pilot School = 1 Pilot Intro = 1			**Older Sibling** Pilot School = 1 Pilot Intro = 0			**All**		
Gender	**Course Choice**	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
		All											
All	MathScience	11,452	0.1684	0.3742	2,220	0.1770	0.3818	5,174	0.1801	0.3843	18,846	0.1726	0.3779
		Older Brother											
Brother	MathScience	2,542	0.2821	0.4501	460	0.3196	0.4668	1,155	0.2753	0.4469	4,157	0.2843	0.4512
Sister	MathScience	2,891	0.0771	0.2669	553	0.0850	0.2791	1,338	0.0949	0.2932	4,782	0.0830	0.2759
		Older Sister											
Brother	MathScience	2,567	0.2778	0.4480	529	0.2609	0.4395	1,164	0.2887	0.4533	4,260	0.2786	0.4484
Sister	MathScience	3,452	0.0797	0.2708	678	0.0900	0.2864	1,517	0.0995	0.2995	5,647	0.0862	0.2807

PANEL B: Cohorts 1988-91 (age gap ≤ 4 years)

Younger Sibling		**Older Sibling** Pilot School = 0 Pilot Intro = 0			**Older Sibling** Pilot School = 1 Pilot Intro = 1			**Older Sibling** Pilot School = 1 Pilot Intro = 0			**All**		
Gender	**Course Choice**	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
		All											
All	MathScience	4,759	0.1792	0.3836	900	0.2144	0.4107	2,600	0.1935	0.3951	8,259	0.1876	0.3904
		Older Brother											
Brother	MathScience	1,055	0.2967	0.4570	196	0.3878	0.4885	563	0.2913	0.4548	1,814	0.3049	0.4605
Sister	MathScience	1,174	0.0792	0.2702	214	0.0841	0.2782	697	0.1019	0.3027	2,085	0.0873	0.2823
		Older Sister											
Brother	MathScience	1,075	0.2958	0.4566	223	0.3363	0.4735	590	0.3237	0.4683	1,888	0.3093	0.4623
Sister	MathScience	1,455	0.0887	0.2843	267	0.0899	0.2866	750	0.1027	0.3037	2,472	0.0930	0.2906

Note: Bold and italics indicate that the mean is significantly different from the mean for Pilot School=0 & Pilot Intro=0 at the 5 % and the 10% level, respectively.

4. Main Results

In Table 6 we present the main results from the empirical analysis. The OLS regressions indicate a strong positive association between math and science course choices of older and younger siblings. The IV estimates suggest that there is also a causal influence of older siblings' course choices, but only when the age distance between the siblings is less than four years the effect is statistically significant. The magnitude of the estimate is 0.5 which suggests a very strong peer effect. Including additional control variables does not significantly affect our point estimates, lending additional support to our exclusion restriction and exogeneity of *PilotIntro*.

The IV point estimates are larger than the OLS estimates, although not significantly so. However, it suggests that older siblings who are at the margin of choosing math science are more influential for their younger siblings than others. This is consistent with sibling competition: if the older sibling is an always taker ("a math science star"), the younger sibling would be more reluctant to compete than if the older sibling is on the margin of choosing math science.

Table A1 in Appendix A presents the results when the maximum age gap is held constant and the cohorts are allowed to vary. These results confirm that the results vary with age gap and not with cohorts.

Table 6. Estimates of Peer Effects: Main Results

		Parameter Estimates (Standard Errors)					
		OLS			IV		
	N Sibling Pairs	(1)	(2)	(3)	(4)	(5)	(6)
Effect on Younger Sibling *MathScience*:							
Older Sibling *MathScience*		0.140 ***	0.133 ***	0.131 ***	0.512 **	0.433 **	0.487 *
Younger Sibling 1988-91, ≤4y	8.259	(0.01)	(0.01)	(0.01)	(0.22)	(0.21)	(0.27)
Older Sibling *MathScience*		0.145 ***	0.137 ***	0.133 ***	0.326 *	0.265	0.211
Younger Sibling 1988-92, ≤5y	11.230	(0.01)	(0.01)	(0.01)	(0.19)	(0.18)	(0.22)
Older Sibling *MathScience*		0.143 ***	0.135 ***	0.133 ***	0.296	0.239	0.274
Younger Sibling 1988-93, ≤6y	13.537	(0.01)	(0.01)	(0.01)	(0.18)	(0.17)	(0.20)
Older Sibling *MathScience*		0.138 ***	0.131 ***	0.128 ***	0.133	0.145	0.195
Younger Sibling 1988-97, ≤10y	18.846	(0.01)	(0.01)	(0.01)	(0.13)	(0.12)	(0.15)
Older Sibling *MathScience* (First-Stage):							
PilotIntro					0.069 ***	0.069 ***	0.067 ***
Younger Sibling 1988-91, ≤4y					(0.02)	(0.02)	(0.02)
PilotIntro					0.064 ***	0.064 ***	0.061 ***
Younger Sibling 1988-92, ≤5y					(0.01)	(0.01)	(0.02)
PilotIntro					0.059 ***	0.059 ***	0.060 ***
Younger Sibling 1988-93, ≤6y					(0.01)	(0.01)	(0.01)
PilotIntro					0.065 ***	0.067 ***	0.063 ***
Younger Sibling 1988-97, ≤10y					(0.01)	(0.01)	(0.01)
Additional control variables:							
Gender			+	+		+	+
Parental variables (for mother and father):							
Highest Completed Education and Income			+	+		+	+
Fixed effects							
Entry Cohort Fixed Effects			+	+		+	+
High School Fixed Effects				+			+

Note: Significance at a 1%, 5%, and 10% level are denoted by ***, ** and *, respectively.

5. Understanding Heterogeneity in Peer Effects

In this section, we seek to better understand heterogeneity in peer effects. We explore differences in peer effects across gender, parental background, sib ship composition as well as across strength of ties. These heterogeneous effects lead towards inference about plausible causal mechanisms.

5.1. Heterogeneity by Gender Composition of the Sib Ship

Already the descriptive statistics revealed heterogeneous patterns across gender composition of the sib ship. In Table 5 we saw that younger siblings are more likely to choose *MathScience* if the older sibling was unexpectedly exposed to the pilot scheme, and that this relationship appears to be particularly strong for boy-boy sib ships. This pattern already suggests that older brothers influence younger brothers greatly.

Table 7 presents estimates of spillover effects by gender composition of the sib ship. The top panel shows spillover effects from older brothers and sisters to younger *sisters*, while the bottom panel shows spillover effects from older brothers and sisters to younger *brothers*. The striking conclusion from this table is that the significantly positive coefficient in Table 6 is entirely driven by the strong influence of older brothers on their younger brothers. The association as estimated from the OLS is much larger than for sib ships of other gender compositions while the causal impact as estimated by IV is large and statistically significant while none of the other sib ships show such a relationship. The table reveals that one reason for this pattern is that the first stage estimate is stronger (z-stat >2.4) and more influential (coefficient = 0.15) for older brothers than for older sisters. The results are smaller in magnitude but qualitatively unchanged for the sample of widely spaced sibling pairs (not shown).

Table 7. Estimates of Peer Effects: Heterogeneous Gender Composition, Cohorts 1988-91 (age gap ≤ 4 years)

	Parameter Estimates (Standard Errors)							
	OLS		IV		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Older Brother				Older Sister			
Effect on Younger Sister *MathScience:*								
Older Sibling *MathScience*	0.108 ***	0.106 ***	-0.305	-0.410	0.138 ***	0.139 ***	0.394	0.539
	(0.02)	(0.02)	(0.63)	(0.80)	(0.02)	(0.02)	(0.44)	(0.50)
Older Sibling *MathScience* (First-Stage):								
PilotIntro			0.107 **	0.053			0.066 *	0.063 *
			(0.04)	(0.06)			(0.04)	(0.04)
Number of Sibling Pairs	**2,085 (Older Brother, Younger Sister)**				**2,472 (Older Sister, Younger Sister)**			
Effect on Younger Brother *MathScience:*								
Older Sibling *MathScience*	0.189 ***	0.190 ***	0.920 **	0.818 *	0.121 ***	0.111 ***	-0.454	0.323
	(0.03)	(0.03)	(0.46)	(0.45)	(0.03)	(0.03)	(1.51)	(1.12)
Older Sibling *MathScience* (First-Stage):								
PilotIntro			0.145 **	0.145 **			0.037	0.045
			(0.06)	(0.06)			(0.05)	(0.05)
Number of Sibling Pairs	**1,814 (Older Brother, Younger Brother)**				**1,888 (Older Sister, Younger Brother)**			
Additional control variables:								
Parental variables (for mother and father):								
Highest Completed Education and Income		+		+		+		+
Fixed effects								
Entry Cohort Fixed Effects		+		+		+		+
High School Fixed Effects		+		+		+		+

Note: Significance at a 1%, 5%, and 10% level are denoted by ***, ** and *, respectively.

5.2. Heterogeneity by Parental Background

In Table 8 we investigate heterogeneity of the peer effect according to parental background. We define an indicator variable for whether the income of the father is in the top quartile and include an interaction term in the first and second stages. Similarly, we define an indicator variable for whether at least one parent is educated in a STEM field according to a narrow and a broad definition, respectively. We find that the older sibling tends to repond less to the pilot scheme when the father has a high income or when a parent is educated in a STEM field, however none of the effects are statistically significant.

Table 8. Estimates of Peer Effects: Heterogeneous Parental Background, Cohorts 1988-91 (age gap ≤ 4 years)

	Parameter Estimates (Standard Errors)					
	Father's income in top quartile		**Parent in STEM (narrow)**		**Parent in STEM (broad)**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
Effect on Younger Sibling *MathScience:*						
Older Sibling *MathScience*	0.125 ***	0.486	0.128 ***	0.481	0.130 ***	0.541
	(0.01)	(0.40)	(0.01)	(0.31)	(0.01)	(0.35)
Interaction	0.019	-0.333	-0.008	-0.355	-0.004	-0.407
	(0.02)	(0.39)	(0.02)	(0.31)	(0.02)	(0.35)
Older Sibling *MathScience* (First-Stage):						
PilotIntro		0.047 **		0.060 ***		0.055 **
		(0.02)		(0.02)		(0.02)
Interaction		-0.027		-0.031		-0.035
		(0.03)		(0.03)		(0.03)
Full set of control variables:	+	+	+	+	+	+
Number of Sibling Pairs	**8,259**					

Note: The narrow definition of STEM fields follows the definition by the US Department of Homeland Security (DHS), which almost entirely includes Math, Engineering, Natural and Technical Sciences, while the broad definition follows the definition by the National Science Foundation (NSF), which also includes some social sciences and life sciences.

5.3. Heterogeneity by Birth Order and Size of the Sib Ship

The previous section (Section 4) already indicated that peer effects may be stronger when the age difference was limited to four years than otherwise. This could suggest that a smaller age difference implies closer ties, but it could also reflect more sibling rivalry or stronger role model effects among closely tied siblings. Unfortunately, the data material does not allow us to draw more detailed inference about the importance of sibling spacing closer than 4 years. In this section we investigate how the effects vary with birth order and size of the sibling ships.

Table 9 shows the results for three separate subsamples of sibling pairs. The first set of columns shows the results for pairs where the oldest sibling is also firstborn. In this case the instrument is weaker, but the point estimate of the peer effect is larger. The second set of columns shows that the effect disappears for older siblings who are second or later born. The third set of columns reveals that the effect is unchanged whether the younger sibling is lastborn or not. We have also divided the sample into families with two siblings versus more siblings which reveal no difference.

Table 9. Estimates of Peer Effects: Heterogeneity by Rank Order in the Sib Ship, Cohorts 1988-91 (age gap ≤ 4 years)

	Parameter Estimates (Standard Errors)					
	Older sibling, first born		**Older sibling, second or later born**		**Younger sibling, last born**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
Effect on Younger Sibling *MathScience:*						
Older Sibling *MathScience*	0.136 ***	1.133	0.115 ***	0.025	0.139 ***	0.528 *
	(0.01)	(0.81)	(0.01)	(0.21)	(0.01)	(0.27)
Older Sibling *MathScience* (First-Stage):						
PilotIntro		0.035 *		0.165 ***		0.078 **
		(0.02)		(0.04)		(0.02)
Full set of control variables:	+	+	+	+	+	+
Number of Sibling Pairs	**6525**		**1734**		**5598**	

Note: Significance at a 1%, 5%, and 10% level are denoted by ***, ** and *, respectively.

5.4. *Strength of Ties*

Granovetter (1973) distinguishes between strong and weak ties as defined by the overlap in network members. In practice the strength of ties are related to the nature and the duration of the relationship as well as the frequency and intensity of interactions between individuals.

In our context of sibling pairs, ties are most likely stronger among closely spaced siblings, among sibling pairs who share both parents and among sibling pairs who grow up together. We explore heterogeneity in peer effects across these characteristics. Furthermore, we investigate if spillover effects exist among class mates as well as among siblings.

While weak ties play an important role in Granovetter's setting because they link small well-defined groups, strong ties are likely to matter the most in our setting. Peers with strong ties are more likely to interact with each other and trust each other's opinions, increasing the chance that information will be transmitted and acted upon. Thus, when making important education and career decisions, strong ties are very influential, while weak ties play a role in job search later on when the benefit of these choices are being reaped.

Table to be added.

5.5. *What Is the Mechanism?*

In order to shed light on which mechanism is more important, we draw upon psychological literature on the social interaction and sib ship size and composition (see Buhrmester, 1992). The importance of birth order was first mentioned by Adler (1927) while Adams (1972), who suggested that second and middle children would often try to catch up with first child and thus compete, while the youngest child would do so less. Conley (2000) has stressed that same-sex sibling ships are more competitive and achievement-oriented than other sibling ships, and in particular if they consist of two boys. Adams (1972) suggest that sib ships that are closely spaced (less than five years apart) are more competitive, while siblings who are more than five years are part tend to work like separate sib ships. Thus, it seems that sibling rivalry and competition is a common denominator which may be important among closely spaced pairs of brothers.

For several reasons it makes sense that high school course choice reflects competitive actions. Various characteristics of math, in particular, but to some extent also science, suggest that it is a competitive discipline (Niederle and Vesterlund, 2010). In the discipline of math, answers are either right or wrong, which makes it easier to claim victory. Furthermore, math skills predict future performance very well, which means that the gains from excellent performance may be sizeable. Finally, the math discipline is dominated by males who are known to be attracted to competition, while females tend to shy away from mixed-sex competition and to do worse in high-stake mixed-sex competition.

6. Robustness and Sensitivity Checks

6.1. Other Channels than Sibling Peer Effects

To be written

6.2. Placebo Tests

To be written

7. Conclusion

In this paper we investigate the importance of peer effects in the decision to pursue advanced math and science in high school. Any policy seeking to induce more students to pursue the highly demanded STEM fields could be seriously dampened or amplified by such social interaction effects. We exploit quasi-experimental variation stemming from a pilot scheme in place in the eighties in Denmark. While the pilot scheme induced some older siblings to pursue advanced math and science and not others, it did not directly influence the course choices of younger siblings or their high school peers. Therefore, any influence of this scheme on the younger siblings or their peers may be attributed to the influence of the older sibling. Our results suggest that spillover effects are stronger among closely spaced siblings, and that their magnitude depends on the gender composition of the sibling pair. Our research agenda will seek to better understand the nature of these gender and age differences, as well as exploring birth order effects.

References

Adermon, A. (2013), *Essays on the Transmission of Human Capital and the Impact of Technological Change*. Department of Economics, Uppsala University. Economics Studies 135. ISBN 978-91-85519-42-2. ISSN 0283-7668.

Akerlof, G. (1997), Social Distance and Social Decisions", *Econometrica* 65(5), 1005-1027.

Akerlof, G. and R. Kranton (2002), Identity and Schooling: Some Lessons for the Economics of Education *Journal of Economic Literature* 40(4): 1167-1201.

Altonji, J. (1995), The Effect of High School Curriculum on Education and Labor Market Outcomes. *Journal of Human Resources* 30(3): 409-438.

Altonji, J. G., E. Blom and C. Meghir (2012), Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics* 4: 185-223.

Altonji, J. G., S. Cattan and I. Ware (2010), Identifying Sibling Influence on Teenage Substance Use. NBER WP# 16508.

Avvisati, F., M. Gurgand, N. Guyon and E. Maurin (forthcoming), Getting Parents Involved: A Field Experiment in Deprived Schools. Forthcoming in *Review of Economic Studies.*

Brock, W. A., and S. N. Durlauf, Interactions-Based Models (vol. 5, pp. 3297–3380), in J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics* (Amsterdam: North-Holland, 2001).

Buhrmester, D. (1992): The Developmental Courses of Sibling and Peer Relationships, in F. Boer and J. Dunn (Eds.), *Children's Sibling Relationships: Developmental and Clinical Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, D. and L. Giuliano (2013), Peer Effects and Multiple Equilibria in the Risky Behavior of Friends. *Review of Economics and Statistics* 95(4) 1130-1149.

Conley, D. (2000), Sibship, Sex Composition and the Educational Attainment of Men and Women. *Social Science Research* 29: 441-457.

Conley, D. (2005), The Pecking Order: A Bold New Look at How Family and Society Determine Who We Become. Vintage.

Dahl, G., K. V. Løken & M. Mogstad (forthcoming), Peer Effects in Program Participation. Forthcoming in *American Economic Review*.

Gardner, D. P. and the National Commission on Excellence in Education (1983), A Nation at Risk: The Imperative for Educational Reform.

Granovetter, M. S. (1973), The Strength of Weak Ties. *American Journal of Sociology*. 1360-1380.

Joensen, J. S. and H. S. Nielsen (2009), Is there a Causal Effect of High School Math on Labor Market Outcomes? *Journal of Human Resources* 44(1): 171-198.

Manski, C. F. (1993), Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

Manski, C. (1995), *Estimation Problems in the Social Sciences*. Cambridge, Massachusetts: Harvard University Press.

Moffitt, R. A. (2001), Policy interventions, low-level equilibria, and social interactions. In S. N. Durlauf and H. P. Young (eds.), *Social Dynamics*. Cambridge: MIT Press, pp. 45–82.

Monstad, K., C. Propper and K. G. Salvanes (2011), Is teenage motherhood contagious? Evidence from a Natural Experiment. NHH Discussion Paper SAM 12-2011.

President's Council of Advisors on Science and Technology (2011), *Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future*. Executive Office of the President, Washington DC.

President's Council of Advisors on Science and Technology (2012), *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. Executive Office of the President, Washington DC.

Qureshi, J. (2011), Additional Returns to Investing in Girls' Education: Impact on Younger Sibling Human Capital. Manuscript, University of Illinois at Chicago.

Rose, H. and J. Betts (2004), The Effect of High School Courses on Earnings. *Review of Economics and Statistics* 86(2): 497-513.

Stinebrickner, R. and T. Stinebrickner (2013), A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. Forthcoming in *Review of Economic Studies.*

Appendix A. Additional Results

Table A1. Estimates of Peer Effects, Cohort Variation

		Parameter Estimates (Standard Errors)					
		OLS			IV		
	N Sibling Pairs	(1)	(2)	(3)	(4)	(5)	(6)
Effect on Younger Sibling *MathScience:*							
Older Sibling *MathScience*		0.141 ***	0.133 ***	0.129 ***	0.515 **	0.428 **	0.415 *
Younger Sibling 1988-90, ≤4y	7.486	(0.01)	(0.01)	(0.01)	(0.20)	(0.19)	(0.22)
Older Sibling *MathScience*		0.140 ***	0.133 ***	0.131 ***	0.512 **	0.433 **	0.487 *
Younger Sibling 1988-91, ≤4y	8.259	(0.01)	(0.01)	(0.01)	(0.22)	(0.21)	(0.27)
Older Sibling *MathScience*		0.139 ***	0.132 ***	0.130 ***	0.531 **	0.448 **	0.488 *
Younger Sibling 1988-92, ≤4y	8.360	(0.01)	(0.01)	(0.01)	(0.22)	(0.21)	(0.27)
Older Sibling *MathScience*		0.139 ***	0.132 ***	0.131 ***	0.527 **	0.446 **	0.491 *
Younger Sibling 1988-93, ≤4y	8.373	(0.01)	(0.01)	(0.01)	(0.22)	(0.21)	(0.27)
Older Sibling *MathScience* (First-Stage):							
PilotIntro					0.079 ***	0.079 ***	0.083 ***
Younger Sibling 1988-90, ≤4y					(0.02)	(0.02)	(0.02)
PilotIntro					0.069 ***	0.069 ***	0.069 ***
Younger Sibling 1988-91, ≤4y					(0.02)	(0.02)	(0.02)
PilotIntro					0.069 ***	0.070 ***	0.066 ***
Younger Sibling 1988-92, ≤4y					(0.02)	(0.02)	(0.02)
PilotIntro					0.070 ***	0.070 ***	0.066 ***
Younger Sibling 1988-93, ≤4y					(0.02)	(0.02)	(0.02)
Additional control variables:							
Gender			+	+		+	+
Parental variables (for mother and father):							
Highest Completed Education and Income			+	+		+	+
Fixed effects							
Entry Cohort Fixed Effects			+	+		+	+
High School Fixed Effects				+			+

Significance at a 1%, 5%, and 10% level are denoted by ***, ** and *, respectively.