DISCUSSION PAPER SERIES

DP14205
(v. 2)

High School Rank in Math and English and the Gender Gap in STEM

Judith Delaney and Paul J. Devereux

LABOUR ECONOMICS

ISSN 0265-8003

High School Rank in Math and English and the Gender Gap in STEM

Judith Delaney and Paul J. Devereux

Discussion Paper DP14205
First Published 17 December 2019
This Revision 06 February 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Labour Economics

High School Rank in Math and English and the Gender Gap in STEM

Abstract

Using unique data on preference rankings for all high school students who apply for college in Ireland, we investigate whether, conditional on absolute achievement at the end of high school, within school-cohort rank in English and math affects choice of college major. We find that higher rank in math increases the likelihood of choosing STEM and decreases the likelihood of choosing Arts and Social Sciences. Similarly, a higher rank in English leads to an increase in the probability of choosing Arts and Social Sciences and decreases the probability of choosing STEM. The effects of subject ranks on STEM are larger for boys than girls while there is no evidence of a gender difference in the effect of subject ranks on Arts and Social Sciences. We also find that English and math rank can explain about 4% of the gender gap in the choice of STEM as a college major and 9% of the gender gap that is not explained by absolute achievement. Overall, the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 10% of the difference in the STEM gender gap between mixed-sex schools and same-sex schools and about 25% of the difference that is unexplained by absolute achievement. Notably, these effects occur even though within-school rank plays no role whatsoever in college admissions decisions. Overall, the findings imply behavioral effects of subject rank on college major choices that go beyond their effects on human capital accumulation in school.

JEL Classification: I20, I23

Keywords: High school rank, STEM, college major choice, Gender Gap, comparative advantage, gender gap in STEM

Judith Delaney - judithmdelaney@gmail.com
University of Bath

Paul J. Devereux - devereux@ucd.ie
University College Dublin and CEPR

High School Rank in Math and English and the Gender Gap in STEM[*]

by

Judith M. Delaney
University of Bath
University College London and IZA
judithmdelaney@gmail.com

Paul J. Devereux
School of Economics and Geary Institute, University College Dublin
NHH, CEPR, and IZA
devereux@ucd.ie

September 2020

Abstract

Using unique data on preference rankings for all high school students who apply for college in Ireland, we investigate whether, conditional on absolute achievement at the end of high school, within school-cohort rank in English and math affects choice of college major. We find that higher rank in math increases the likelihood of choosing STEM and decreases the likelihood of choosing Arts and Social Sciences. Similarly, a higher rank in English leads to an increase in the probability of choosing Arts and Social Sciences and decreases the probability of choosing STEM. The effects of subject ranks on STEM are larger for boys than girls while there is no evidence of a gender difference in the effect of subject ranks on Arts and Social Sciences. We also find that English and math rank can explain about 4% of the gender gap in the choice of STEM as a college major and 9% of the gender gap that is not explained by absolute achievement. Overall, the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 10% of the difference in the STEM gender gap between mixed-sex schools and same-sex schools and about 25% of the difference that is unexplained by absolute achievement. Notably, these effects occur even though within-school rank plays no role whatsoever in college admissions decisions. Overall, the findings imply behavioral effects of subject rank on college major choices that go beyond their effects on human capital accumulation in school.

[*] We are grateful to the Central Applications Office for providing access to the data used in this paper and to the State Examinations Commission for helpful information. We would also like to thank Richard Blundell, Ben Elsner, Eric French, Richard Murphy, Cormac O'Dea, Fabien Postal-Vinay, Barra Roantree, Uta Schoenberg, Michela Tincani and seminar participants at NUI Galway, NUI Maynooth, UCD Dublin, European University Institute, University of Bath, Lancaster University, and University of Manchester. This is a heavily revised version of a paper previously circulated as "The Effect of High School Rank in English and Math on College Major Choice". This work was partially supported by the Research Council of Norway through its Centres of Excellence Scheme, FAIR project No 262675.

1. Introduction

The choice of college major is one of the most important decisions made by young people and can have a great impact on later earnings in the labor market (Altonji et al., 2016). It is well established that academic preparation and student interests are predominant determinants of college major choice – students tend to enter fields that they enjoy and in which they are likely to do well. However, recent research has found that, in addition to absolute skills and achievement, relative class rank in school matters for human capital accumulation and for educational behavior. In this paper, we use Irish data to investigate whether, conditional on achievement at the end of high school, rank in English and math affects choice of college major. Given its importance to the economy and its large gender gap, we place particular emphasis on whether higher rank in math and lower rank in English causes high school students to be more likely to choose Science, Technology, Engineering, and Math (STEM) fields in college.

There are two major motivations for our study. First, many papers have studied the role of comparative advantage (in particular, in math and English) in determining college major.[1] However, little work has considered whether students additionally consider their within-school rankings in math and English when making choices. Students may lack information about their academic ability (Zafar, 2011; Stinebrickner and Stinebrickner, 2012, 2014; Bobba and Frisancho, 2014) and this uncertainty may lead students to infer their comparative advantage across subjects from their rank across subjects in school.[2] While relative achievement is informative, it can also lead students astray if the distribution of achievement in their class is not typical. If ordinal rank is important, it could motivate policy interventions to provide

[1] Studies include Speer (2017), Card and Payne (2017), Delaney and Devereux (2019), Breda and Napp (2019), and Aucejo and James (2019). Delaney and Devereux (2020a) have shown that relative achievement in math and English is also important for performance at university.

[2] Tincani (2015) and Bursztyn and Jensen (2015) argue that students care about rank and status and are more willing to invest effort to improve if it will increase their rank within their school. Azmat and Iriberri (2010) and Azmat et al. (2019) find that providing feedback on relative performance in school affects subsequent student performance.

information to school students about their absolute achievement level. Second, women are still greatly underrepresented in STEM college programs with serious implications for gender earnings gaps (Card and Payne, 2017).[3] Previous research has found that this is partly due to female comparative advantage in English compared to math (Speer, 2017; Card and Payne, 2017; Delaney and Devereux, 2019; Breda and Napp, 2019; Aucejo and James, 2019). However, relatively little is known about whether the STEM gender gap can be further explained by the tendency for women to be higher ranked within-school in English and lower ranked within-school in math than men.

In this paper, we study whether rank in math and English affects college major decisions, after controlling for multiple measures of achievement and academic interests at the end of high school. There are several reasons why math and English ranks may affect college major choices. Students may develop confidence in a particular area of study from a higher class rank or despondence from a lower one.[4] The advice students obtain from peers, teachers, or family members about college major choice may depend on school-cohort rank in math and English.[5] Also, students may be uncertain about their ability and so may rely on rank to infer ability and comparative advantage across subjects.

While the recent rank literature has found strong impacts of class rank on many outcomes including earnings, high school graduation, college enrollment, and risky behavior (Murphy and Weinhardt, 2020; Denning et al., 2018; Elsner and Isphording, 2017, 2018), there

[3] There is a large literature studying the gender gap in STEM. Recent papers include Nollenberger et al. (2016), Mouganie and Wang (2020), McDool and Morris (2020), Shi (2018), Astorne-Figari and Speer (2018), and Friedman-Sokuler and Justman. (2016).

[4] Using survey data, Elsner and Isphording (2017) find that higher ranked students believe themselves to be more intelligent and have better mental health than other equally able students. Murphy and Weinhardt (2020) find that subject rank has positive effects on student self-confidence in that subject.

[5] Pop-Eleches and Urquiola (2013) show that behavior of teachers and parents is affected by the student's rank within the school (parents provide less help if the child is in a better school; teachers are found to pay more attention to higher ranked students). However, Elsner and Isphording (2017) find no evidence for this mechanism in their study of US high school students. Kinsler and Pavan (2020) show that parental beliefs and investments in kindergarten in the US are influenced by the child's skill relative to that of other children in the same class. However, Murphy and Weinhardt (2020) find no effect of primary school rank on parental investment.

has been little focus on the relationship between rank in English and math in high school and choice of college major. Using UK data, Murphy and Weinhardt (2020) find that students who are ranked higher in a subject in primary school are more likely to complete that subject at A-Level. However, they do not examine rank at the end of high school or the choice of college major. Denning et al. (2018), using data from Texas, find that math rank in 3rd grade has a positive effect on doing a STEM major in college; however, they don't consider the effects of rank at the end of high school.[6] Elsner et al. (2019) find that rank within teaching sections in a Dutch university affects subsequent choice of courses. However, their focus is on different choices within the field of business rather than across very distinct fields.

We add to the literature in several ways. First, our focus differs from these papers in that, rather than studying the effect of rank at early ages on much later choices, we abstract from human capital effects of rank and examine the purely behavioral effects of rank at the end of high school on college choices, conditional on absolute performance at the end of high school. We have grades for each of the 7 or 8 subjects taken in the Leaving Certificate examinations. These high-stakes exams are centrally set and graded and so are comparable across all students. They provide a detailed description of academic readiness at the end of high school and allow us to control for absolute achievement in a variety of subjects as well as for academic interests as revealed by high school subject choices. Thus, we can isolate the effect of math and English ranks on student choices, conditional on their academic interests and achievements.

Second, our data include preference rankings over college majors for all high school students who apply for college and, if relevant, the program accepted. Thus, we can study desired college program of study for all persons who consider college, not just for the sample

[6] Also, because they give a 0 for STEM to all people who do not go to college, it is hard to disentangle the effect of math rank on college major choice from that of math rank on college enrollment. In Texas, the top 10% of students in each high school are guaranteed college admission and this may influence their estimates to the extent that overall rank at the end of high school correlates with math rank in 3rd grade.

who actually attend. As such, we can see how math and English ranks affect desired college major for all applicants. Third, compared to the U.S., there are several features of the Irish system that make it conceptually easier to study the effects of high school rank on college choices. Unlike in the US, college admission decisions are never influenced by class rank but are predominantly determined by Leaving Certificate points that are solely based on scores in the student's best 6 subjects. Also, both English and math are compulsory subjects throughout high school so we can calculate within-school ranks in these for all students who apply to college. Finally, we add to the literature on the gender gap in STEM. In mixed-sex schools girls tend to be lower ranked in math and higher ranked in English than boys. We examine whether these differential ranks in English and math by gender have explanatory power for the gender gap in the choice of STEM as a college major and for the larger gender gap in STEM in mixed-sex schools compared to same-sex schools.

We find that, conditional on achievement at the end of high school, within school-cohort percentile rank in English and math is predictive for field choice, particularly for STEM and Arts and Social Sciences -- higher English rank is positively associated with choosing Arts and Social Sciences and negatively with STEM; higher math rank is positively associated with STEM and negatively with Arts and Social Sciences. This finding implies behavioral effects of subject rank that go beyond their effects on human capital accumulation in school. Subject ranks have some explanatory power for the gender gap in the choice of STEM as a college major in mixed-sex schools – the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 4% of the STEM gender gap in mixed-sex schools and about 10% of the difference in the STEM gender gap between mixed-sex schools and same-sex schools.

The structure of the paper is as follows: In the next section, we describe the institutional background and data, and, in Section 3, we describe the empirical methodology. In Section 4, we present our main results. Section 5 outlines a set of robustness checks. Section 6 shows that our estimates can help explain the gender gap in STEM. Finally, Section 7 concludes.

2. Institutional Background and Data

We use data from the Central Admissions Office (CAO) that include all individuals who did their Leaving Certificate (the terminal high school exam in Ireland) and applied to an Irish college in the years 2015 to 2017.[7] The CAO is an independent company that processes applications for undergraduate courses in Irish colleges, issues offers to applicants, and records all acceptances. The CAO centralized system means that applicants do not have to apply separately to different colleges and that data are processed and collected in one place. When applying for a college course, applicants can list up to 10 level 8 courses (honors bachelor's degrees) and 10 level 6/7 courses (ordinary bachelor's degrees and higher certificates). For the majority of courses, whether or not an applicant is accepted depends solely on their performance in the Leaving Certificate.[8] At the end of the last year of high school, students sit the Leaving Certificate, typically in 7 or 8 subjects, and grades in the student's 6 best subjects are combined to form their total Leaving Certificate points.[9] Each college program has a minimum points level that is required to enter. The required points vary from year to year depending on the preferences of students and the number of available places in the program. If the student has points equal to or above the minimum for their first-ranked program, they are offered that program. If not, they are offered the highest ranked program for which they have enough points.

[7] This section draws heavily from Delaney and Devereux (2019).
[8] There are a small number of college programs that do admissions based on information other than Leaving Certificate points. For example, music programs typically require an audition, and arts/architecture programs may require a portfolio.
[9] The Leaving Certificate examinations are written exams that are centrally set and are anonymously graded by external examiners.

English, Irish and math are compulsory high school subjects and the student chooses another 4 or 5 subjects to study based on their interests, aptitudes, and future college and career plans.[10] Some college programs have subject requirements, for example, pharmacy and veterinary medicine require Leaving Certificate chemistry, and most engineering and science programs require at least one Leaving Certificate science subject.

All subjects are offered at a higher or lower level.[11] The grades awarded and mapping from grades to points changed in 2017. Appendix Table A1 shows how points/grades are awarded during our three-year period. Since 2012, to induce more students to study higher level math, an additional 25 points bonus is given in math to those who pass the subject at higher level.[12] Better students do subjects at higher level and it is unlikely that good students would choose to study math or English at lower level for strategic reasons. For math, unless a student fails the exam (scores less than 40%), then the points awarded at higher level strictly dominate the points awarded at lower level so it would make no sense to choose to study lower level math unless the student expected to fail higher level math. For English and all other subjects, the points awarded at higher level are greater than the highest possible points at lower level provided the student gets more than 50% at higher level (55% in 2017). Just 2.7% of those taking lower level English achieve the highest grade. In contrast, of those taking higher level, 82% score a grade that gives higher points than the maximum points from lower level. Overall,

---

[10] While Irish is compulsory, there are exemptions available for children who have lived for a sufficient time outside of Ireland or who have a learning disability (https://www.education.ie/en/Circulars-and-Forms/Active-Circulars/ppc10_94.pdf). Therefore, in practice, many students do not study Irish for Leaving Certificate.

[11] A student who takes math at higher level and English at lower level will be in the same math class as others who also take math at higher level and the same English class as those who take English at lower level. Therefore, a student can have different peers in their math and English classes. However, students will also interact with other students as they take classes in other subjects.

[12] Other than the bonus points for mathematics, grades in all Leaving Certificate subjects count equally for points, irrespective of the college programs to which the student applies. For example, an H1 grade in history provides the same points as an H1 grade in physics whether the student applies to a humanities program or a science program. Note, also, that students are not required to use their grades in English and mathematics for their points, although in practice most do.

it is reasonable to assume that, in any subject, the stronger students take higher level and the weaker students choose lower level.

It is also unlikely that students strategically choose to make low effort in English or math. While students need not include English and math amongst the best 6 subjects used to calculate points, over 99% of students have either math or English included in their best 6 subjects. Overall, 8% of students do not use English when calculating their final Leaving Certificate points while 37% do not include math. Of the 37% who do not include math scores, almost all of them took lower level math for Leaving Certificate.[13] The most likely explanation for the relatively poor performance of many people in math is that they are just not very good at it and tend to do better in other, less quantitative, subjects. Overall, while we cannot rule it out, it is unlikely that many applicants made little effort to do well in English or math.

The CAO data include information on the applicant's age, gender, high school, Leaving Certificate subjects and grades, county of origin, year they sat the Leaving Certificate, and whether they have a foreign qualification. Our baseline sample includes 137,708 individuals who apply to the CAO in the same year as they sit the Leaving Certificate. We restrict the sample to applicants between the ages of 16 and 20 which reduces the sample size by 1,542 observations. We also drop those who took the Leaving Certificate exams more than once, reducing the sample by a further 3,372 observations. In addition, we drop 518 applicants who took fewer than six subjects in their Leaving Certificate.[14] We omit a few schools that are "grinds" schools – private schools that are aimed at students who wish to do just the last year (or two years) of high school at an exam-oriented school – as we do not have the requisite information to calculate ranks in these schools. This reduces the sample by a further 3,273

[13] While it is possible that some of these students put in little effort, it should be noted that all colleges require students to pass math so it would be risky for students to put little effort into studying math in case they do not pass.

[14] We also delete cases with missing information on high school attended (161 observations) or where the number of students taking the Leaving Certificate exams is not available for the school (117 observations), and a further 76 cases where the grade in English or math is missing.

observations. Finally, we drop 1,723 observations that are missing information on preferences over college programs. This results in a sample with 126,962 observations.

We allocate all programs to one of four fields (STEM; Arts and Social Sciences; Business, Administration and Law; and other) using the International Standard Classification of Education (ISCED).[15] While applicants can list up to 10 level 6/7 and 10 level 8 programs, in practice, the most important decisions are what programs to place at or near the top of the lists. In our main analysis, we focus on the college program listed as first choice by the student. If the student listed both level 6/7 and level 8 programs (and so had a preference ordering for two distinct lists), we use the first-choice level 8 program, otherwise we use the first-choice program on the list used by the student (97% of students list at least one level 8 program).[16]

We rank students in math and English based on their Leaving Certificate grades in these subjects. They sit the Leaving Certificate exams in June and final college application choices must be made by July. Thus, choices are made after sitting the exams but before receiving the results. While students do not know their results, we assume that students have an estimate of where they rank in the Leaving Certificate achievement distribution for the school-cohort. Exams are given throughout the year in each year of high school and “mock Leaving Certificate exams” are provided to the student.[17] Thus, students have regular feedback on their performance in English and math and tend to be aware also about how their classmates are

[15] In general, we denote a program as STEM if it is in Natural Sciences, Math, and Statistics (ISCED-05), Information and Communication Technologies (ISCED-06), or Engineering, Manufacturing, and Construction (ISCED-07); however, following Delaney and Devereux (2019), we adjust the categories slightly as we think some programs are more likely to fall under STEM than others. Therefore, we include Dentistry (0911), Medicine (0912), Pharmacy (0916), and Veterinary (0841) as STEM and remove Wildlife (0522), Food Processing (0721), and Materials (0722).

[16] Typically students list level 6/7 programs as their safe or back up options as these programs generally have much lower points requirements. Given that level 8 programs are honours-degree programs and are much more selective than level 6/7 non-honours-degree programs, using the highest-ranked level 8 program best captures the student’s “dream” program abstracting from the probability of admission.

[17] The “mock” exams are taken about 4 months prior to the Leaving Certificate and are a complete rehearsal for the Leaving Certificate. Students sit the full set of exams under the same conditions that they later face in the Leaving Certificate. These exams are strongly predictive of actual Leaving Certificate performance and there is usually much discussion among class peers about performance in the “mock” exams.

performing. They also have feedback from sitting the Leaving Certificate exams. Note that it is standard in the literature to assume that students have some perception of their rank even though it is not formally reported to them.[18]

Consistent with the literature, we use a percentile measure of rank that is calculated as follows:

$$Rank = \frac{(n_i - 1)}{(N_i - 1)}$$

where n_i is the student's ordinal rank in the subject in the school-cohort and N_i is the number of students in the school-cohort.[19] We percentilize the ordinal rank with the above transformation because a simple ordinal rank measure would not be comparable across schools of different sizes. Our percentile rank measures are approximately uniformly distributed, and are bounded between 0 and 1, where 0 denotes the lowest ranked student in a subject in a school-cohort and 1 denotes the highest ranked student in a subject in a school-cohort. We do separate rankings for math and for English based on grades achieved in these subjects.

While, as mentioned above, we exclude some observations from our estimating sample, such as omitting students aged over 20, we include all students (except repeat students) when calculating ranks.[20] This is important as otherwise we could erroneously assign a student as top ranked if the actual highest ranked student was dropped from the sample due to, for example, an age restriction. We know the total number of students who sit the Leaving Certificate exams

[18] Given that individuals do not know their exact rank in Leaving Certificate achievement, in practice we will be estimating the reduced form effects of perceived rank using actual rank. There is mixed evidence on how perceived rank relates to actual rank. Azmat et al. (2019) find that the majority of college students underestimate their rank in the grade distribution while Tincani et al. (2020) find that high school students tend to overestimate their school rank. To the extent that the actual rank in achievement differs from perceived rank, and that it is perceived rank that matters, this would lead our estimates to be attenuated.

[19] In the event of ties, we follow Denning et al. (2018) and assign individuals the average rank. For example, if three people are joint top in a school-cohort, we give each of them an ordinal rank of 2 and the next in line then has an ordinal rank of 4. Later, we show that our results are robust to instead giving all students who tie, the highest ranking or the lowest ranking, amongst the group who are tied.

[20] We drop Leaving Certificate repeaters from the calculation of rank as these students often go to a different school to repeat and it is unlikely that non-repeating students compare themselves to repeaters. However, we find that the estimates are very similar if we include repeat students.

in each school in each year (and the number of these who are repeat students) from data provided by the State Examinations Commission (SEC). Thus, we know the number of non-repeat students in each school in each year.

The major issue we face in calculating ranks is that we do not know Leaving Certificate grades for students who do not apply to college – 83% of Leaving Certificate students apply to the CAO. In our main analysis, we assume that those who have not applied to the CAO and, so, are not in our sample, come from the bottom of the distribution and would have ranked lower in English and math than those who apply. This is not as strong an assumption as it appears as even persons who plan to go to college abroad generally also apply to the CAO.[21] So, non-applicants are generally the least academically inclined students.[22] To reduce the measurement error problem, we remove observations in which less than 75% of the school-cohort applied to the CAO; this reduces our sample by 18% and reduces the number of school-cohorts from 2,029 to 1,409. In the remaining schools, over 88% of students apply to the CAO. Later in the paper, we provide evidence that remaining measurement error in rank due to non-applicants is not likely to be large.[23] The reduction in measurement error in rank means that our results are more likely to be internally valid. We also verify that leaving out schools with a low percentage of applicants is unlikely to affect the external validity of our estimates.

As seen in Appendix Table A1, the grading scheme changed somewhat in 2017. To use all available information, we form the ranks in each year using the grades in that year. Both math and English are compulsory subjects for Leaving Certificate so there is no selection problem due to different students taking different subjects. However, a complication is that

[21] In addition, students who plan to defer college (take a gap year) are encouraged to apply anyhow in case they change their mind.
[22] We have compared the distribution of grades in math and English in the CAO data with information on the distribution of grades for all Leaving Certificate students that is publicly available on the SEC website. Consistent with our assumption, we find that students from the top end of the grade distribution are relatively more likely to apply to college.
[23] Note that, because we are using administrative data, our grade measures are very accurate and unlikely to contain error, so we believe any measurement error in rank will arise because of non-applicants.

students can take these subjects at either a higher or lower level, each level has a different exam paper and a different mapping from grades to points. As discussed earlier, we believe that it is appropriate to rank students who study at higher level above those who study at lower level. Within schools, students who do higher level math will likely be perceived as being better at math than students in lower level, and so we rank even those who did badly in higher level math higher than those who did well in lower level math. Generally, there are separate classes for higher and lower level students and, so, it is reasonable that students who do lower level assume that they are worse than those who do higher level.[24] At each level, we rank those who obtain an A1 higher than those who obtained an A2, and rank those who obtain an A2 higher than those who obtained a B1, etc. In the robustness checks, we show estimates using alternative methods of dealing with the higher and lower level grades.

Descriptive statistics for our sample are in Table 1. Because we assume that non-applicants have lower rank than applicants, the average percentile rank in our sample is 0.56 for both English and math. Two-thirds of applicants list a university program as top choice but only 42% end up enrolling in a university.[25] Points range from zero to 625 but only 5 students (.005% of the sample) score zero points in their Leaving Certificate.

[24] Under the assumption that all higher level students rank above lower level students, it is easier for students to know their rank as they only need to know their rank amongst other students at the same level. So, misperception of rank is less likely to be a problem if this assumption is reasonable.

[25] During this period, there were seven universities: University College Dublin (UCD), Trinity College Dublin (TCD), Dublin City University (DCU), Maynooth University (MU), National University of Ireland, Galway (NUIG), University College Cork (UCC), and University of Limerick (UL). The remaining colleges are mostly institutes of technology and teacher training colleges.

Table 1: Descriptive Statistics

	Mean	SD	Min	Max	Observations
Age	17.40	0.63	16	20	104116
Year	2015.99	0.81	2015	2017	104116
Female	0.51	0.50	0	1	104116
Leaving Certificate Points	385.69	115.76	0	625	104116
Math Points	49.52	35.02	0	125	104116
English Points	61.20	20.14	0	100	104116
Math Rank	0.56	0.26	0	1	104116
English Rank	0.56	0.26	0	1	104116
Overall Rank based on Total Points	0.56	0.26	0	1	104116
List Both Level 7 and 8 Programs	0.66	0.47	0	1	104116
List Level 7 Programs Only	0.03	0.18	0	1	104116
List Level 8 Programs Only	0.31	0.46	0	1	104116
First Choice is a University	0.65	0.48	0	1	104116
First Choice is STEM	0.30	0.46	0	1	104116
First Choice is Business and Law	0.21	0.41	0	1	104116
First Choice is Arts and Social Sciences	0.20	0.40	0	1	104116
First Choice is Other Field	0.30	0.46	0	1	104116
Enroll in Top Choice	0.33	0.47	0	1	104116
Enroll in Any Program	0.73	0.44	0	1	104116
Enroll in University Program	0.42	0.49	0	1	104116
Enroll in Top Choice (given enroll)	0.45	0.50	0	1	75939
Enroll in STEM (given enroll)	0.31	0.46	0	1	75939
Enroll in Business and Law (given enroll)	0.23	0.42	0	1	75939
Enroll in Arts and Social Sciences (given enroll)	0.23	0.42	0	1	75939
Enroll in Other Field (given enroll)	0.22	0.42	0	1	75939

Sample: Central Admissions Office (CAO) 2015 – 2017

3. Methodology

Given absolute achievement is highly correlated with school-cohort rank, the key to isolating the effect of rank is to control for the absolute level of achievement. We do this flexibly by controlling for indicator variables for obtaining each possible grade in English and math, both of which are compulsory subjects for Leaving Certificate. We further control for academic interests and subject-specific abilities by adding indicator variables for whether the student took each of the 25 most popular subjects for Leaving Certificate and further indicators for grades achieved in each of the 25 subjects (interacted with indicators for taking the subjects).[26] Also, because entry to most college programs is determined by points achieved in the Leaving Certificate, we include a quartic function of points. Taken together, these variables account for the absolute factors that should influence choice of college major.

We also include a full set of school-cohort indicators. The inclusion of the school-cohort indicators is important as, otherwise, our rank estimates could be biased by correlations with school-specific factors such as the quality of teachers, facilities, and peers. Conditional on grades, students who are highly ranked will tend to be in low-achieving schools, and school quality is an omitted variable that could cause bias. Therefore, it is important to include school-cohort fixed effects as these eliminate all the potential confounders mentioned above by absorbing all mean differences between school-cohorts (see Murphy and Weinhardt (2020) for further discussion on this point).

Identification of rank effects when including school-cohort indicators

Given we include grade indicators and school-cohort indicators, rank effects are identified due to the exclusion of interactions between grade indicators and school-cohort indicators. For simplicity, consider identifying the effect of a single subject rank (the effect of rank in math). We abstract from individual-level variation and consider variation by school-cohort (c) and by math grade (g) as math rank for any individual depends only on their school-cohort and their math grade. Denoting the percentile rank in math as RM, we write the relationship between the outcome and math rank as

$$Y_{cg} = \alpha + \beta RM_{cg} + v_{cg}. \qquad (1)$$

Then, the critical identifying assumption is that

[26] While English, mathematics, and Irish are compulsory subjects for Leaving Certificate, students choose an additional 4 or 5 subjects. These choices (for example, choosing physics rather than history) provide a lot of information about academic interests. Additionally, grades in these subjects (in addition to grades in English and mathematics) provide much information about student ability across a range of subjects. Taken together, we believe that the subject choices and grades provide a rich set of controls for achievement and academic interests.

$$E(v_{cg}|c,g) = \gamma_g + \theta_c. \tag{2}$$

This assumption states that differences in the outcome variable across combinations of math grades and school-cohorts can be summarized by an additive school-cohort effect and an additive grade effect. If the outcome is choosing STEM, it allows STEM probabilities to differ systematically across school-cohorts and to differ systematically by math grades. However, it posits that, other than math rank, functions of interactions between school-cohorts and math grades do not belong in the model. This allows the identification of math rank from cases where differences in math rank across grades are not homogenous across schools. Given the assumption in (2), the presence of indicators for subject grades and indicators for school-cohort provide consistent estimation of subject rank effects.

For example, consider two schools that have the same distribution of English grades. Given that the distribution of English grades is the same in both schools, there is no variation in English rank conditional on grade indicators and so we cannot identify the effect of English rank. Suppose, however, that the math grade distribution differs between the two schools and that going from an A grade to a C grade in math in one school leads to math rank falling by 0.5; while going from an A to a C in math in the other school leads to math rank falling by 0.25. We have identifying variation in math rank as the differences in math rank between the two schools is not the same for each math grade. That is, so long as math rank cannot be written as the sum of a school-cohort effect and a math grade effect, the effect of math rank is identified.

Estimating Equation

We use the following linear specification (later, we also show estimates for a non-linear specification):

$$Y_{ic} = \alpha + \beta_1 RM_{ic} + \beta_2 RE_{ic} + X\delta + \theta_c + \varepsilon_{ic}, \tag{3}$$

where Y_{ic} represents the college field choice of individual i in school-cohort c, RM is the percentile rank in math, RE is the percentile rank in English, X includes a vector of controls including age and gender and the controls for grades in math and English, Leaving Certificate subject choices and subject-specific grades, and Leaving Certificate points described above, and θ_c represents school-cohort fixed effects. We also include a control for overall rank as measured by the within school-cohort percentile rank based on total Leaving Certificate points. This allows us to isolate the effect of math and English rank abstracting from any effect of overall rank. Omitting the control for overall rank might lead us to ascribe the effects of overall rank to math rank or English rank.[27] As we show later, whether or not we include this variable has very little effect on our estimates. We cluster the standard errors at the school level and, so, allow for both serial and school level correlation in the errors.

Appendix Figure A1 shows that there is a distribution of ranks at each grade level for English and math. Each box plot displays the distribution of the subject rank for a particular subject grade. The variation in subject rank is strongest in the middle of the grade distribution as students with mediocre grades are widely dispersed in terms of rank due to variation in the grades of their peers. There is less variation in rank at the highest level of achievement. Appendix Table A2 shows the variation in the residual after regressing rank on school-cohort indicators, gender, age indicators, and our achievement controls. We find that the standard deviation in rank is approximately 0.05 for each of our rank measures, which is non-trivial given that rank is bounded between 0 and 1. Math rank and English rank are positively correlated, the correlation coefficient is 0.55.

We consider subject ranks to be predetermined at the point that students make final decisions about college major in July. At that point, while students have not received their

[27] The correlation between math rank and overall rank is 0.82 and that between English rank and overall rank is 0.75.

Leaving Certificate results, they have completed the exams and have also received a lot of feedback about their likely performance from in-school tests and, as described earlier, from the "mock Leaving Certificate examinations" done earlier in the year. While students are not made explicitly aware of their class rank, we assume, in common with the literature, that they are aware of their rank from repeated interactions with other students. Indeed, we believe that achievement rank will generally be more salient than ability rank as it is repeatedly revealed through test performance.

Grades may be a function of many factors including ability at the start of high school, quality of teaching, study habits of the student, effects of peers on human capital accumulation, and effects of desired field of study on effort in math and English. Therefore, by controlling for grades, our rank estimates measure the effect of rank conditional on achievement at the end of high school.

A potential problem is that students who desire to do STEM may work harder at math and students who desire to do ASSc may work harder at English. This type of motivation would constitute a potential omitted variable bias as desire for STEM is correlated with the outcome variable (choosing STEM) and is also correlated with grades in math and English and, hence, with rank in these subjects. This will not lead to bias so long as these unobserved preferences are captured by our controls for achievement so that there is no correlation between desire for STEM and rank in math, conditional on achievement controls. Note that we also control for Leaving Certificate subject choices and subject grades. These should absorb much of the effect of desired field of study as those who are motivated to do STEM may be likely to choose STEM-friendly subjects in high school such as physics or chemistry and, conditional on subject choices, may get better grades in these subjects.

A second potential issue is that of peer effects. Subject rank is a type of peer effect, but peers may also influence choice of college major in other ways. We approach this potential issue in several ways: firstly, we control for school-cohort fixed effects which account for any peer effects that have the same effect on all students; secondly, in our robustness checks, we capture potential non-linear peer effects by controlling for interactions of math and English grades with the mean and standard deviation of subject achievement in the school-cohort.[28] The results are robust to accounting for these non-linear peer effects.

Choice of Dependent Variables

Our main specifications focus on the field of the top ranked level 8 program listed (or level 6/7 program if the student lists no level 8s) because that is likely to reflect the program that is most preferred by the student. The college admissions allocation mechanism used in Ireland is a "serial dictatorship" allocation mechanism – the algorithm allocates the applicant with the highest points his/her first preference, then the second-ranked applicant gets an offer for his/her top ranked program amongst those still available, and so on. Candidates are accepted to the highest ranked program for which they have sufficient points for admission. If the student has points equal to or above the minimum for their first-ranked program, they are offered that program. If not, they are offered the highest ranked program for which they have enough points.[29] As shown by Svensson (1999), this type of allocation mechanism is strategy-proof and induces applicants to provide a ranking that reflects their preferences if additional choices are costless and there are no limits on how many programs students can rank. The logic is that there is no cost to listing the most preferred program first as, if the applicant does not receive

[28] These controls allow for peer effects to be heterogeneous by individual achievement. See Booij et al. (2017) and Bertoni and Nistico (2019).
[29] A student can be offered both a level 6/7 and a level 8 program if they list programs on both lists.

enough points to be admitted to that program, they are then considered for their second-choice program, and so on.

In our setting, students can list 20 choices on the CAO form (10 choices for each of level 6/7 and 8 programs) and empirically we find students do not exhaust this full list of choices – just 26% list all 20 choices. In addition, marginal applications are costless – it costs the same to list 20 programs as it does to list a single program. Because students can only rank a finite number of programs, it may be optimal for them to include programs towards the end of the preference list that are less preferred than some omitted programs but for which there is a very high probability of admittance for the student. However, there is no reason not to list their most-preferred program as first choice unless they believe that they have zero probability of obtaining this program.[30] Therefore, by focusing on the top ranked choice, we are most likely studying the program that is most preferred by the student. However, we also show robustness checks where we study second and third choices as well as the proportion of all program choices that are in a particular field.

4. Results

It is well established that English and math grades are predictive of choice of college field with an emphasis in the literature on how they affect whether students choose to do STEM (Speer, 2017; Card and Payne, 2017; Delaney and Devereux, 2019; Aucejo and James, 2019). In this section, we advance this literature by studying whether, conditional on a broad array of student achievement measures, within school-cohort ranks in English and math are associated

[30] Also, the CAO illustrate on their website that, when filling out the forms, students should list their "dream" programs in their top choices, "more realistic" programs as their middle choices, and "banker" programs for which they are more certain of acceptance further down the list.

with field choice. Our expectation is that persons with a higher rank in math and/or a lower rank in English may be more likely to choose a STEM program. Likewise, a higher rank in math and/or a lower rank in English may be associated with a lower likelihood of choosing Arts and Social Sciences (ASSc).

Table 2 reports our estimates for the effect of rank in these subjects on field choice. A one decile increase in math rank leads to a 1 percentage point increase in the probability of listing STEM as first preference and a 0.7 percentage point decline in the probability of listing ASSc. These compare to baseline first preference probabilities of 0.30 and 0.20, respectively. On the other hand, a one decile increase in English rank decreases the probability of listing STEM by 0.4 percentage points and increases the probability of listing ASSc by 0.7 percentage points. We find small effects of math and English ranks on listing a Business Administration and Law (BAL) major and on listing a major from some other field, and none of the rank coefficients for these fields are statistically significant at the 5% level. We conclude that math and English ranks affect college major choice mainly through their effects on choosing STEM and Arts and Social Sciences. Given that we are controlling for absolute achievement at the end of high school, we consider the magnitudes of the school-cohort rank effects to be important and they suggest meaningful behavioral responses to within school subject rank.[31]

[31] Elsner et al. (2019) find that, in a Dutch business school, a one decile increase in rank in a teaching section in a compulsory subject increases the probability of subsequently choosing to major in that subject by 1 percentage point. Our estimates are similar in magnitude to theirs.

Table 2: Effect of Rank in Math and English on First Preference Field of Study

VARIABLES	(1) STEM	(2) BAL	(3) Arts & Social	(4) Other
Math Rank	0.103***	0.025	-0.066**	-0.062*
	(0.029)	(0.027)	(0.029)	(0.034)
English Rank	-0.044*	0.006	0.069***	-0.031
	(0.025)	(0.022)	(0.026)	(0.028)
Observations	104,116	104,116	104,116	104,116
R-squared	0.312	0.197	0.170	0.178
Mean Outcome	0.299	0.208	0.197	0.296

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

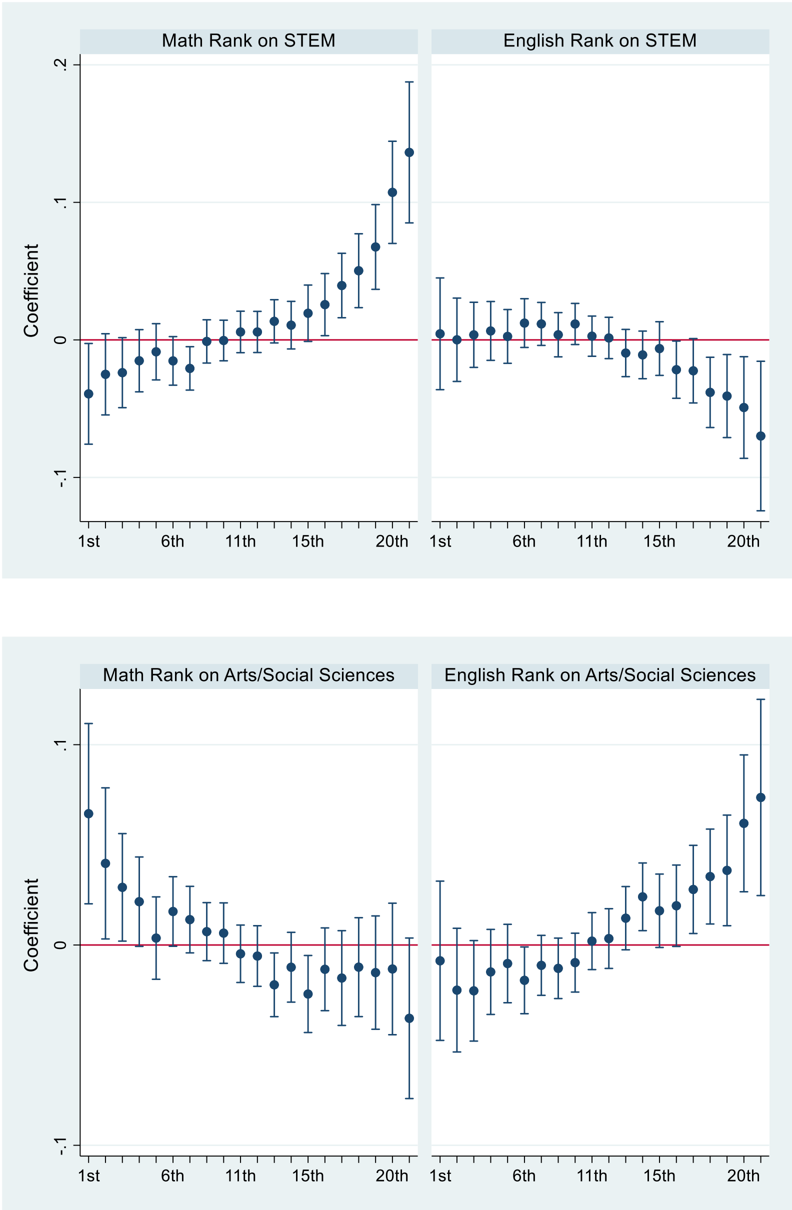
4.1 Heterogeneous Effects

Given our estimates in Table 2 show that math rank and English rank are particularly significant for STEM and Arts and Social Sciences (ASSc), in the rest of the paper, for parsimony, we focus our analysis on these two fields. We examine heterogeneous effects across the subject rank distribution, by gender, and by size of school.

Non-linearities

We replace the linear subject rank variables with indicator variables for being in each ventile of the rank distributions plus indicators for being the top person(s) in the subject in the school-cohort, with the 10^{th} ventile being the omitted category. We plot the estimates and 95% confidence intervals in Figure 1. The effect of subject rank is approximately linear for Arts and Social Sciences and is also close to linear for the effect of math rank on STEM. In contrast, we only see a negative effect of English rank on STEM in the top half of the English rank distribution; the relationship is quite flat in the bottom half of the distribution.

Figure 1: Effect of Rank Ventiles and Top Ranked Person in Math and English on First Preference Field of Study



Estimates from regressions where subject rank is entered in ventiles, with an additional category for the top ranked person(s). The omitted category is the 10th ventile. Point estimates and 95% confidence intervals are shown.

Effects by Gender

There are several reasons why the effects of rank may differ by gender.[32] To estimate differential effects by gender, we interact English and math rank with indicators for male. We also include interactions of gender with grade indicators for English and math to take account of correlations between subject rank and absolute achievement in the subject. We plot the estimates and 95% confidence intervals in Figure 2. We find that the effect of math rank on STEM is larger for boys than for girls (0.11 versus 0.06) but this difference is not statistically significant. Likewise, there is a negative effect of English rank on STEM of -0.08 for boys while the effect is effectively zero for girls (the gender difference is significant at the 5% level). This is consistent with previous literature that found larger effects of rank for males than females (Murphy and Weinhardt, 2020).[33] Interestingly, this effect only appears for the effect of subject rank on choosing STEM. There do not appear to be any large gender differences in the effect of math or English rank on choosing Arts and Social Sciences.

These findings may relate to differential gender norms in which females are less expected to choose STEM and are less confident about their ability to succeed in STEM at college. Another possible explanation is that girls are more mature and pay less attention to relative rank in school (as they should) when making choices.[34] The gender differences for rank relate somewhat to previous findings about the responsiveness of STEM choice to absolute performance in English and math. Delaney and Devereux (2019) find that boys are more likely to make decisions on STEM based on their comparative advantage in English and math

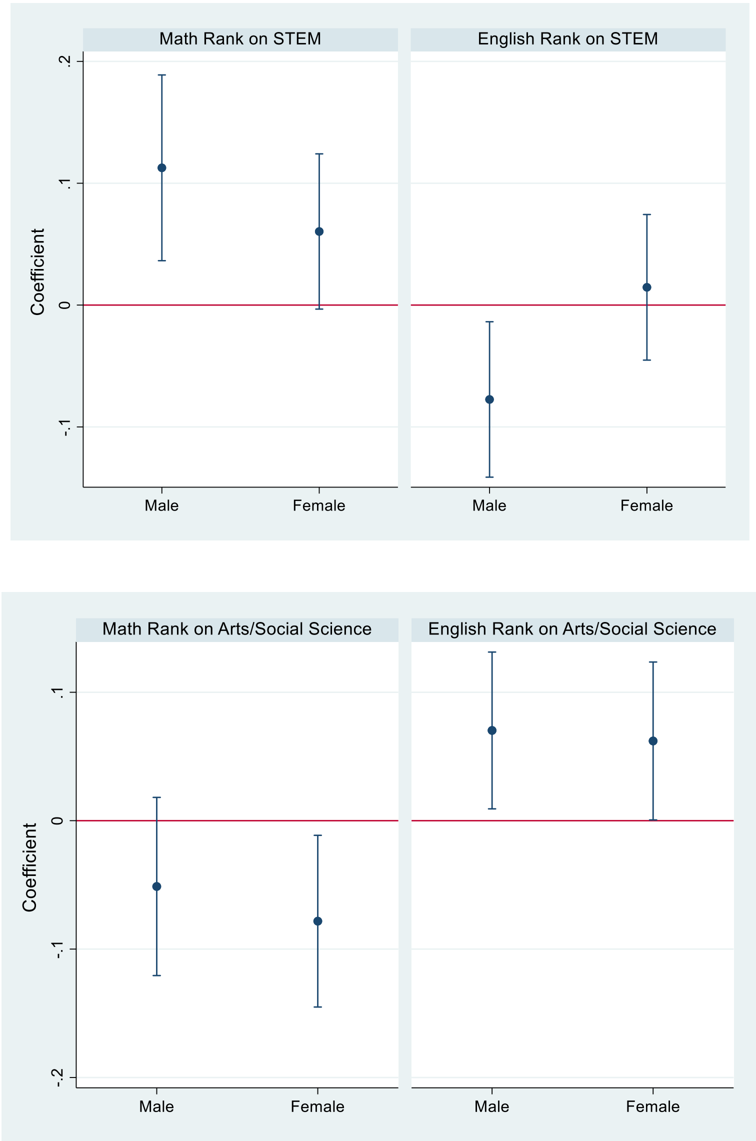
[32] A large literature has found that behavior of boys and girls differs along many dimensions with several papers finding that girls are less competitive than boys (Buser et al. 2017), are more risk averse (Reuben et al. 2015), are more sensitive to grades (Rask and Tiefenthaler, 2017), and are less confident in math (Bordalo et al. 2019).

[33] Elsner et al. (2019) find that the effect of rank on study effort is larger for males in a college tutorial setting. However, Denning et al. (2018) find no evidence of gender heterogeneity in their study of Texas schools.

[34] Delaney and Devereux (2020b) find that girls appear to make more sensible decisions than boys when applying to college in Ireland.

whereas girls are more likely to focus purely on their absolute advantage in math and do not respond much to their English grades.

Figure 2: Effect of Rank in Math and English on First Preference Field of Study by Gender



Estimates from regressions where subject rank is interacted with gender. Point estimates and 95% confidence intervals are shown.

Heterogeneity by Size of Schools

Subject rank may be more salient in smaller schools. We restrict the sample to school-cohorts with at most 60 students to examine this as these school-cohorts typically have at most two classes.[35] Consistent with our prior, Table 3 shows that the subject rank estimates for STEM are much larger in absolute value for the small schools while there is little difference in the effect of subject ranks on ASSc by school size; however, while some of the differences are substantial, none of them are statistically significant.[36] Our finding of larger coefficient sizes in small schools is consistent with rank being more observable in these schools and suggests that our estimates may be attenuated in the full sample due to misperceptions of rank by students.

Table 3: Effect of Rank in Math and English on First Preference Field of Study by School Size

	Size<=60		**Size>60**	
	(1)	(2)	(3)	(4)
VARIABLES	STEM	Arts & Soc	STEM	Arts & Soc
Math Rank	0.184***	-0.098	0.083**	-0.057*
	(0.063)	(0.064)	(0.034)	(0.033)
English Rank	-0.097*	0.045	-0.026	0.066**
	(0.053)	(0.050)	(0.028)	(0.030)
Observations	14,739	14,739	89,377	89,377
R-squared	0.330	0.217	0.314	0.167

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

Interaction of Subject Ranks

It may be the case that choice of college major depends on the interaction between math rank and English rank. We test for this by interacting math and English rank in the regression (and also controlling for interactions between math and English grades). The results displayed

[35] Students in small schools may still have different classmates for math and English depending on whether they study higher or lower level.

[36] We find a similar pattern if we look at school cohorts with at most 50 students or 70 students. One might expect rank effects to be particularly salient in school-cohorts with fewer than 30 students, so only one class. There are too few of these schools in our sample to test this possibility.

in Table A3 in the appendix show that the interaction term is small and insignificant for STEM but there is a positive and statistically significant interaction effect for Arts and Social Sciences. Therefore, there exist important cross subject rank effects for Arts and Social Sciences – the effect of English rank on choosing ASSc is greater for people with higher math rank, and the effect of math rank on ASSc is less negative for students with higher English rank.

5. Robustness checks

We do a series of robustness checks. For brevity, we focus on our main outcomes of interest – whether the student lists STEM or Arts and Social Sciences as their first preference.

Moving Beyond First Preference

As discussed in Section 3, our main specifications focus on the field of the top ranked level 8 program listed (or level 6/7 program if the student lists no level 8s) because that is likely to reflect the program that is most preferred by the student. However, we now show additional specifications where the dependent variable is the field of the student's second choice, field of the third choice, the proportion of the top 3 choices that are in the field, as well as the proportion of all choices (including both levels if both level 6/7 and level 8 programs are listed) that are in a specific field.[37] The estimates are in Appendix Table A4.

When we look at the field of the second and third choice programs, the magnitude of the coefficients tend to be smaller than for first choices, but the overall pattern remains the same. Looking at the proportion of the top 3 choices that are in the field or the proportion of all choices listed that are STEM or Arts/Social Sciences gives very similar results to those

[37] As before, we use the level 8 choices if the student lists both level 6/7s and level 8 programs. 45% of enrolees enter their top choice program, 15% enter their second choice, and 9% enter their third choice. So, in total, about 70% of enrolees enter one of their top 3 choices. Therefore, the top 3 choices provide a good description of what matters. However, we also report estimates for the proportion of all choices that are in the field for completeness.

using the first-choice level 8 program. We conclude that our findings are not simply an artifact of our focus on the top ranked choice.

Calculating Rank when there are non-applicants

While we restrict our sample to school-cohorts where at least 75% of students apply to the CAO, there remains a concern about our assumption that non-applicants are lower-ranked than applicants. As a test of our assumption, we have experimented by assuming that a proportion of non-applicants are missing at random rather than coming from the bottom of the distribution. We first assume that all non-applicants are missing randomly. This is an extreme assumption that we do not think is realistic; however, it informs about how important the treatment of non-applicants could be for our estimates. The estimates displayed in column (1) of Table 4 are very similar to those assuming that non-applicants are lower-ranked than applicants. In columns (2) – (4) of Table 4, we allow various combinations of the proportion of non-applicants assumed to come from the bottom of the grade distribution and the proportion assumed to be missing randomly. In each case, we find very similar estimates. We conclude that our assumption about the ranks of non-applicants is not crucial for our estimates.

How we deal with ties

In our main analysis, we assign ties the average rank so, for example, if 3 people have the highest score in a school-cohort, we assign an ordinal rank of 2 to each of them. In column (5) of Table 4, we show estimates where, instead, we assign the highest rank to ties, for example, if 3 people have the highest score, they would all be assigned an ordinal rank of 1 rather than an ordinal rank of 2. Column (6), on the other hand, shows the effect of assigning the lowest rank to ties. In each case, we find that the estimates are quite robust to the way we deal with ties.

Using Points in English and Math to assign ranks

Admission to college depends on Leaving Certificate points obtained. An alternative to using subject grades to assign ranks in English and math would be to use the points assigned to each grade for that subject (see Appendix Table A1 for the mapping from subject grades to points). In column (7) of Table 4, we show that using points to calculate rank tends to reduce the estimates slightly. This is unsurprising as we believe that our original assumption that students consider persons who do higher level to be better than those doing lower level provides a better measure of subject rank.[38]

[38] We have also evaluated the extent to which our rank effects may be picking up the effect of doing the subject at different levels by adding rank in levels in math and English as controls to our regression. We find that the resulting estimates are similar to our main estimates in Table 2.

Table 4: Robustness Checks – Measurement of Rank

VARIABLES	(1) Assume non-applicants random	(2) Assume 50% of non-applicants random	(3) Assume 30% of non-applicants random	(4) Assume 70% of non-applicants random	(5) Ties (Highest Rank)	(6) Ties (Lowest Rank)	(7) Using Points for Rank
	First Preference Field of Study is STEM						
Math Rank	0.106***	0.107***	0.108***	0.107***	0.083***	0.099***	0.085***
	(0.027)	(0.028)	(0.028)	(0.029)	(0.028)	(0.027)	(0.029)
English Rank	-0.051**	-0.049**	-0.049**	-0.047*	-0.025	-0.053**	-0.038
	(0.024)	(0.025)	(0.024)	(0.025)	(0.022)	(0.024)	(0.026)
R-squared	0.312	0.312	0.312	0.312	0.312	0.312	0.312
Observations	104,116	104,116	104,116	104,116	104,116	104,116	104,116
Mean Outcome	0.299	0.299	0.299	0.299	0.299	0.299	0.299
	First Preference Field of Study is Arts and Social Sciences						
Math Rank	-0.034	-0.054*	-0.047*	-0.060**	-0.066**	-0.051*	-0.064**
	(0.027)	(0.028)	(0.027)	(0.028)	(0.027)	(0.027)	(0.029)
English Rank	0.085***	0.076***	0.080***	0.072***	0.049**	0.071***	0.058**
	(0.023)	(0.025)	(0.024)	(0.025)	(0.023)	(0.024)	(0.026)
R-squared	0.170	0.170	0.170	0.170	0.170	0.170	0.170
Observations	104,116	104,116	104,116	104,116	104,116	104,116	104,116
Mean Outcome	0.197	0.197	0.197	0.197	0.197	0.197	0.197

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise. Columns (1) – (4) vary the proportion of non-applicants assumed to be missing randomly; remaining non-applicants are assumed to come from the bottom of the distribution.

Table 5: Robustness Checks – Specification Checks

VARIABLES	(1) Omit Overall Rank	(2) Omit English Rank	(3) Omit Math Rank	(4) Interact Grades with School Characteristics	(5) Interact Grades with Mean Achievement	(6) Interact Grades with SD of Achievement	(7) Triple Interact Grades with mean and SD of Achievement
First Preference Field of Study is STEM							
Math Rank	0.087***	0.103***		0.125***	0.169***	0.080**	0.134***
	(0.027)	(0.029)		(0.031)	(0.035)	(0.031)	(0.039)
English Rank	-0.052**		-0.044*	-0.040	-0.043	-0.051*	-0.055
	(0.024)		(0.025)	(0.027)	(0.031)	(0.028)	(0.035)
R-squared	0.312	0.312	0.312	0.317	0.313	0.313	0.315
Observations	104,116	104,116	104,116	104,116	104,116	104,116	104,116
Mean Outcome	0.299	0.299	0.299	0.299	0.299	0.299	0.299
First Preference Field of Study is Arts and Social Sciences							
Math Rank	-0.114***	-0.066**		-0.058*	-0.070**	-0.105***	-0.093**
	(0.026)	(0.029)		(0.031)	(0.032)	(0.030)	(0.038)
English Rank	0.045*		0.069***	0.079***	0.102***	0.076**	0.069*
	(0.025)		(0.026)	(0.024)	(0.031)	(0.029)	(0.038)
R-squared	0.170	0.170	0.170	0.175	0.171	0.171	0.172
Observations	104,116	104,116	104,116	104,116	104,116	104,116	104,116
Mean Outcome	0.197	0.197	0.197	0.197	0.197	0.197	0.197

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

Omitting Overall Rank

In column (1) of Table 5, we omit the control for overall rank as measured by the within school-cohort percentile rank of total Leaving Certificate points. We included this variable because omitting the control for overall rank might lead us to ascribe the effects of overall rank to math rank or English rank. When we exclude this variable as a control, we obtain quite similar estimates to those in Table 2, suggesting that this is not a major concern.

Omitting Subject Rank

In columns (2) and (3) of Table 5, we omit controls for rank in English and math, respectively. Given the correlation between these two variables is 0.55, it is interesting to see how this impacts their coefficients. We see that it has very little effect on the estimates.

Interacting School Characteristics with Grades

We saw earlier that subject rank effects are identified in our model so long as subject rank cannot be written as an additive function of school-cohort indicators and subject grade indicators. As such, there are numerous sources of identification including differences in mean achievement across school-cohorts, differences in variances of achievement across school-cohorts, and differences in higher-order moments across school-cohorts. In this set of robustness checks, we eliminate some of these sources of variation to see how this affects the estimates.

First, we interact math and English grades with school characteristics -- school-cohort size terciles, whether it is a mixed-sex school, and the type of school (whether it is a fee-paying school, whether it is a DEIS (disadvantaged) school, and whether it is a Secondary, Vocational, Comprehensive, or Irish-language school).[39] By interacting subject grades with these school-

[39] There are several different types of post-primary schools in Ireland including secondary schools (both non-fee-paying and fee-paying), vocational schools, and community or comprehensive schools. Most students attend

type indicators, we eliminate certain types of identifying variation such as from a particular math grade in a fee-paying school leading to lower rank than the same math grade in a disadvantaged (DEIS) school.[40] In column (4) of Table 5, we see the estimates are robust to this change in specification.

In column (5), we remove identifying variation that comes from mean differences in achievement in math and English across school-cohorts by interacting grades in each subject with mean achievement in that subject in the school-cohort.[41] This leads to a slight increase in the positive math rank effect of STEM and a slight increase in the positive English rank effect for ASSc. In column (6), we similarly eliminate identification coming from differences in the standard deviation of achievement across school-cohorts and find that this has very little impact on outcomes. Finally, in column (7), we include interactions of subject grades with both the mean and the standard deviation of subject achievement in the school-cohort and, further, include triple interactions of subject grades with the mean and standard deviation of subject achievement in the school-cohort. Once again, we find quite similar estimates. This is reassuring as it suggests that our rank effects are robust to relying on identification from higher-order and idiosyncratic variation in subject grade distributions across school-cohorts.[42]

---

secondary schools. These are privately owned and managed but largely funded by the state. Most do not charge fees, but there is a set of secondary schools that are partially funded by student fees (typically around €6,000 per year) and tend to attract students from disproportionately affluent backgrounds. Vocational schools and community colleges are owned by the local Education and Training Board. They do not charge fees and tend to focus more on technical education than secondary schools. Community or comprehensive schools were often established through the amalgamation of secondary and vocational schools. These are all free, are fully funded by the state, and offer a wide range of academic and technical subjects. Many schools that attract students from relatively deprived backgrounds have been designated as "DEIS" schools and these receive extra supports from the state (somewhat lower pupil-teacher ratios and extra state funding for other purposes). Irish-medium post-primary schools, "Gaelscoileanna", have become more common in recent years and teach all subjects through the Irish language. See Doris et al. (2019) for further information about Irish post-primary schools.

[40] Delaney and Devereux (2020c) have shown that there are large differences in college application behaviour across these school types with students from advantaged schools displaying more ambition.

[41] We calculate mean achievement in each subject by translating grades into points (see Appendix Table A1) and calculating the average points in the subject in the school-cohort.

[42] Also, these additional controls capture many types of non-linear peer effects and, so, make it less likely that our rank estimates are confounded by some type of non-linear peer effect. See Booij et al. (2017) and Bertoni and Nistico (2019).

Balancing Tests

Our assumption is that, conditional on the controls, subject ranks are orthogonal to other factors that affect college major choice. Given rank is not randomly assigned, this cannot be guaranteed. We have information on two predetermined student characteristics – age and gender – and we do balancing tests using these as dependent variables. We also use these variables to create variables for predicted STEM probability and predicted Arts and Social Science (ASSc) probability by regressing college major choices on gender, age, school-cohort indicators, and the controls for absolute achievement. In Appendix Table A5, we show that while subject ranks have no predictive power for student age, there is a negative relationship between math rank and female. Likewise, math rank is related to predicted STEM (but not to predicted ASSc). Reassuringly, the effect of math rank on predicted STEM is very small (-0.005). This is about 20 times smaller than the effect of math rank on STEM (Table 2) and is also opposite in sign. In any case, we control for gender and age in all regressions.[43]

Enrollment Effects

So far, we have analyzed the college program listed as first choice by applicants. Next, we verify that we find similar results if we use the sample of persons who accept a program and enroll in college (73% of applicants). There are many reasons why applicants may not end up enrolling in college. They may not satisfy the program requirements and required points for any of their listed college programs and thus may not be offered any college program.[44] Alternatively, students may decide not to enter college despite being offered a college program. One reason is that the student was not offered their first choice (about 80% of non-enrolees

[43] We test whether gender affects our estimates by excluding controls for gender from the regression and find that the estimates are not sensitive to the exclusion of gender. This suggests that any correlation between gender and rank has minimal effect on our rank estimates.

[44] While we do not have data on what college programs were offered (with the exception of the program enrolled in), we can infer programs that were offered based on a combination of Leaving Certificate points and the program required points cut-offs. Using this information, we find that, of those who did not enrol in any college program, 30% were not offered a program.

were not offered their top choice) or preferred field of study and so decided to repeat the Leaving Certificate exams and reapply the following year. Students may also decide to study abroad (as discussed earlier, students generally apply to the CAO even if they also apply and prefer to go abroad). In addition, there are Post Leaving Certificate (PLC) courses which cater to students who are less academically inclined and offer a mixture of practical work, academic work, and work experience. Finally, many students decide to enter the labour force without any further education.

In theory, the effect of rank on the enrolled field of study may differ from that for first preferences as there may be selection in terms of the students that end up going to college and, additionally, those with different ranks may choose to list programs differently on the CAO form. For example, those who are higher ranked might be more ambitious and more likely to list programs for which they are unlikely to get sufficient points.

Table 6 shows the regression results when restricting the sample to students who enroll in a program. In columns (1) and (2), the dependent variable is the type of college program accepted by the student. The results for math rank on STEM enrollment are similar to those for having STEM as first preference. However, the effect of English rank is smaller and no longer significant while, for ASSc enrollment, math rank has a larger negative effect. These differences could result from our more selected sample as we only include persons who enroll in a college program. We examine this in columns (3) and (4) by showing estimates for first preference field of study for the sample who enroll. These estimates are quite similar to those for the full sample, suggesting that subject rank may have slightly different effects on field of study enrollment than it does on first preference field of study.[45]

[45] There is no effect of English or math rank on whether a student enrolls in college. Therefore, our results are not due to rank affecting overall enrollment.

Table 6: Rank in Math and English on Field of Study Enrollment

	Effect on Enrollment in Field of Study		**Effect on First Preference Field of Study for Enrollment Sample**	
	(1)	(2)	(3)	(4)
VARIABLES	STEM	Arts & Soc	STEM	Arts & Soc
Math Rank	0.089**	-0.109***	0.082**	-0.074**
	(0.034)	(0.037)	(0.033)	(0.034)
English Rank	-0.016	0.059**	-0.044	0.055*
	(0.029)	(0.028)	(0.028)	(0.030)
Observations	75,939	75,939	75,939	75,939
R-squared	0.338	0.206	0.343	0.181

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise. The sample is restricted to students who enroll in a college program.

External Validity

Our sample is restricted to school-cohorts in which at least 75% of students apply to college. While this restriction reduces measurement error in the subject ranks and maintains internal validity, it may imply that our estimates are not representative of Irish high school students in general. We address this issue using inverse-probability weighting, using observable characteristics of school-cohorts to generate the weights.

We calculate weights using the following procedure: First, we carry out a school-cohort level logit regression in which the dependent variable is an indicator for whether the school-cohort is included in our estimation sample. So that the estimates from the logit are representative of students, we weight each observation by the number of students who sit the Leaving Certificate in that school-cohort. The controls we include to predict sample inclusion are school-cohort size terciles, whether it is a mixed-sex school, and the type of school – whether it is a fee-paying school, whether it is a DEIS school, and whether it is a Secondary, Vocational, Comprehensive, or Irish-language school. Using the estimated logit coefficients, we form the propensity score and use this to weight our 75%+ sample by $\left(\frac{1}{p}\right)$, where p is the estimated propensity score.[46] This re-weighting makes the students in our sample more similar to students in general by putting relatively more weight on students who are in school-cohorts that are similar to school-cohorts that are excluded from our sample. Reassuringly, we find similar results (shown in Appendix Table A7) to our main estimates in Table 2.

Varying the School-level Application Threshold

We restrict the sample to schools in which at least 75% of applicants apply to college to reduce measurement error in the subject ranks. However, our choice of 75% is quite arbitrary and, therefore, we do additional robustness checks to examine the sensitivity of our estimates to varying the threshold. Appendix Table A8 shows the results where we limit the sample to schools with at least 50%, 65%, 85%, or 90% of students being college applicants. We find that the estimates are lower when we decrease the CAO-proportion threshold which is likely due to greater measurement error in subject rank when we have greater numbers of non-applicants. On the other hand, as we increase the threshold, the sample size decreases, and we have larger standard errors. Overall, the difference in estimates is not very large across these different samples, suggesting that the choice of a 75% threshold is not important to our findings.

6. Math and English Rank and the Gender Gap in STEM

In this section, we examine whether differential ranks in English and math by gender have significant explanatory power for the gender gap in the choice of STEM as a college major. There are two stylized facts that may be influenced by math and English rank (see

[46] The logit model has strong predictive power. The pseudo R^2 is 0.29 and the predicted probability of being in the sample is over 0.5 for 84% of sample members. The estimates are shown in Appendix Table A6.

Appendix Table A9).[47] First, boys are more likely than girls to list STEM as their first preference (the gender gap in our sample is 21 percentage points) and, second, the gender gap is larger in mixed-sex schools (25 percentage points) than in the sample of same-sex schools (16 percentage points). Conceptually, rank could explain both facts to some extent given that subject ranks differ between boys and girls in mixed-sex schools but not in same-sex schools -- about 58% of our sample attend mixed-sex schools, the remainder come from schools that either enroll only boys or only girls. By definition, average ranks are the same for boys and girls in same-sex schools.

Table 7 shows how ranks vary by sex in mixed-sex schools. As expected, females have higher rank in English (by 9 percentage points), but males have higher rank in math (by 3 percentage points).[48]

Table 7: Average Ranks by Gender in Mixed-sex Schools

	Within School-Cohort Rank	
	Female	**Male**
Math Rank	0.558	0.584
English Rank	0.617	0.531
N	27,431	29,876

Note: Average ranks are greater than 0.50 due to our assumption that missing observations come from the bottom of the school-cohort achievement distribution.

Rank Effects in Mixed-sex Schools

Because effects of rank may differ between same-sex and mixed-sex schools, we begin by estimating the main specification on a sample of mixed-sex schools. There may be

[47] In Appendix Table A9, we report the gender gap in STEM by school type. For each type, we first show the female coefficient without controls (the raw gender gap) and then the female estimate with controls for achievement. The gender gap in mixed-sex schools is 25.3 percentage points without controls and falls to 9.7 percentage points with controls for absolute achievement.

[48] Indeed, female and male subject ranks depend on the proportion of males in mixed-sex schools. Appendix Table A10 shows that an increase in the proportion of males in a school-cohort of 0.1 decreases the average math rank of females and males by almost a percentile.

differential effects in same-sex and mixed-sex schools for a variety of reasons. One possibility is that girls (boys) mostly compare themselves to other girls (boys) in mixed-sex schools, perhaps because students have people from the same gender in their social circle and, so, within-gender ranks are more salient. If this is the case, we would find that, in mixed-sex schools, the effect of own-gender rank within a school-cohort is larger than the effect of overall school-cohort rank. We have tested for this (Appendix Table A11) and found that overall school-cohort rank is more important than own-gender rank so we do not believe that this is an important consideration. Another possibility is that the presence of members of the opposite sex affects behavior.

In Table 8, we show that there are no statistically significant differences in the effects of math or English rank on STEM between mixed-sex and same-sex schools (column (1)) When we split the samples by gender, we find the effect of subject ranks on STEM are larger for boys than girls in mixed-sex schools and also tend to be larger for girls than boys in same-sex schools. However, the standard errors are quite high, and the only statistically significant gender gap is that the effect of English rank on STEM is larger for boys than for girls in mixed-sex schools. We also find no evidence for gender differences in subject rank effects on ASSc in either type of school. The larger effect of subject rank for boys in mixed-sex schools may be due to their behavior being influenced by the presence of members of the opposite sex. For example, boys may be more likely to behave competitively and display more masculine characteristics when surrounded by girls.[49]

[49] Sullivan (2009) finds that boys in mixed-sex schools display a higher math self-concept and a lower English self-concept than boys in same-sex schools.

Table 8: Rank in Math and English and First Preference Field of Study by School Gender-mix

	STEM			ASSc		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All	Males	Females	All	Males	Females
			Mixed-sex Schools			
Math Rank	0.128***	0.143**	0.090*	-0.065*	-0.085*	-0.053
	(0.040)	(0.059)	(0.049)	(0.039)	(0.048)	(0.063)
English Rank	-0.063*	-0.128**	0.032	0.058*	0.098***	0.011
	(0.034)	(0.051)	(0.044)	(0.031)	(0.037)	(0.051)
Observations	57,307	29,876	27,431	57,307	29,876	27,431
R-squared	0.315	0.281	0.296	0.183	0.206	0.195
Mean Outcome	0.314	0.436	0.183	0.193	0.156	0.234
			Same-sex Schools			
Math Rank	0.078*	-0.000	0.101*	-0.087**	-0.064	-0.116*
	(0.044)	(0.080)	(0.051)	(0.044)	(0.065)	(0.066)
English Rank	-0.045	-0.009	-0.028	0.101**	0.078	0.114*
	(0.038)	(0.052)	(0.050)	(0.042)	(0.059)	(0.063)
Observations	46,809	20,642	26,167	46,809	20,642	26,167
R-squared	0.315	0.293	0.318	0.164	0.170	0.169
Mean Outcome	0.279	0.369	0.209	0.201	0.169	0.226

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

The Gender Gap in STEM in Mixed-sex Schools

We analyze how gender differences in STEM relate to rank differences by multiplying the effect of math rank on STEM by the average difference in math rank between boys and girls and multiplying the effect of English rank on STEM by the average difference in English rank between boys and girls. Adding these gives an estimate of how much the gender gap in preferences for STEM would be reduced in mixed-sex schools if boys and girls had the same ranks in both these subjects, while holding their absolute levels of academic achievement constant.[50] Using the coefficients for mixed-sex schools from Column (1) of Table 8, the amount explained by the rank variables is given by:

$$\beta_{\text{Mathrank}} * (\text{Mathrank}_{\text{Male}} - \text{Mathrank}_{\text{Female}}) + \beta_{\text{Englishrank}} * (\text{Englishrank}_{\text{Male}} - \text{Englishrank}_{\text{Female}})$$

Table 9 shows the differences in the gender gap in STEM explained by differential gender ranks is 0.9 percentage points, compared to the 25 percentage point gender gap in STEM in mixed-sex schools. We conclude that, in Ireland, the tendency for girls to be lower ranked in math and higher ranked in English can explain about 4% (0.9/25) of the gender gap in preferences for STEM in college.[51] If we focus on the 10 percentage point unexplained gender gap in mixed-sex schools after controlling for absolute achievement (see Appendix Table A9), the rank variables can account for 9% of the unexplained gap.[52] Overall, subject rank can account for a relatively small percentage of the gender gap in STEM in mixed-sex schools.

Table 9: Proportion of the Gender Gap in STEM in Mixed-Sex Schools explained by English and Math Rank

Percentage Points (standard error)	**Percent of Overall Gap**	**Percent of Unexplained Gap**	**Percent of Difference in Gender Gap between Mixed-sex and Same-sex Schools**	**Percent of Unexplained Difference in Gender Gap between Mixed-sex and Same-sex Schools**
0.88 (0.3)	3.5	9.1	9.5	24

The overall gender gap in first preference for STEM in mixed-sex schools is 25.3 pp and is 16pp in same-sex schools. Adding controls for subjects, grades and points reduces this gap to 9.7pp and 6pp, respectively.

[50] This exercise shows what happens to the predicted probabilities of STEM when we equalize the subject ranks, holding all other variables constant. It should not be seen as representing the effects of a particular policy intervention as, in practice, it is not clear how a policy maker could intervene in such a way as to change the subject ranks in this way while holding all else fixed. Note, however, that boys and girls can have different average ranks even if they have the same performance so long as average performance differs across schools and there are different proportions of girls and boys in different schools.

[51] Murphy and Weinhardt (2020) find that subject ranks in primary school explain about 0.66 percentage points (7%) of the STEM-gender gap in A-levels in the UK. However, their estimate is not directly comparable as it does not condition on absolute achievement levels at the point when the field choice is being made.

[52] We find very similar results if we do the calculation using the non-linear specification that includes indicator variables for each ventile of the subject rank distributions.

Difference in the Gender Gap between Mixed-sex and Same-sex Schools

Ireland has a mix of same-sex and mixed-sex schools and there is a sizeable difference in the STEM gender gap between the two types of schools – 25 percentage points in mixed-sex schools versus 16 percentage points in same-sex schools. The rank variables can explain about 10% (0.9/9.3) of the difference between the STEM gender gap in mixed-sex schools versus same-sex schools and about 24% (0.9/3.7) of the difference in the unexplained gender gap between these two types of schools. Thus, within school-cohort ranks in English and math can explain some portion of the larger gender gap in choice of STEM as a college major in mixed-sex schools compared to same-sex schools.

7. Conclusions

We draw three main conclusions from our analysis. First, conditional on achievement at the end of high school, within school-cohort percentile ranks in English and math are predictive for college field choice, particularly for STEM and Arts and Social Sciences -- higher English rank is positively associated with choosing Arts and Social Sciences and negatively with STEM; higher math rank is positively associated with STEM and negatively with Arts and Social Sciences. Second, the effects of subject ranks on STEM are larger for boys; there is no evidence of a gender difference in the effect of subject ranks on ASSc. Third, the subject rank effects can explain about 4% of the gender gap in the choice of STEM as a college major in mixed-sex schools and 10% of the difference in the STEM gender gap between students from mixed-sex and same-sex schools. Notably, these effects occur even though we control for an extensive set of measures of absolute achievement at the end of high school, and the institutional setup implies that within-school rank plays no role whatsoever in college admissions decisions. Our findings imply behavioral effects of subject rank that go beyond their effects on human capital accumulation in school.

Our results are important as research has found long-run effects of field of study on earnings. Kirkeboen et al. (2016) find that choice of field of study in college is potentially as relevant to future earnings as the decision to enroll in college, and the payoff to a STEM degree is typically much larger than to an Arts or Social Science degree. So, math and English rank within school-cohorts may have implications for future earnings trajectories and for the gender-earnings gap. The results suggest a role for information provision such that high school students are made more aware of their absolute achievement in math and English relative to that of others in the nation. This is important as students may be in a high school cohort that is atypical in terms of the math and English grade distribution and therefore may inadvertently choose college majors to which they are not well matched. Providing information on where students stand in their overall cohort may help them to make better and more informed decisions.

References

Altonji, J. G., Arcidiacono, P. and Maurel, A. (2016) “The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects.” *Handbook of the Economics* of Education. Vol. 5. Elsevier, 2016: 305–396.

Astorne-Figari, C., and J. D. Speer. (2018). "Drop out, Switch Majors or Persist? The Contrasting Gender Gaps". *Economic Letters* 164: 82-85.

Aucejo, E. M. and James, J. (2019). ‘The Path to College Education: The Role of Math and Verbal Skills’. Working Paper.

Azmat, G, Bagues Fernandez, M, Cabrales, A & Iriberri, N (2019). "What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education", *Management Science,* vol. 65, no. 8, pp. 3714-3736.

Azmat, G. & Iriberri, N. (2010). “The importance of relative performance feedback information: Evidence from a natural experiment using high school students” *Journal of Public Economics*, 94(7-8), p. 435–452.

Bertoni, M. & Nisticò, R. (2019). Ordinal Rank and Peer Composition: Two Sides of the Same Coin? IZA DP No. 12789.

Bobba, M., & Frisancho, V. (2014). "Learning About Oneself: The effects of Signaling Academic Ability on School Choice". Inter-American Development Bank, mimeo.

Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability Peer Effects in University: Evidence from a Randomized Experiment. *Review of Economic Studies*, 84(2), 547-578.

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2019). "Beliefs about Gender." *American Economic Review*, 109 (3): 739-73.

Breda, T., and C. Napp (2019). "Girls’ comparative advantage in reading can largely explain the gender gap in math-related fields". *Proceedings of the National Academy of Sciences,* July, 116 (31) 15435-15440.

Bursztyn, L. and Jensen, R. (2015). "How Does Peer Pressure Affect Educational Investments?" *The Quarterly Journal of Economics*, 130(3):1329–1367.

Buser, T., Noemi P., and S.C. Wolter. (2017). "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review,* 107 (5): 125-30.

Card, D. and Payne, A. A. (2017). High school choices and the gender gap in STEM. NBER Working paper 23769.

Delaney J.M. and P.J. Devereux (2019). "Understanding Gender Differences in STEM: Evidence from College Applications." *Economics of Education Review,* Volume 72, October 2019, Pages 219-238.

Delaney, J.M. and P.J. Devereux, (2020a), "Math matters! The importance of mathematical and verbal skills for degree performance", *Economics Letters,* 186, issue C.

Delaney, J.M. and P.J. Devereux, (2020b), "Gender Differences in College Applications: Aspiration, Risk, and Coherence", Working Paper.

Delaney, J.M. and P.J. Devereux, (2020c), "Choosing Differently? College Application Behavior and the Persistence of Educational Advantage", *Economics of Education Review,* Volume 77, August 2020, 101998.

Denning, J. T., Murphy, R., and Weinhardt, F. (2018). "Class Rank and Long-run Outcomes", No. 11808, IZA Discussion Paper.

Doris, A., D. O'Neill, and O. Sweetman, (2019). "Good Schools or Good Students? The Importance of Selectivity for School Rankings" *Economics, Finance and Accounting Department Working Paper Series* n293-19.pdf, Department of Economics, Finance and Accounting, National University of Ireland - Maynooth.

Elsner, B., and Isphording, I. E. (2017). "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics*, 35(3), 787-828.

Elsner, B., and Isphording, I. E. (2018). “Rank, Sex, Drugs and Crime”, *Journal of Human Resources*, 53(2), 356-381.

Elsner, B., Isphording, I.E. & Zolitz, U. (2019). “Achievement Rank Affects Performance and Major Choices in College” University of Zurich, Department of Economics, Working Paper.

Friedman-Sokuler, N., and M. Justman. (2016). "Gender Streaming and Prior Achievement in High School Science and Mathematics". *Economics of Education Review* 53: 230-253.

Kinsler, J. and Pavan, R. (2020), “Local Distortions in Parental Beliefs over Child Skill. *Journal of Political Economy (*Forthcoming).

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, 131(3), 1057–1111.

McDool, E. and D. Morris (2020). "Gender and Socio-Economic Differences in STEM Uptake and Attainment". Centre for Vocational Education Research. LSE Discussion Paper. No. 29.

Mouganie, P. and Y. Wang (2020). "High Performing Peers and Female STEM Choices in School". *Journal of Labor Economics,* 38(3).

Murphy, R. and Weinhardt, F. (2020) “Top of the Class: The importance of ordinal rank”. Forthcoming in *Review of Economic Studies*.

Nollenberger, N., N. Rodríguez-Planas, and A. Sevilla (2016). "The math gender gap: The role of culture." *The American Economic Review* 106 (5), 257–261.

Pop-Eleches, C., & Urquiola, M. (2013). "Going to a Better School: Effects and Behavioral Responses". *American Economic Review*, 103(4), 1289-1324.

Rask, K. and Tiefenthaler, J. (2008). “The role of grade sensitivity in explaining the gender imbalance in undergraduate economics”. *Economics of Education Review*, Volume 27, Issue 6. December 2008, pages 676-687.

Reuben, E., Wiswall, M. and Zafar, B. (2017). “Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender”. *Economic Journal,* 127: 2153-2186. doi:10.1111/ecoj.12350

Shi, Y. (2018). "The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs and Preferences". *Economics of Education Review*. 64: 129-143.

Speer, J. D. (2017). “The Gender Gap in College Major: Revisiting the Role of Pre-College Factors", *Labour Economics* 44, p. 69-88.

Stinebrickner, T., & Stinebrickner, R. (2012). "Learning about Academic Ability and the College Drop-out Decision". *Journal of Labor Economics*, 30(4), 707-748.

Stinebrickner, T., & Stinebrickner, Ralph. (2014). A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *Review of Economic Studies*, 81(1), 426-472.

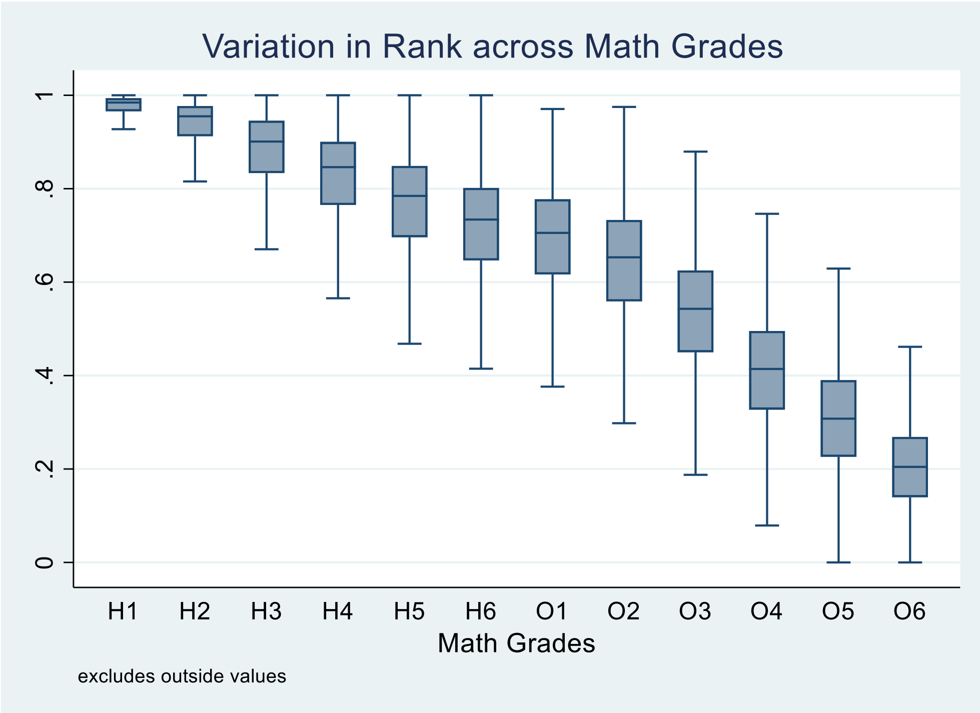
Sullivan, A. 2009. ‘Academic self-concept, gender and single-sex schooling’. *British Educational Research Journal* 35(2) 259-288.

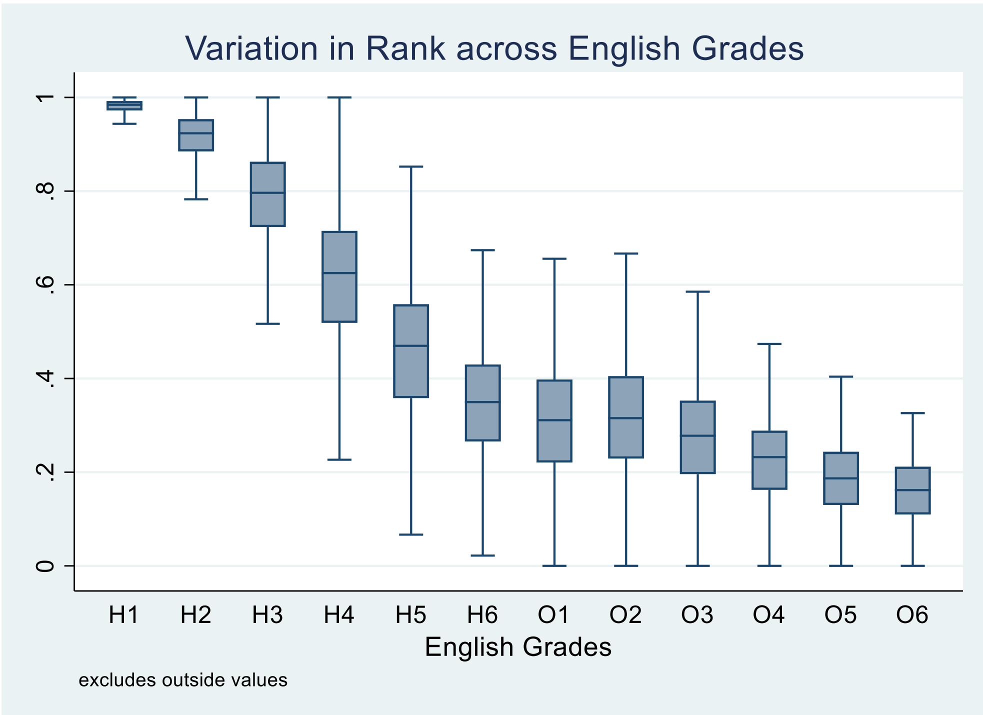
Tincani, M. M. (2015). "Heterogeneous peer effects and rank concerns: Theory and evidence". Mimeo, University College London.

Tincani, M. M., Kosse, F. and E. Miglino (2020). “Student Beliefs and the (Perverse) Incentives of Preferential College Admissions”, Working Paper.

Zafar, B. (2011). "How do College Students Form Expectations?" *Journal of Labor Economics*, 29(2), 301-348.

Appendix Figure A1: Box Plots of Variation in Rank





These box plots show variation in rank for each grade. We have converted all grades to 2017 grades here. Due to the small number of students who fail either subject, we combine fail grades at higher level (H7 and H8) with the O1 grade and combine fail grades at lower level (O7 and O8) with the O6 grade. The horizontal line in the center of each box denotes the median rank at that grade, the lower and upper bound of the box displays the 25th and 75th percentile rank, and the top and bottom of each line represents the smallest and largest rank.

Appendix Table A1: Mapping from Grades to Leaving Certificate Points

2015 and 2016

Grade	Marks (%)	Points	Points (Math)
Higher Level			
A1	90% to 100%	100	125
A2	85% to 89%	90	115
B1	80% to 84%	85	110
B2	75% to 79%	80	105
B3	70% to 74%	75	100
C1	65% to 69%	70	95
C2	60% to 64%	65	90
C3	55% to 59%	60	85
D1	50% to 54%	55	80
D2	45% to 49%	50	75
D3	40% to 44%	45	70
E	25% to 39%	0	0
F	10% to 24%	0	0
NG	0% to 9%	0	0
Lower Level			
A1	90% to 100%	60	60
A2	85% to 89%	50	50
B1	80% to 84%	45	45
B2	75% to 79%	40	40
B3	70% to 74%	35	35
C1	65% to 69%	30	30
C2	60% to 64%	25	25
C3	55% to 59%	20	20
D1	50% to 54%	15	15
D2	45% to 49%	10	10
D3	40% to 44%	5	5
E	25% to 39%	0	0
F	10% to 24%	0	0
NG	0% to 9%	0	0

2017

Grade	Marks (%)	Points	Points (Math)
Higher Level			
H1	90% to 100%	100	125
H2	80% to 89%	88	113
H3	70% to 79%	77	102
H4	60% to 69%	66	91
H5	50% to 59%	56	81
H6	40% to 49%	46	71
H7	30% to 39%	37	37
H8	0 to 29%	0	0
Lower Level			
O1	90% to 100%	56	56
O2	80% to 89%	46	46
O3	70% to 79%	37	37
O4	60% to 69%	28	28
O5	50% to 59%	20	20
O6	40% to 49%	12	12
O7	30% to 39%	0	0
O8	0 to 29%	0	0

Appendix Table A2: Identifying Residual Variation

	Standard Deviation
Variation in math rank no controls	0.258
Variation in math rank controlling for age, gender, grades, points, subjects, and school-cohort fixed effects	0.047
Variation in math rank controlling for age, gender, grades, points, subjects, and school-cohort fixed effects and overall rank and English rank	0.041
Variation in English Rank no controls	0.257
Variation in English rank controlling for age, gender, grades, points, subjects, and school-cohort fixed effects	0.054
Variation in English rank controlling for age, gender, grades, points, subjects, and school-cohort fixed effects and overall rank and math rank	0.051

This table shows the variation in the residual after regressing math and English rank on each set of control variables.

Appendix Table A3: Effect on First Preference Field of Study Allowing for Interaction of Math and English Rank

VARIABLES	(1) STEM	(2) Arts/Soc
Math Rank	0.104***	-0.102***
	(0.034)	(0.031)
English Rank	-0.040	0.054*
	(0.026)	(0.032)
Math Rank*English Rank	-0.003	0.074**
	(0.034)	(0.033)
Observations	104,116	104,116
R-squared	0.317	0.177
Mean Outcome	0.104***	-0.102***

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). Regressions also include the interaction of grades in math with grades in English. The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

Appendix Table A4: Effect of Rank in Math and English on Field of Study Preferences

	2nd Choice		3rd Choice		% of Top 3 Choices		% of All Choices	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	STEM	Arts & Soc	STEM	Arts & Soc	STEM	Arts & Soc	STEM	Arts & Soc
Math Rank	0.091***	-0.072**	0.045	-0.070**	0.077***	-0.072***	0.092***	-0.073***
	(0.031)	(0.031)	(0.032)	(0.028)	(0.025)	(0.024)	(0.021)	(0.016)
English Rank	-0.013	0.048*	-0.024	0.018	-0.037*	0.043**	-0.012	0.041***
	(0.024)	(0.026)	(0.025)	(0.025)	(0.021)	(0.021)	(0.016)	(0.014)
Observations	101,108	101,108	97,434	97,434	97,434	97,434	104,116	104,116
R-squared	0.315	0.161	0.302	0.159	0.408	0.235	0.440	0.303
Mean Outcome	0.294	0.210	0.283	0.219	0.291	0.208	0.269	0.158

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable in columns 1 and 2 (3 and 4) equals 1 if the second-choice (third-choice) college program is in the field and equals 0 otherwise. The dependent variable in columns 5 and 6 (7 and 8) is the proportion of the Top 3 choices (all choices) that are in the field of study.

Appendix Table A5: Balancing Tests

VARIABLES	(1) Age 18	(2) Female	(3) STEM prediction	(4) ASSc prediction
Math Rank	-0.045	0.073***	-0.005**	0.000
	(0.035)	(0.027)	(0.002)	(0.000)
English Rank	-0.000	0.018	-0.002	-0.000
	(0.027)	(0.022)	(0.002)	(0.000)
Observations	104,116	104,116	104,116	104,116
R-squared	0.139	0.661	0.990	0.999
Female	Yes	No	No	No
Age Dummies	No	Yes	No	No

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). Age 18 is a dummy for the individual being at least 18 years old. Age dummies and female are included in the regression that calculates predicted STEM/ASSc.

Appendix Table A6: Logit Regression for whether at least 75% of students in the School-Cohort apply to College

VARIABLES	(1) CAO Proportion at least 75%
Comprehensive/Vocational School (omitted category = DEIS)	0.329***
	(0.043)
Secondary School	0.472***
	(0.046)
Irish-medium School	0.513***
	(0.100)
Fee-Paying School	0.746***
	(0.132)
School-Cohort Size Middle Tercile	0.079***
	(0.028)
School-Cohort Size Top Tercile	0.101***
	(0.031)
Same-Sex School	-0.022
	(0.033)
Observations	2,029

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; *0.10. The reported estimates are marginal effects computed at the means. Each observation is a school-cohort and observations are weighted by the number of persons in the school-cohort. School-cohort size terciles: 10-63; 64-103; 104-275 students.

Appendix Table A7: Effect of Rank on First Preference Field of Study (Weighting by the Inverse Probability of being in the Sample)

	(1)	(2)
VARIABLES	STEM	Arts/Soc
Math Rank	0.099***	-0.058*
	(0.034)	(0.032)
English Rank	-0.054*	0.085***
	(0.029)	(0.028)
Observations	104,116	104,116
R-squared	0.310	0.172
Mean Outcome	0.30	0.20

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise. Regressions are estimated using inverse probability weighting.

Appendix Table A8: Omitting Schools with Varying Proportions Applying to College

	CAO-Prop<0.50		CAO-Prop<0.65		CAO-Prop<0.85		CAO-Prop<0.90	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	STEM	Arts & Soc	STEM	Arts & Soc	STEM	Arts & Soc	STEM	Arts & Soc
Math Rank	0.063**	-0.074***	0.088***	-0.072***	0.132***	-0.056	0.159***	-0.058
	(0.027)	(0.025)	(0.028)	(0.026)	(0.035)	(0.035)	(0.044)	(0.046)
English Rank	-0.051**	0.064***	-0.037	0.065***	-0.037	0.044	-0.087**	0.097**
	(0.022)	(0.022)	(0.023)	(0.024)	(0.032)	(0.032)	(0.041)	(0.039)
Observations	124,000	124,000	117,158	117,158	70,304	70,304	43,711	43,711
R-squared	0.310	0.167	0.311	0.168	0.320	0.174	0.330	0.180

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

Appendix Table A9: Effect of Female on Choosing STEM as First Preference by School Type

	Overall		Mixed-Sex		Same-Sex	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM	STEM	STEM	STEM	STEM	STEM
Female	-0.213***	-0.082***	-0.253***	-0.097***	-0.160***	-0.060***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.008)	(0.008)
Observations	104,116	104,116	57,307	57,307	46,809	46,809
R-squared	0.055	0.296	0.074	0.299	0.033	0.300
Grades, Subjects, and Points	No	Yes	No	Yes	No	Yes
Mean Outcome	0.299	0.299	0.315	0.315	0.279	0.279

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05 * p<0.10. Age and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). Points included as a quartic polynomial. The dependent variable equals 1 if the first-choice college program is STEM and equals 0 otherwise.

Appendix Table A10: Effect of Proportion of Males in School Cohort on Math and English Rank

	(1)	(2)	(3)	(4)
VARIABLES	Male Math Rank	Female Math Rank	Male English Rank	Female English Rank
Proportion Males in School-Cohort	-0.069***	-0.094***	0.065***	0.042***
	(0.012)	(0.012)	(0.017)	(0.016)
English Rank	Yes	Yes	No	No
Math Rank	No	No	Yes	Yes
School-Cohort FE	No	No	No	No
Observations	29,876	27,431	29,876	27,431
R-squared	0.957	0.949	0.920	0.910

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject).

Appendix Table A11: Effect of Same-gender Rank in Mixed-sex schools on First Preference Field of Study

VARIABLES	(1) STEM	(2) Arts & Social
Same-gender Math Rank	0.013	0.013
	(0.044)	(0.042)
Same-gender English Rank	-0.006	-0.007
	(0.038)	(0.035)
Math Rank	0.114*	-0.079
	(0.061)	(0.062)
English Rank	-0.060	0.062
	(0.050)	(0.050)
Observations	56,118	56,118
R-squared	0.315	0.185
Mean Outcome	0.315	0.193

Robust standard errors clustered by school are in parentheses. *** p<0.01; ** p<0.05; * p<0.10. Age, gender, indicator variables for grades in math and English (interacted with 2017), subject indicators, grades in all subjects, overall rank and a quartic in points (interacted with 2017) and school-cohort fixed effects included in all regressions. Subject fixed effects are indicators for doing each of the 25 most popular subjects for Leaving Certificate. Grade fixed effects are grades in these 25 subjects (interacted with an indicator for doing the subject). The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise. Same-gender rank is rank calculated just using persons in the school-cohort who have the same gender. Overall rank and same-gender overall rank also included in the regressions. The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.