



## Working Paper No. 650

January 2020

### *The effect of high school rank in English and math on college major choice*

*Judith M. Delaney<sup>a,b,c</sup> and Paul J. Devereux<sup>c,d,e,f</sup>*

*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, 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 rank effects are substantial, being about one third as large as the effects of absolute performance in math and English. We identify subject choice in school as an important mediator – students who rank high in math are more likely to choose STEM subjects in school and this can partly explain their subsequent higher likelihood of choosing STEM for college. We also find that English and math rank have significant explanatory power for the gender gap in the choice of STEM as a college major--they can explain about 36% as much as absolute performance in English and math. Overall, the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 6% of the STEM gender gap in mixed-sex schools and about 16% of the difference in the STEM gender gap between mixedsex schools and same-sex schools. Notably, these effects occur even though within-school rank plays no role whatsoever in college admissions decisions.*

Corresponding author: [judith.delaney@esri.ie](mailto:judith.delaney@esri.ie)

*Acknowledgements: 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, Cormac O’Dea, Fabien Postal-Vinay, Barra Roantree, Uta Schoenberg, Michela Tincani and seminar participants at NUI Galway and Maynooth.*

- 
- a University College London
  - b Economic and Social Research Institute
  - c IZA Institute of Labor Economics
  - d School of Economics and Geary Institute, University College Dublin
  - e NHH Norwegian School of Economics
  - f Center for Economic and Policy Research

*ESRI working papers represent un-refereed work-in-progress by researchers who are solely responsible for the content and any views expressed therein. Any comments on these papers will be welcome and should be sent to the author(s) by email. Papers may be downloaded for personal use only.*

# The Effect of High School Rank in English and Math on College Major Choice\*

by

Judith M. Delaney  
University College London  
Economic and Social Research Institute and IZA  
judithmdelaney@gmail.com

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

December 2019

## 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, 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 rank effects are substantial, being about one third as large as the effects of absolute performance in math and English. We identify subject choice in school as an important mediator – students who rank high in math are more likely to choose STEM subjects in school and this can partly explain their subsequent higher likelihood of choosing STEM for college. We also find that English and math rank have significant explanatory power for the gender gap in the choice of STEM as a college major--they can explain about 36% as much as absolute performance in English and math. Overall, the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 6% of the STEM gender gap in mixed-sex schools and about 16% of the difference in the STEM gender gap between mixed-sex schools and same-sex schools. Notably, these effects occur even though within-school rank plays no role whatsoever in college admissions decisions.

---

\* 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, Cormac O’Dea, Fabien Postal-Vinay, Barra Roantree, Uta Schoenberg, Michela Tincani and seminar participants at NUI Galway and Maynooth.

## 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, 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.<sup>1</sup> 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 Frisano, 2014) and this uncertainty may lead students to infer their comparative advantage across subjects from their rank across subjects in school.<sup>2</sup> 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

---

<sup>1</sup> Studies include Speer (2017), Card and Payne (2017), Delaney and Devereux (2019), and Aucejo and James (2019).

<sup>2</sup> 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 Iriberry (2010) and Azmat et al. (2015) find that providing feedback on relative performance in school leads to better short-term performance by students.

interventions to provide 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). 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; 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.

While the recent rank literature has found strong impacts of class rank on many outcomes including earnings, high school graduation, college enrolment, and risky behavior (Murphy and Weinhardt, 2018; Denning et al., 2018; Elsner and Isphording, 2017, 2018), there has been little focus on the relationship between rank in English and math and choice of college major. Using UK data, Murphy and Weinhardt (2018) 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 don't examine choice of college major. Denning et al. (2018), using data from Texas, find that math rank in 3<sup>rd</sup> grade has a positive effect on doing a STEM major in college; however, because they give a 0 for STEM to all people who don't go to college, it is hard to disentangle the effect of math rank on college major choice from that of math rank on college enrolment. Also, 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 3<sup>rd</sup> grade.

We add to the literature in several ways. First, 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 who actually attend. As such, we can see how math

and English ranks affect desired college major for all applicants.<sup>3</sup> Second, a primary concern in the literature is that class rank may be correlated with absolute achievement, even conditional on control variables. We have grades from high-stakes exams at the end of high school (the Leaving Certificate examinations). These exams are centrally set and graded so are comparable across all students and provide a detailed description of academic readiness at the end of secondary schooling. Thus, we may be better able to control for absolute levels of academic interests and achievement than other papers in the literature. 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 significant 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 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. The magnitudes of the effects are substantial, being about one third as large as the effects of absolute performance in English and math. They also have substantial explanatory

---

<sup>3</sup> We also differ from the prior literature by studying the effect of math and English rank towards the end of high school, when college major decisions are made, rather than when students are much younger.

power for the gender gap in the choice of STEM as a college major in mixed-sex schools – they can explain about 36% as much as absolute performance in English and math. Overall, the tendency for girls to be higher ranked in English and lower ranked in math within school-cohorts can explain about 6% of the STEM gender gap in mixed-sex schools and about 16% of the difference in the STEM gender gap between mixed-sex schools and same-sex schools.

While we are limited in our ability to study mechanisms, we identify subject choice in school as an important mediator – students who are highly ranked in math are more likely to choose STEM subjects in school and this can partly explain their subsequent higher likelihood of choosing STEM in college. However, even when we control for the subjects taken and grades obtained in each subject in the Leaving Certificate, we still find that math and English rank affect college major choice. This finding implies behavioral effects of subject rank that go beyond their effects on human capital accumulation in school.

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 examines subject choice in school as a potential mechanism, and Section 7 shows that our estimates can help explain the gender gap in STEM. Finally, Section 8 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.<sup>4</sup> 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

---

<sup>4</sup> This section draws heavily from Delaney and Devereux (2019).

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.<sup>5</sup> 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.<sup>6</sup> 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.

English, Irish and math are compulsory high school subjects and the student can then choose other subjects to study. All subjects are offered at a higher or lower level. 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.

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

---

<sup>5</sup> 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.

<sup>6</sup> High school lasts 5 or 6 years (depending on whether the student does a "transition year" in year 4).

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.<sup>7</sup> 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 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).<sup>8</sup>

We assume that students are aware of where they rank in the school-cohort and that the Leaving Certificate grades are reflective of this, for example, those who are top of the class throughout the year will more than likely end up scoring highest in the Leaving Certificate exams. This assumption is plausible as we expect student peer interactions and teacher feedback to conform with the rankings.<sup>9</sup> Exams are given throughout the year in each year of high school and “mock Leaving Certificate exams” are provided to the student.<sup>10</sup>

---

<sup>7</sup> 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.

<sup>8</sup> 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).

<sup>9</sup> To the extent that individuals do not know their exact rank, in practice we will be estimating the reduced form effects of perceived rank using actual rank.

<sup>10</sup> 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.



College applications are made by May of the year of entry and students can change the programs they list until July, after they sit the Leaving Certificate exams.

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.<sup>11</sup> 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.<sup>12</sup> 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 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.

---

<sup>11</sup> 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.

<sup>12</sup> 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.

The major issue we face in calculating ranks is that we don't 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.<sup>13</sup> So, non-applicants are generally the least academically inclined students. 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.<sup>14</sup> 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 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. 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 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

---

<sup>13</sup> In addition, students who plan to defer college (take a gap year) are encouraged to apply anyhow in case they change their mind.

<sup>14</sup> 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.

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. 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.<sup>15</sup>

*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 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
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 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

<sup>15</sup> 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.

### 3. Methodology

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 (over 95% of students list at least one level 8 program).

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. Delaney and Devereux (2019) found significant differences between how girls and boys responded to comparative advantage in math and English when choosing whether to do STEM. For this reason, and because we later emphasize specifications in which we allow rank effects to differ by gender, we interact the subject grade indicators with a gender indicator.

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 (2018) for further discussion on this point).

There remains the possibility that parental investment or other factors that affect college major choice may differ by rank. Students may develop confidence from a higher class rank or despondence from a lower one.<sup>16</sup> More highly ranked students may also receive different levels of encouragement (or different advice about college major choice) from peers, teachers, or family.<sup>17</sup> However, if these processes operate through achievement, they do not cause biases so long as rank remains quasi-randomly assigned conditional on achievement and school-cohort indicators. If they operate independently of achievement (such as a parent suggesting a college major because of the student’s class rank in math), then we consider this as a mechanism rather than a confounder.

*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

$$Y_{cg} = \alpha + \beta RM_{cg} + v_{cg}. \tag{1}$$

Then, the critical identifying assumption is that

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

---

<sup>16</sup> 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.

<sup>17</sup> 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 (2016) 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.

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):

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

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 described above, and  $\theta_c$  represents school-cohort fixed effects. 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.

#### **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 performance in math and English, within school-cohort ranks in English and math are associated 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 1.3 percentage point increase in the probability of listing STEM as first preference and a 1.5 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 0.9 percentage points and increases the probability of listing ASSc by 0.8 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.

**Table 2: Effect of Rank in Math and English on Choice of Field of Study**

VARIABLES	(1) STEM	(2) BAL	(3) Arts & Social	(4) Other
Math Rank	0.133*** (0.031)	0.041 (0.027)	-0.147*** (0.028)	-0.027 (0.030)
English Rank	-0.089*** (0.027)	0.045* (0.024)	0.076*** (0.027)	-0.032 (0.028)
Observations	104,116	104,116	104,116	104,116
R-squared	0.187	0.053	0.092	0.111
Mean Outcome	0.299	0.208	0.197	0.296

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. The dependent variable equals 1 if the first-choice college program is in the field and equals 0 otherwise.

### *Comparison to Absolute Percentile Rank*

Delaney and Devereux (2019) have shown that math and English grades are very important in predicting STEM choices in Ireland. To further assess the magnitudes of the



rank effects, we compare them to the absolute effect of scoring well in English and math. For comparability, we translate math and English grades to population-year percentiles, and we regress the outcomes on these variables. The estimates are in Appendix Table A3. We see the effect of absolute rank in math on listing a STEM program as first preference is 0.53 and that of absolute rank in English is -0.25. These contrast to the equivalent school-cohort rank effects of 0.13 and -0.09. So, the school-cohort rank effects on choosing STEM are about 3 to 4 times smaller than the effects of absolute achievement in math and English. Similarly, for choosing ASSc, the effects of within school-cohort English and math rank are about 2 to 4 times smaller than the absolute subject rank effects. Given that we are controlling for absolute achievement, we consider the magnitudes of the school-cohort rank effects to be substantial and they suggest meaningful behavioral responses to within school subject rank.

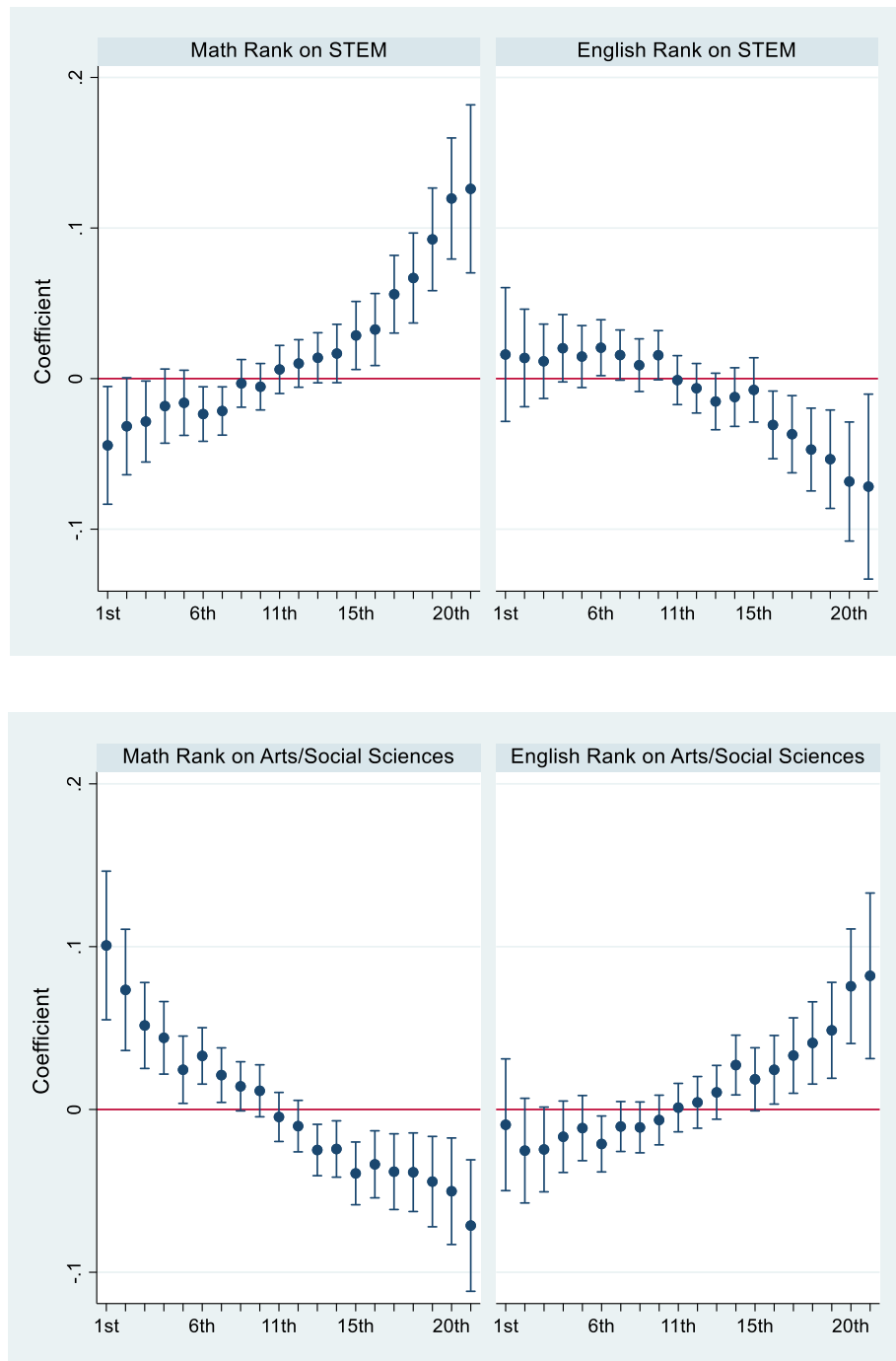
#### **4.1 Heterogeneous Effects**

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<sup>th</sup> 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: Rank Ventiles & Top Ranked Person in Math and English and 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 10<sup>th</sup> ventile. Point estimates and 95% confidence intervals are shown.

### *Effects by Gender*

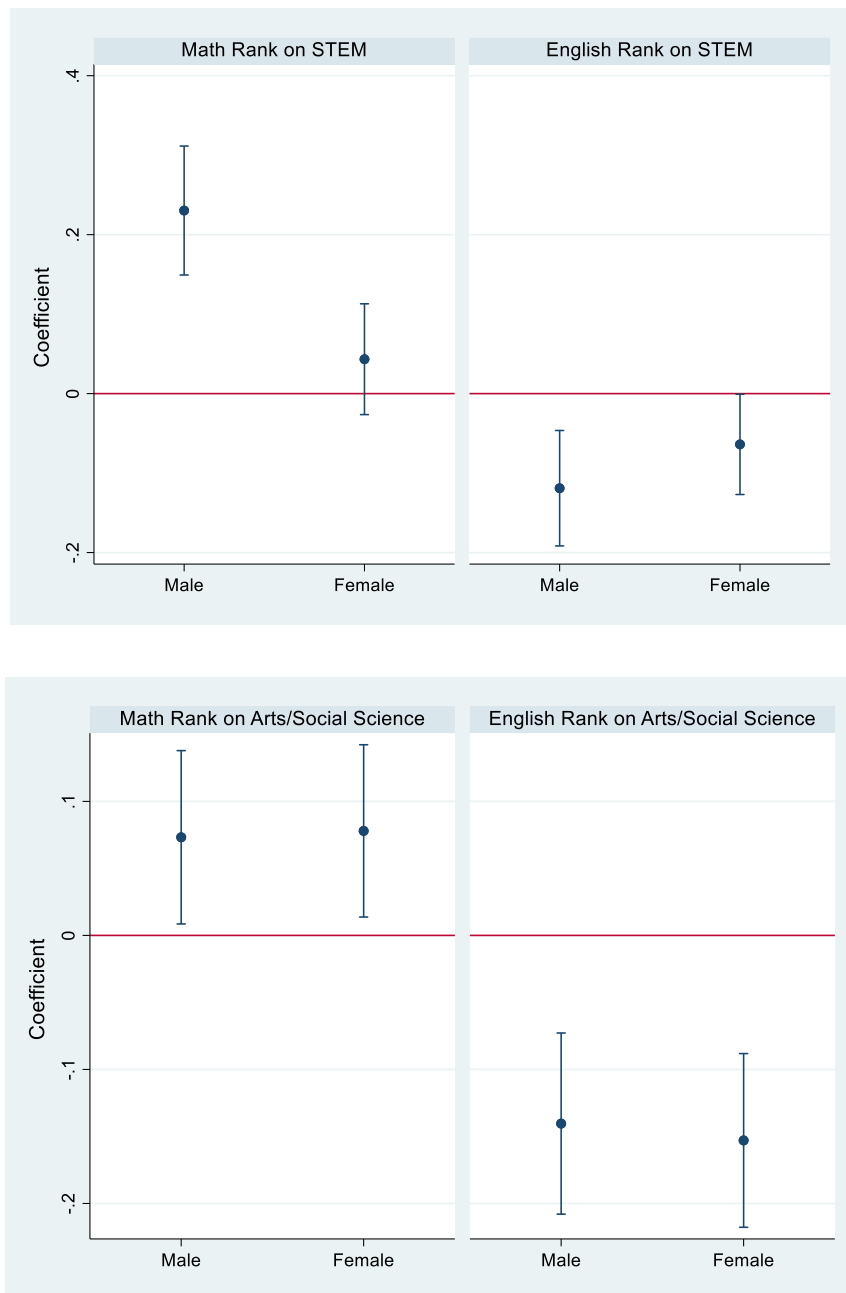
Some previous literature has found that boys are more competitive than girls (Buser et al. 2017), and so it is likely that they care more about their rank. Therefore, we might expect that the effect of rank is larger for males. However, it is not clear if this will hold across both subjects as the superior performance of girls in English may make them more sensitive to their rank in that subject.

To estimate differential effects by gender, we interact English and math rank with indicators for male. As mentioned earlier, 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 much larger for boys than for girls (0.23 versus 0.04) and this difference is statistically significant at the 1% level. This is consistent with previous literature that found larger effects of rank for males than females (Murphy and Weinhardt, 2018).<sup>18</sup> Interestingly, this effect only appears for the effect of math rank on choosing STEM. There does not appear to be any large gender differences in the effect of English rank on STEM or in the effect of math or English rank on choosing Arts and Social Sciences.

---

<sup>18</sup> Elsner et al. (2019) find that the effect of rank is larger for males in a college tutorial setting whereby males respond to their rank by increasing their study effort in contrast to females. However, Denning et al. (2018) find no evidence of gender heterogeneity in their study of Texas schools.

**Figure 2: Effect of Rank in Math and English on 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.<sup>19</sup> Consistent with our prior, Table 3 shows that the subject rank estimates are larger in absolute value for the small schools; however, while some of the differences are substantial, none of them are statistically significant.

**Table 3: The Effect of Rank in Small versus other School-Cohorts**

VARIABLES	Less than or equal to 60 students		More than 60 students	
	(1) STEM	(2) Arts & Social	(3) STEM	(4) Arts & Social
Math Rank	0.197*** (0.057)	-0.168*** (0.057)	0.115*** (0.036)	-0.143*** (0.032)
English Rank	-0.155*** (0.057)	0.095* (0.051)	-0.068** (0.030)	0.063** (0.031)
Observations	14,739	14,739	89,377	89,377
Number of School Cohorts	437	437	972	972
R-squared	0.210	0.132	0.185	0.086
Mean Outcome	0.306	0.202	0.297	0.196

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions.

## 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.

### *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 don't think is realistic; however, it informs about how important the treatment of non-applicants could be for our estimates. The estimates displayed in column

<sup>19</sup> 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.

(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.

**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
<b>First Preference is STEM</b>							
Math Rank	0.140*** (0.028)	0.140*** (0.030)	0.142*** (0.029)	0.138*** (0.030)	0.101*** (0.029)	0.137*** (0.029)	0.104*** (0.031)
English Rank	-0.103*** (0.025)	-0.097*** (0.026)	-0.100*** (0.026)	-0.094*** (0.026)	-0.070*** (0.023)	-0.084*** (0.026)	-0.076*** (0.027)
R-squared	0.187	0.187	0.187	0.187	0.187	0.187	0.187
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
<b>First Preference is Arts and Social Sciences</b>							
Math Rank	-0.092*** (0.026)	-0.124*** (0.027)	-0.113*** (0.027)	-0.135*** (0.028)	-0.137*** (0.026)	-0.124*** (0.027)	-0.140*** (0.028)
English Rank	0.106*** (0.024)	0.090*** (0.025)	0.097*** (0.025)	0.084*** (0.026)	0.053** (0.024)	0.079*** (0.025)	0.060** (0.027)
R-squared	0.091	0.092	0.091	0.092	0.091	0.091	0.091
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 in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. 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.

### *Controlling for Overall Points*

It is possible that points achieved in the Leaving Certificate (based on the best 6 subjects) may affect choice of college major. We have not included points as a control in our main specification as rank in math and English may affect performance in other subjects and, hence, total points. However, in column (1) of Table 5, we add a quartic in points (interacted with a 2017 indicator because the mapping of grades to points changed in 2017) to show that our estimates are robust to including these additional controls.

**Table 5: Robustness Checks – Specification Checks**

VARIABLES	(1) Control for Overall Points	(2) Control for Overall Rank	(3) Interact Grades with School Characteristics	(4) Interact Grades with Mean Achievement	(5) Interact Grades with SD of Achievement	(6) Triple Interact Grades with mean and SD of Achievement
<b>First Preference is STEM</b>						
Math Rank	0.129*** (0.031)	0.149*** (0.034)	0.151*** (0.033)	0.192*** (0.037)	0.129*** (0.033)	0.164*** (0.042)
English Rank	-0.078*** (0.027)	-0.068** (0.028)	-0.082*** (0.028)	-0.089*** (0.033)	-0.085*** (0.030)	-0.087** (0.039)
R-squared	0.189	0.189	0.193	0.188	0.188	0.190
Observations	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
<b>First Preference is Arts and Social Sciences</b>						
Math Rank	-0.140*** (0.028)	-0.099*** (0.031)	-0.125*** (0.030)	-0.145*** (0.033)	-0.186*** (0.029)	-0.161*** (0.039)
English Rank	0.091*** (0.027)	0.112*** (0.028)	0.096*** (0.025)	0.135*** (0.032)	0.086*** (0.031)	0.102** (0.040)
R-squared	0.095	0.095	0.096	0.092	0.092	0.094
Observations	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

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions.

### *Controlling for Overall Rank*

In column (2) of Table 5, we include a control for overall rank as measured by the within school-cohort percentile rank of 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.<sup>20</sup> However, we exclude the control for overall rank from our main specifications as higher rank in English or math may lead to better performance in a range of subjects and, hence, to higher overall rank. If that is the case, overall rank is an intermediate

<sup>20</sup> The correlation between math rank and overall rank is 0.82 and that between English rank and overall rank is 0.75.



variable and it is inappropriate to control for it. We find that adding this variable has very little effect on our 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).<sup>21</sup> By interacting subject grades with these school-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. In column (3) of Table 5, we see the estimates are robust to this change in specification.

---

<sup>21</sup> 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 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.

In column (4), 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.<sup>22</sup> 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 (5), 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 (6), 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.<sup>23</sup>

#### *More detailed Field Categories*

In Appendix Table A4 we show estimates for the 10 main ISCED field of study categories. We see that there are statistically significant positive effects of math rank on listing college programs in Technology and Engineering. There is a negative effect of English rank on listing Science (including mathematics) and Engineering. So, the rank effects on STEM appear in each of the main constituent categories. Similarly, the effects we have found show up for both Social Sciences and for Arts.

---

<sup>22</sup> 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.

<sup>23</sup> 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).

## *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 actually accept a program and enroll in college. 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 STEM enrollment are similar to those for having STEM as a first preference. However, for ASSc enrollment, math rank has a larger negative effect and English rank has a smaller positive 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 more like those for first preference choices for the full sample, suggesting that subject rank may have slightly different effects on enrollment than it does on first-choice program.<sup>24</sup>

---

<sup>24</sup> The only substantial difference is for the effect of math rank on ASSc. The larger effect of math rank on ASSc enrollment compared to listing ASSc as first preference is likely due to a number of factors including the persistent negative effects of math rank on ASSc throughout the list of preferences, the correlation between ranks and overall points given that admissions are solely dependent on overall points, and the higher likelihood for those with higher math rank to turn down offers of ASSc programs.

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

VARIABLES	Effect on Enrollment		Effect on First Preference for Enrollment Sample	
	(1) STEM	(2) Arts & Social	(3) STEM	(4) Arts & Social
Math Rank	0.158*** (0.036)	-0.273*** (0.037)	0.140*** (0.036)	-0.160*** (0.032)
English Rank	-0.057* (0.032)	0.015 (0.033)	-0.088*** (0.030)	0.052* (0.031)
Observations	75,939	75,939	75,939	75,939
R-squared	0.204	0.123	0.202	0.104
Mean Outcome	0.31	0.23	0.31	0.19

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. 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.<sup>25</sup> 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 A6) to our main estimates in Table 2.

## 6. School Subject Choice as a Mechanism

There are many possible reasons why ranks in math and English may impact college major decisions. Many are untestable with our data (such as effects on confidence or influence of parents). However, one important potential mediator is choice of subjects for Leaving Certificate. While English, math, and Irish are compulsory subjects for Leaving Certificate, students also choose 4 or 5 option subjects that they study for the 2-year senior cycle. Delaney and Devereux (2019) show that the subjects chosen for Leaving Certificate have very strong predictive power for the subsequent choice of college major. A priori, it is reasonable to believe that subject choice may be influenced by rank in English and math. For example, students who are highly ranked in math may be more likely to choose STEM subjects as options for Leaving Certificate and, hence, may be more likely to subsequently choose STEM programs in college.

To investigate this issue, we need to assume that ranks are persistent so that ranks in English and math at the end of high school are similar to the equivalent ranks two years earlier when students are choosing subjects for Leaving Certificate. Given that there is very little mobility between schools over the last two years, and that talent in math and English is

---

<sup>25</sup> 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 Tables A5.

unlikely to vary much over time, we believe that it is reasonable to believe that ranks are persistent over this period.

In Table 7, we study the relationship between rank and choosing the four main STEM subjects available in school. We find that higher math rank is associated with a higher probability of choosing physics, chemistry, and applied math for Leaving Certificate and that higher English rank is negatively associated with the likelihood of choosing physics and applied math. We also find a weak negative relationship between math rank and choosing biology.

We found that the effect of subject rank on STEM is larger for boys than girls (see Figure 2) and this may be related to how math and English ranks affect subject choices in school. In the bottom panel of Table 7 we show the effect of rank interacted with gender. Interestingly, we find that rank in math has a larger effect on choosing physics and applied math for boys. Additionally, we find that rank in English has a larger negative effect on choosing physics, biology and applied math for boys. These results are consistent with our previous finding that boys are more influenced by subject rank and provide further support for subject choices being an important mechanism for the effect of subject rank on choosing a STEM major in college.

**Table 7: Effect of Rank on Subject Choices in School**

VARIABLES	(1) Physics	(2) Chemistry	(3) Biology	(4) Applied Math
Math Rank	0.147*** (0.028)	0.114*** (0.030)	-0.067* (0.038)	0.079*** (0.027)
English Rank	-0.082*** (0.019)	-0.012 (0.022)	-0.001 (0.031)	-0.032*** (0.011)
R-squared	0.289	0.215	0.142	0.345
<b>Rank Interacted with Gender</b>				
Math Rank	0.065** (0.032)	0.126*** (0.036)	-0.038 (0.042)	0.058** (0.028)
Math Rank*Male	0.166*** (0.038)	-0.023 (0.037)	-0.068 (0.053)	0.042** (0.021)
English Rank	-0.014 (0.023)	-0.025 (0.028)	0.067* (0.034)	-0.007 (0.013)
English Rank*Male	-0.140*** (0.034)	0.027 (0.032)	-0.131*** (0.046)	-0.049*** (0.019)
R-squared	0.289	0.215	0.143	0.345
Observations	104,116	104,116	104,116	104,116
Mean Outcome	0.158	0.191	0.656	0.046

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions.

Given, these findings, we investigate to what extent subject choices can “explain” the relationship between math and English rank and college major choice. To do this, we augment our base specification from Table 2 with indicator variables for whether the student took each of the 25 most popular subjects for Leaving Certificate. A limitation of our data is that we cannot rule out the possibility that subject choices affect rank in math and English – for example, a student may do better in math because they studied physics. However, even if subject choice affects performance in math or English, this should not affect rank in math or English conditional on grades. The estimates are in the first two columns of Table 8. When we look at math and English rank, we see that the effects remain statistically significant but with smaller coefficients than before – approximately half the size for STEM. Our interpretation is that students with high math rank are more likely to choose STEM-friendly

subjects in high school and this is a major reason for their higher likelihood of choosing STEM programs in college.

**Table 8: Rank in Math and English Controlling for Subjects & Grades**

VARIABLES	(1) STEM	(2) Arts/Soc	(3) STEM	(4) Arts/Soc
Math Rank	0.062** (0.028)	-0.116*** (0.027)	0.067** (0.027)	-0.121*** (0.026)
English Rank	-0.065*** (0.024)	0.049* (0.025)	-0.042* (0.023)	0.036 (0.025)
Subject Fixed Effects	Yes	Yes	Yes	Yes
Grade Fixed Effects	No	No	Yes	Yes
Observations	104,116	104,116	104,116	104,116
R-squared	0.296	0.156	0.314	0.169
Mean Outcome	0.299	0.197	0.299	0.197

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, 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).

In columns (3) and (4) of Table 8, we add further indicators for grades achieved in each of the 25 subjects (interacted with indicators for taking the subjects). These added controls have relatively little impact on the subject rank estimates. Our estimates with these controls can be interpreted as the effect of math and English ranks on college major choice behavior at the end of high school, taking account of absolute achievement and academic interests at that point (as measured by subjects chosen for Leaving Certificate and performance in each of these subjects). As such, they abstract from the effect of subject rank on human capital accumulation and focus on the effect of subject rank on choice behavior. Thus, the findings in Table 8 imply behavioral effects of subject rank that go beyond effects on human capital accumulation.



## 7. 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 Appendix Table A7).<sup>26</sup> 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 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.<sup>27</sup>

Table 9 shows how ranks vary by sex in mixed-sex schools. As expected, girls have higher rank in English (by 9 percentage points), but boys have higher rank in math (by 3 percentage points). The table also shows that there are similar differences in absolute subject ranks (where the ranking is done across all students in a cohort rather than just across students in the same school-cohort).

*Table 9: Average Ranks by Gender in Mixed-sex Schools*

	<u>Within School-Cohort Rank</u>		<u>Absolute Rank</u>	
	Female	Male	Female	Male
Math Rank	0.558	0.584	0.457	0.489
English Rank	0.617	0.531	0.503	0.411
N	27,431	29,876	27,431	29,876

<sup>26</sup> In Appendix Table A7, 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 grades in English and math. The gender gap in mixed-sex schools is 25.3 percentage points without controls and falls to 19.8 percentage points with controls for absolute performance in math and English.

<sup>27</sup> By definition, average ranks are the same for boys and girls in same-sex schools.

### *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 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 A8) and found that overall school-cohort rank is more important than own-gender rank so we don't believe that this is an important consideration. Another possibility is that the presence of members of the opposite sex affects behavior.

In Table 10, we show that there are no statistically significant differences in the effects of math or English rank between mixed-sex (columns (1) and (2)) and same-sex schools (columns (5) and (6)). When we allow for gender interactions, we find that the subject rank effects on STEM are larger for boys in mixed-sex schools but there is no evidence for a gender difference in same-sex schools. We find no evidence for gender differences in rank effects on ASSc in either type of school.

**Table 10: Rank in Math and English by School Gender-mix**

VARIABLES	Mixed-Sex		Mixed-Sex		Same-Sex		Same-Sex	
	(1) STEM	(2) Arts/Soc	(3) STEM	(4) Arts/Soc	(5) STEM	(6) Arts/Soc	(7) STEM	(8) Arts/Soc
Math Rank	0.117*** (0.044)	-0.124*** (0.037)	-0.021 (0.049)	-0.130*** (0.045)	0.127*** (0.043)	-0.180*** (0.044)	0.154*** (0.053)	-0.200*** (0.060)
English Rank	-0.140*** (0.036)	0.111*** (0.032)	-0.090** (0.044)	0.106*** (0.040)	-0.055 (0.040)	0.074* (0.044)	-0.069 (0.053)	0.093 (0.065)
Math Rank*Male			0.280*** (0.054)	0.014 (0.044)			-0.057 (0.087)	0.040 (0.088)
English Rank*Male			-0.111** (0.056)	0.008 (0.041)			0.028 (0.079)	-0.039 (0.089)
Observations	57,307	57,307	57,307	57,307	46,809	46,809	46,809	46,809
R-squared	0.192	0.101	0.192	0.101	0.180	0.084	0.180	0.084
Mean Outcome	0.315	0.193	0.315	0.193	0.279	0.201	0.279	0.201

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions.

#### *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. Using the coefficients for mixed-sex schools from Column (1) of Table 10, 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 11 shows the differences in the gender gap in STEM explained by differential gender ranks is 1.5 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 6% (1.5/25) of the gender gap

in preferences for STEM in college.<sup>28</sup> If we focus on the 20 percentage point unexplained gender gap in mixed-sex schools after controlling for absolute performance in math and English (see Appendix Table A7), the rank variables can account for 8% of the unexplained gap.<sup>29</sup>

**Table 11: 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
Baseline Rank Coefficients	1.5 (0.3)	5.9%	7.6%	16.1%
Male Rank Coefficients	2.4 (0.4)	9.5%	12.1%	25.8%
Female Rank Coefficients	0.7 (0.4)	2.8%	3.5%	7.5%

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 math and English grades reduced this gap to 19.8pp and 11.2pp, respectively.

We saw, in Table 10, that the effect of subject rank on STEM is greater for males. We can do separate calculations using the male rank coefficients and using the female rank coefficients. We get 0.7 percentage points using the female coefficients and 2.4 percentage points using the male coefficients. These provide a lower and upper bound on the effect of gender rank in mixed-sex schools on the gender gap in STEM. We conclude that gender differences in gap in first math and English can explain between 0.7 and 2.4 percentage points of the gender gap in STEM in mixed-sex schools.

It is natural to compare the estimates to the amount that can be explained by absolute performance in math and English in mixed-sex schools. Appendix Table A9 shows the effect of absolute performance ranks in English and math for students in mixed-sex schools. Using these coefficients and the gender differences in absolute ranks in Table 9, we obtain an effect of differential absolute performance ranks of 4.2 percentage points. Thus, the explanatory

<sup>28</sup> Murphy and Weinhardt (2018) 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.

<sup>29</sup> 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.

power of school-cohort rank in English and math for the gender gap in STEM is about 36% of that of absolute rank in English and math.

### *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 16% (1.5/9.3) of the difference between the STEM gender gap in mixed-sex schools versus same-sex schools. Thus, it appears that within school-cohort ranks in English and math have substantial effects on the larger gender gap in choice of STEM as a college major in mixed-sex schools compared to same-sex schools.

## **8. Conclusions**

We draw three main conclusions from our analysis. First, within school-cohort percentile ranks in English and math are 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. 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 magnitudes of the effects are substantial -- they are about 25-44% as large as the effects of absolute performance in English and math. They also can explain about 6% of the gender gap in the choice of STEM as a college major in mixed-sex schools and 16% of the difference in the gender gap between mixed-sex and same-sex schools. Notably, these effects occur even though within-school rank plays no role whatsoever in college admissions decisions.

While we are limited in our ability to study mechanisms, we identify subject choice in school as an important mediator – students who rank high in math are more likely to choose STEM subjects in school and this can partly explain their subsequent higher likelihood of choosing STEM for college. However, even when we control for the subjects taken and grades obtained in each subject in the Leaving Certificate, we still find that math and English rank affect college major choice. This finding implies 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 substantial 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 ability in math and English. 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 high school subjects and 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, Joseph G., Peter Arcidiacono, and Arnaud Maurel. 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.
- Aucejo, Esteban M. and Jonathan James. 2019. ‘The Path to College Education: The Role of Math and Verbal Skills’. Working Paper.
- Azmat, G., Bagues, M., Cabrales, A. & Iriberry, N. (2015). “What You Know Can’t Hurt You, For Long: A Field Experiment on Relative Performance Feedback” Mimeo, Queen Mary University.
- Azmat, G. & Iriberry, 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, Matteo, & Frisancho, Veronica. 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.
- Bursztyn, L. and Jensen, R. (2015). How Does Peer Pressure Affect Educational Investments? *The Quarterly Journal of Economics*, 130(3):1329–1367.
- Buser, Thomas, Noemi Peter, and Stefan 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.

Denning, J. T., Murphy, R., and Weinhardt, F. (2018). “Class Rank and Long-run Outcomes”, IZA Discussion Paper.

Doris, Aedin, Donal O'Neill, and Olive 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” Mimeo.

Kinsler, Josh and Ronni Pavan, “Parental Beliefs and Investment in Children: The Distortionary Impact of Schools,” Technical Report, Unpublished Manuscript 2016.

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

Murphy, Richard. and Weinhardt, Felix. (2018) “Top of the Class: The importance of ordinal rank”. NBER Working paper. August 2018.

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

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



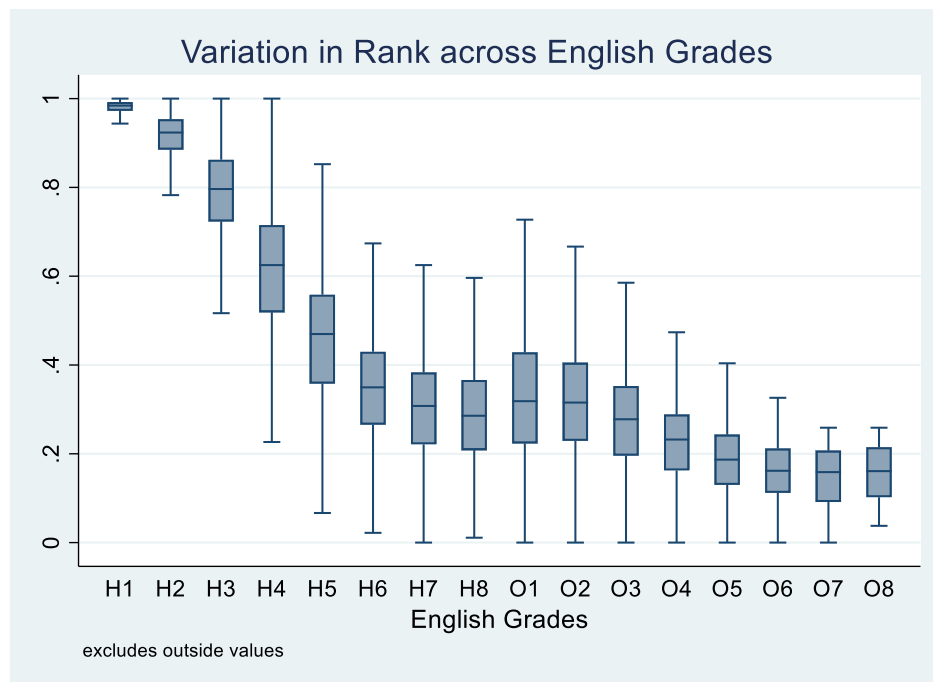
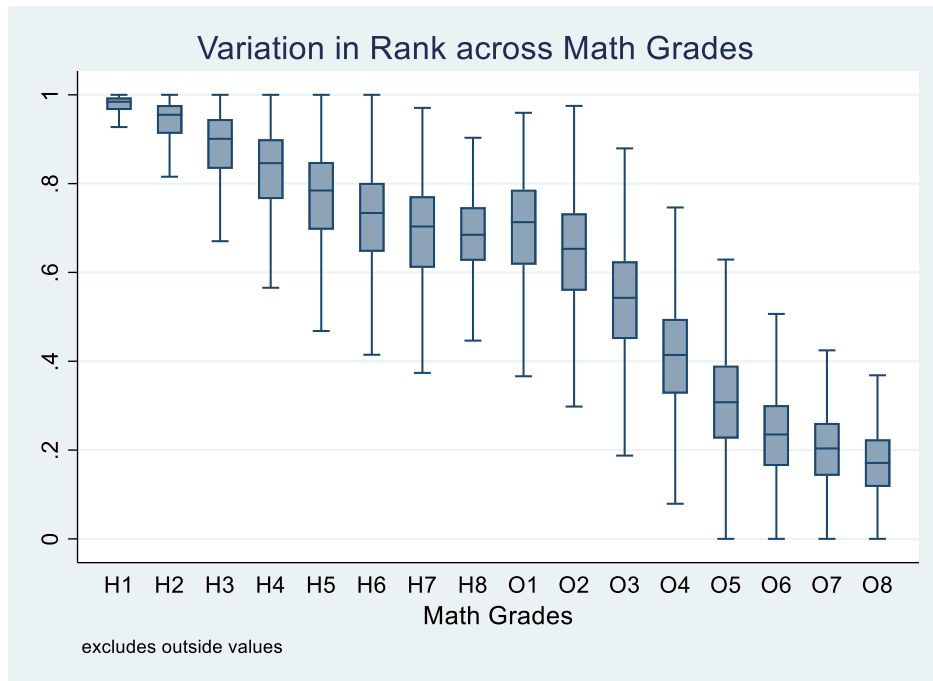
Stinebrickner, Todd, & Stinebrickner, Ralph. 2012. Learning about Academic Ability and the College Drop-out Decision. *Journal of Labor Economics*, 30(4), 707-748.

Stinebrickner, Todd, & 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.

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

Zafar, Basit. 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. 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 25<sup>th</sup> and 75<sup>th</sup> 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*

<b>Grade</b>	<b>Marks (%)</b>	<b>Points</b>	<b>Points (Math)</b>
<i>Higher Level</i>			
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
<i>Lower Level</i>			
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

<b>Grade</b>	<b>Marks (%)</b>	<b>Points</b>	<b>Points (Math)</b>
<i>Higher Level</i>			
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
<i>Lower Level</i>			
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*

	<i>Standard Deviation</i>
Variation in math rank no controls	0.258
Variation in math rank controlling for age, gender, math and English grades (interacted with gender), and school-cohort fixed effects	0.047
Variation in math rank controlling for age, gender, math and English grades (interacted with gender), school-cohort fixed effects and subjects and grades	0.047
Variation in English Rank no controls	0.257
Variation in English rank controlling for age, gender, math and English grades (interacted with gender), and school-cohort fixed effects	0.055
Variation in English rank controlling for age, gender, math and English grades (interacted with gender), school-cohort fixed effects and subjects and grades	0.054

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

*Appendix Table A3: Absolute Rank in Math and English and Choice of Field*

VARIABLES	(1) STEM	(2) BAL	(3) Arts/Soc	(4) Other
Absolute Math Rank	0.532*** (0.007)	-0.038*** (0.006)	-0.338*** (0.007)	-0.156*** (0.007)
Absolute English Rank	-0.250*** (0.008)	0.090*** (0.007)	0.308*** (0.007)	-0.148*** (0.007)
Observations	104,116	104,116	104,116	104,116
R-squared	0.151	0.045	0.083	0.105
Mean Outcome	0.299	0.208	0.197	0.296

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age and gender dummies and school-cohort fixed effects included in all regressions. Absolute rank in a subject is the percentile rank of the student in that subject in the entire cohort, not just in his/her school. BAL refers to Business Administration and Law.

*Appendix Table A4: Effect of Rank on First Preference Field of Study (ISCED 2-digit classification)*

VARIABLES	(1) Education	(2) Social	(3) Arts	(4) BAL	(5) Sciences	(6) Technology	(7) Engineering	(8) Agriculture	(9) Health	(10) Services
Math Rank	0.024 (0.019)	-0.065*** (0.018)	-0.081*** (0.024)	0.041 (0.027)	0.034 (0.027)	0.036** (0.016)	0.062*** (0.023)	-0.014* (0.008)	-0.036 (0.022)	0.000 (0.014)
English Rank	-0.036** (0.016)	0.010 (0.016)	0.066*** (0.022)	0.045* (0.024)	-0.050*** (0.019)	0.000 (0.014)	-0.038* (0.020)	-0.011 (0.009)	0.023 (0.019)	-0.011 (0.014)
Observations	104,116	104,116	104,116	104,116	104,116	104,116	104,116	104,116	104,116	104,116
R-squared	0.056	0.042	0.071	0.053	0.133	0.064	0.130	0.050	0.099	0.072
Mean Outcome	0.084	0.063	0.134	0.208	0.142	0.058	0.098	0.017	0.131	0.051

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. BAL refers to Business Administration and Law.

*Appendix Table A5: 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 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 A6: Effect of Rank on Field of Study (Weighting by the Inverse Probability of being in the Sample)**

VARIABLES	(1) STEM	(2) Arts/Soc
Math Rank	0.138*** (0.035)	-0.144*** (0.030)
English Rank	-0.084** (0.033)	0.085*** (0.030)
Observations	104,116	104,116
R-squared	0.187	0.093
Mean Outcome	0.30	0.20

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. Regressions are estimated using inverse probability weighting.

**Appendix Table A7: Effect of Female on Choosing STEM in College by School Type**

VARIABLES	Overall		Mixed-Sex		Same-Sex	
	(1) STEM	(2) STEM	(3) STEM	(4) STEM	(5) STEM	(6) STEM
Female	-0.213*** (0.005)	-0.162*** (0.006)	-0.253*** (0.006)	-0.198*** (0.006)	-0.160*** (0.008)	-0.112*** (0.010)
Observations	104,116	104,116	57,307	57,307	46,809	46,809
R-squared	0.055	0.153	0.074	0.163	0.033	0.146
Math and English Grades	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 in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age and school-cohort fixed effects included in all regressions.



*Appendix Table A8: The Effect of same-gender rank in mixed-sex schools*

VARIABLES	(1) STEM	(2) Arts & Social
Same-gender Math Rank	0.044 (0.041)	0.023 (0.039)
Same-gender English Rank	0.010 (0.040)	-0.009 (0.035)
Math Rank	0.073 (0.061)	-0.148*** (0.054)
English Rank	-0.155*** (0.054)	0.115** (0.051)
Observations	56,118	56,118
R-squared	0.192	0.101
Mean Outcome	0.315	0.193

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age, indicator variables for grades in math and English interacted with gender, and school-cohort fixed effects included in all regressions. Same-gender rank is rank calculated just using persons in the school-cohort who have the same gender.

*Appendix Table A9: Absolute Rank in Math and English and Choice of Field in Mixed-Sex Schools*

VARIABLES	(1) STEM	(2) Arts/Soc
Absolute Math Rank	0.520*** (0.010)	-0.030*** (0.008)
Absolute English Rank	-0.276*** (0.008)	0.102*** (0.008)
Observations	57,307	57,307
R-squared	0.160	0.038
Mean Outcome	0.315	0.193

Robust standard errors clustered by school in parentheses. \*\*\* p<0.01; \*\* p<0.05 \* p<0.10. Age and gender dummies and school-cohort fixed effects included in all regressions. Absolute rank in a subject is the percentile rank of the student in that subject in the entire cohort, not just in his/her school.