Gender Differences in College Major Choice: Preferences versus Performance

Ş. Pelin Akyol [*] Kala Krishna[†] Sergey Lychagin [‡]

April 2, 2017

Abstract

When admissions are performance based there seem to be no obvious gender biases. Despite this, the college major choice decisions of students vary considerably by gender. This could be because of performance differences by gender, and/or differences in preferences across majors. We documents the differences in major choice in college in Turkey and argue that differences in exam performance especially performance in math and science, are important in driving the differences in college major choices. Our results suggest that policies that bring the performance of women to the that of men would reduce the gender gap in engineering and technical science majors by half.

[*]Bilkent University, e-mail:pelina@bilkent.edu.tr
[†]The Pennsylvania State University, NBER, IGC and CES-IFO, e-mail:kmk4@psu.edu
[‡]Central European University

1 Introduction

Gender differences in college completion have been falling in countries all over the world. In some countries, women are even more likely to graduate from college than men! (Gemici and Wiswall [2014]) However, this should not result in complacency regarding gender equality as women tend to major in subjects with lower returns but possibly more friendly to women. These differences in major choice are large and significantly impact their earning potential. While males are more likely to major in engineering, females are more likely to major in education and the humanities. (Turner and Bowen [1999] , Zafar [Forthcoming]).

These differences in college major choice have important implications for labor market outcomes. Arcidiacono [2004] find large earning differences across majors even after controlling for selection. Loury [1997] and Loury and Garman [1995] also document large earning differences across majors. It is important to understand the factors behind differences in college major choice if we are to develop policies targeted to reducing the gap in earnings across gender. Differences in preferences as well as differences in pre-college performance both contribute to the gender differences in college major choice.

Turner and Bowen [1999] find that for the US, SAT score differences account for 45% of the gender gap in terms of enrollment in math-physical science fields and 32% of the gender gap in engineering under the assumption that the college and major are chosen together. This may be different in Turkey, as the educational institutions as well as the role of women in society (which may impact their preferences) may differ in Turkey and the US. The US system is much more flexible than the Turkish one: students choose their majors after they start college and can switch until quite late in their studies. The time of choosing a major, the number of required and elective courses taken before the major choice decision, vary across settings and may affect program choice decisions. (Stinebrickner and Stinebrickner [2014]). In addition, norms regarding work and family may differ in the two countries which may well impact preferences across majors.

In Turkey, as in most countries, there is a large gender gap especially in engineering and technical science programs. As admission to university in the Turkish system is based only on performance (high school GPA and university entrance exam scores) and stated preferences, it provides a simpler setting in which to study the role of these two factors in driving gender differences across majors.[1] In contrast, admission decisions in the US are more opaque and major choices are made after admission, not before making such an analysis more difficult.

In this paper, we first investigate how female and male students differ from each other in terms of their performance on the college entrance exam and their preferences submitted. Then, we attempt to disentangle the effects of differences in pre-collegiate preparation and preferences of students on gender differences in college major choice by using a representative sample of 2002 University entrance exam takers in Turkey. To understand the effect of score differences on different portion of the exam on college majors, we use a method suggested by Fairlie [2006]. This method allows us to identify the contribution of performance differences on gender gap in each field of study.

Our results suggest that in addition to students' scores, which are the most important factor, their gender and income level also affect their major choice decision. Performance differences in math and science portion of the exam account for 53.1% of the 9.2% gender gap in engineering and technical science programs. This suggests that policies that help improve female students' achievement in math and science fields can increase the number of female students in engineering and technical science programs.

Score difference by gender come from two sources: those that arise from differences in preparation in High School and those that arise from differences in retaking of the university entrance exam by gender. We find that controlling for observables, males are much more likely to retake the exam, —, which contributes to their higher scores in the entrance exam. Among first time takers, the gender gap is only – while overall it is –. Women may be discouraged form retaking if they are seen as less likely to gain from better placements as they are less likely to use their education. This constraint is more likely to apply to students coming from less educated families who live in less developed areas and the data is consistent with this intuition. Different retaking behavior may also arise due to different attitudes toward risk and competition. Women seem to be more risk averse and less willing to compete (Niederle and Vesterlund [2011], Niederle and Vesterlund [2007], Eckel and Grossman [2008]).

[1] After students learn their scores they submit a preference list and are allocated to college programs with higher scores given priority.

Retaking is a risky business, since the last score obtained is used.

The rest if the paper is organized as follows. The next section describes the data, and gives the necessary background information regarding the college entrance system in Turkey. We present patterns found in the data in section 3. In section 4, we investigate the effect of students' score, income, gender, and family education level on their major choice decision. Section 5 investigates whether score differences account for the gender differences in college majors. Section 6 looks at the factors affecting students' decision to retake the exam. Section 7 concludes.

2 Data and Background

In Turkey, the university entrance system is highly centralized. Students who want to attend a university program need to take an annual, nation-wide, central university entrance exam. This exam is conducted by the Student Selection and Placement Center (ÖSYM) every year. High school seniors and high school graduates can take the exam. Students are free to repeat the exam, but the score obtained in a year can be used only in that year. A weighted average of the students' test scores and their high school grade point average (GPA) is the only determinant of college admission. The exam includes four tests, Turkish, social science, math, and science. Students' scores are calculated as a weighted average of their standardized raw scores in each test. For each student three different scores, Quantitative (OSS-SAY), Turkish-Math (OSS-EA) and Social (OSS-SOZ), are constructed by using different weights for each test.

After students start high school, they choose one of the four - Science, Turkish-Math, Social Science, or Language - tracks. In each track, students study a different curriculum. The track they have chosen matters when their allocation scores are constructed as there is an attempt to keep people in their chosen track. If a student applies for a program that is outside the list of college programs that are open to student's track, he is punished by attaching a lower weight to his high school performance.

Students' allocation scores (Y-ÖSS) are calculated by using their high school performance and scores in each portion of the exam as follows

$$Y\_\ddot{O}SS\_X_i = \ddot{O}SS\_X_i + \eta AOBP\_X_i$$

where X$\in${SAY,SÖZ,EA,DIL}, and η is a constant that changes according to the student's track, preferred department and whether the student was placed (accepted) into a regular program in the previous year or not. ÖSYM publishes the lists of departments open to students according to their tracks.

After the exam, students are informed about their raw scores, weighted scores and allocation scores. Students who get at least 120 in a score type, are eligible to submit preference for all 2-year and 4-year college programs that admit students based on that type of score. Students whose scores are between 105 and 120 are only allowed to submit preference for 2-year college programs. Students can submit up to 24 preferences, and at most 18 of these can be for 4-year or 2-year programs. Students are well informed as they are provided with information regarding each program's cutoff admission score in the past years, the available number of seats, tuition, and the type of the score the program requires. The system is relatively stable in this period so that it is not unreasonable to think of students having a fairly good idea what their feasible set is.

As different departments require different type of scores, and each student has three different scores, the Multi-Category Serial Dictatorship algorithm is used to allocate students to college programs.[2]

The data used in this study comes from multiple sources. Our main source of data is the institutional data of the 2002 university entrance exam takers and 2002 University entrance exam candidate survey, which was filled by all students while they are making their application for the exam. This data set includes students' raw test scores in each test, weighted test scores (ÖSS_X), high school, track, high school GPA, gender, if they submit a preference list, their ranked preference list, the college they are assigned, if any. The survey data includes information on students' family background information.

We received a random sample of around 40,000 students from each track (Social Science, Turkish-Math, Science). In this study, we focus on the science track students. These are

[2] In Turkey college and program choice is made simultaneously.

the students who can choose STEM majors without punishment and are relatively homogeneous. We observe 39847 students from science track. In our administrative data, although we observe students' high school GPA, we do not observe students' weighted GPA (AOBP), which normalizes GPA across schools and is used to construct student's allocation scores. We develop a way to construct AOBP scores that is explained in the Appendix B. After we drop students whose AOBP scores cannot be calculated, we end up with 37270 observations. However, we observe the preference list of only first and second time taker students. Therefore, in this study we will focus on these two groups of students. Summary statistics for the data set are presented in Table 5.

The second source of data is the booklet published by ÖSYM that includes minimum cutoff scores, maximum admission scores and available number of seats in all college programs for the years 2000, 2001, and 2002. This data also includes tuition cost of each department, amount of the scholarship, if provided. [3] In addition, we collected the distance between all cities from the General Directorate of Highways. Table 6 and 7 present the summary statistics of the college programs that are open to science track students separately for programs that admit students with the quantitative score (Y-OSS-SAY) and the Turkish-Math score (Y-OSS-EA).

3 Data Patterns

In Turkey, around 1.5 million students take the University Entrance Exam every year, and only one third of these get chance to be placed in a university program. Students face fierce competition. On the other hand, there is no limit on retaking the exam. Students, who are not placed in a college program, can take the exam next year without penalty. After students learn their scores, they make a decision about which programs/options to choose including the option of retaking or quitting. Students are allocated to their preferences centrally by using the multi-category serial dictatorship allocation algorithm. In this allocation process, the allocation scores of the students and their preferences are the only criteria. Students know their allocation scores and the previous years' cutoff scores while submitting their

[3] Tuition cost in public universities does not vary across universities, but it varies according to the major.

preference lists. So each student has an idea about his chance of being admitted into a college program.

As in many other countries, in Turkey also there are gender differences in college majors. Figure 1 presents percentage of female and male students in each major according to the major students placed and their first ranked major in their preference list. As the figure shows, 74.6% of students who are placed in an Engineering and Technical Science program are males, on the other hand, 43.9% of students majoring in education are males. The largest gender gap exists in the medical services programs which include nursing, midwifery, and health visitor. This pattern is the same for students' first ranked majors in their preference list. The pattern could be the result of either systematic differences in the preferences of female and male students or differences in their score distributions.

Figure 2 shows the percentage of female and male students, placed into an education program, in the sub-fields of education programs: Technical Education, Math-Science Education, and Primary School Education. This figure is particularly important because public sector jobs are the main source of employment opportunity for the graduates of education programs. So, they work under the same contract and earn the same. However, figure 2 shows that even within education programs, technical education programs are dominated by male students while primary school education programs are dominated by female students.

Figure 3 shows the distribution of allocation scores of female and male students separately for the first and second time taker students. We also present score distributions of all students and students who are eligible to submit preference for 2-years and 4-years college programs (OSS score above 120) separately. In Turkey, almost all high school seniors take the university entrance exam, therefore; the group of first time takers is free of selection.[4] As the figure shows, in all groups there are more male students in the upper tail of the distribution, and the gender score gap gets more prominent in the group of second time takers.

Figure 4 presents the raw score distributions of females and males, and first time and second time taker students in each test. The first thing to note is male students perform better than female students in the math and science part of the exam. However, female students' better performance in Turkish part of the exam and their better high school per-

[4]There may be selection until students reach high school level.

formance (see Figure 5) make up for it and reduce the score gap in the allocation score, which is a weighted sum of the performance on each portion of the exam and the high school GPA. So, the decomposition of allocation score (Y-OSS-X) gives larger differences in terms of female and male students' performances on the exam.

Figure 6 shows the share of female and male students' major preferences relative to students' allocation score (YOSS-SAY) bins and the rank of the preference.[5] In almost all score bins, male students are more likely to submit preference for engineering and technical science programs. However, the preferences of female students depend more on their scores. While female students in the top bins more likely to apply to engineering and technical science programs, female students in the middle of the distribution are more likely to submit preference for education departments. In addition, female students are more likely to apply to medicine departments. Low performing female students submit preferences for medical service majors, however even with lower ranks; males do not put these majors into their preference list.

We also investigate whether there is any difference in the selectivity of the departments female and male students submit. Figure 7 presents the mean of the lowest, highest and the median selectivity (cutoff) of the departments in the students preference lists. The first column of figure 7 shows average selectivity of departments female and male students submit in their preference lists. Female students on average submit preferences for less selective departments, however this pattern is driven by their scores. When we look into differences between the selectivity of the department and students' score, the difference between female and male students gets less significant. However, female students on average list less selective science track departments given to their scores. On the other hand, males submit less selective Turkish-Math track departments conditional on submitting a preference for a program in this track. In figure 7 we dropped distance education programs, since they are open to all students who are eligible to apply a program.

In this section, we showed that there are considerable gender differences in college major choice. There are two main reasons that can explain this difference: Differences in score distributions or differences in preferences. We show that female and male students' score

[5] We construct score bins of width 5 starting from 120.

distributions are significantly different from each other. In addition, female and male students have systematic differences in their preference of majors. In the next section, we will investigate the underlying factors that effect students major preferences including retaking decision.

4 Major Choice

In this section, we investigate how student characteristics affect their program choice decisions. We define the latent variable

$$y_{ij}^* = \gamma X_i + u_i$$

where u_i is standard normally distributed error term, and

$$y_{ij} = \begin{Bmatrix} 1 & \text{if } y_{ij}^* > 0 \\ 0 & \text{if } y_{ij}^* \leq 0. \end{Bmatrix}$$

y_{ij} is an indicator function that shows whether student i has chosen the major j. So, student i chooses major j with probability

$$L(y_{ij}) = \Pr(y_{ij} = 1) = \Pr(\gamma X_i + u_i > 0) = 1 - \Phi(\gamma X_i) = \Phi(-\gamma X_i)$$

Then, we can write the likelihood function of getting the observed choices in the data for a given γ as follows

$$L_j(y|\gamma) = \prod_i L(y_{ij})$$

Coefficient estimates are obtained by maximizing logarithm of the likelihood function

$$\hat{\gamma} = \arg\max_{\gamma} \log(L_j(y|\gamma))$$

We estimated the model for each major separately to understand the effect of students' gender, score, family education level, and income. Table 1 and Table 2 show estimation results for placement programs and first choice programs, respectively. We specify students' test score as a categorical variable to understand the demand coming from different score groups for each major.

The results suggest that students' gender, income and score are the most important factors in their college program choice decisions.

As students' score increases, they are more likely to submit engineering and medicine programs as their first choice. However, as students' score increases, their likelihood of choosing medical service, science, distance education, and agriculture programs decreases. Students with score above 195 are less likely to submit preference for an education program as first choice, however they are more likely to placed into an education program. This suggests that students like engineering or medicine programs more, but they put education programs in latter ranks in their preference list. So that, as it is shown in Table 1 they are more likely to placed into these programs. Essentially, high performing students are competing for engineering and medicine departments.

The income level of students is another important factor in college major choice decision. For the students in the highest income group, engineering and management and economics are the most attractive departments. On the other hand, education and medical service, even medicine programs are more attractive to lowest income students. As discussed in Caner and Okten [2010] graduates of these programs are mostly hired by public sector, and these majors are perceived as less risky in terms of earnings and employment opportunities. Lower income students may prefer to take the less risky route and choose these departments because of their labor market outcomes. Finally, male students are more likely to be placed into technical science and engineering majors, while other majors are more attractive to female students.

Table 1: Placement Department

	Tech. & Eng.	Education	Medicine	Mang. & Econ	Med. Service	Science	Distance Edc.	Agriculture
Male	0.758***	-0.597***	-0.161**	0.000467	-1.355***	-0.125**	0.260*	0.0899
	(0.0395)	(0.0473)	(0.0539)	(0.0494)	(0.11)	(0.0483)	(0.132)	(0.11)
Income								
<250 TL (base)								
$[250TL, 500TL]$	0.120*	-0.241***	0.105	0.275***	-0.196*	0.00365	-0.0949	0.0749
	(0.0561)	(0.0582)	(0.0778)	(0.0809)	(0.0937)	(0.0674)	(0.176)	(0.161)
>500TL	0.432***	-0.769***	-0.0929	0.477***	-0.637***	0.0624	-0.0073	-0.0261
	(0.0625)	(0.0745)	(0.0898)	(0.0859)	(0.127)	(0.0764)	(0.194)	(0.179)
Y-OSS-SAY								
<165 (base)								
$[165, 180)$	0.139	0.817***	0.522**	0.0794	-0.176	0.152	-1.327***	-0.928***
	(0.0828)	(0.118)	(0.186)	(0.0907)	(0.106)	(0.0829)	(0.162)	(0.125)
$[180, 195)$	0.501***	1.355***	-0.106	-0.12	-1.105***	-0.104	-1.765***	-1.405***
	(0.0773)	(0.114)	(0.2)	(0.0878)	(0.126)	(0.0806)	(0.205)	(0.15)
$\geq$ 195	0.699***	0.661***	1.668***	-0.417***	-1.152***	-0.728***		
	(0.0748)	(0.114)	(0.173)	(0.0869)	(0.125)	(0.0854)		
Mother's Education Level								
Primary or less (base)								
Middle/High school	0.220***	-0.497***	-0.0982	0.281***	-0.238*	0.130*	0.0185	0.346*
	(0.0525)	(0.0618)	(0.0726)	(0.0686)	(0.102)	(0.0656)	(0.176)	(0.155)
2-year higher education	0.0444	-0.252*	-0.00746	0.403***	-0.38	0.0365		0.398
	(0.0865)	(0.107)	(0.115)	(0.105)	(0.24)	(0.117)		(0.28)
College/Master/Phd	0.251***	-0.748***	-0.201	0.309**	-0.292	0.207*	-0.0616	0.640**
	(0.0751)	(0.115)	(0.102)	(0.0941)	(0.201)	(0.1)	(0.268)	(0.226)
Missing	0.433**	0.0242	-0.460*	0.25	-0.582*	-0.172	-0.572	0.158
	(0.148)	(0.16)	(0.219)	(0.202)	(0.259)	(0.177)	(0.413)	(0.338)
Father's Education Level								
Primary or less (base)								
Middle/High school	0.0942	0.0666	-0.0258	-0.0188	-0.123	-0.0915	-0.193	0.0543
	(0.0591)	(0.0636)	(0.0838)	(0.0807)	(0.102)	(0.0704)	(0.185)	(0.171)
2-year higher education	-0.1	0.290**	0.0255	0.022	-0.265	-0.103	-0.0229	-0.057
	(0.0879)	(0.098)	(0.12)	(0.113)	(0.193)	(0.109)	(0.304)	(0.268)
College/Master/Phd	0.147*	-0.0218	0.123	0.012	-0.159	-0.322***	0.144	-0.0491
	(0.069)	(0.0819)	(0.0948)	(0.0906)	(0.141)	(0.0881)	(0.224)	(0.21)
Missing	-0.157	-0.0898	0.241	-0.0575	0.227	-0.0225	0.315	0.286
	(0.138)	(0.148)	(0.198)	(0.19)	(0.223)	(0.158)	(0.319)	(0.301)
Constant	-1.622***	-0.979***	-2.190***	-1.569***	-0.171	-0.835***	-1.209***	-1.495***
	(0.0884)	(0.114)	(0.183)	(0.107)	(0.112)	(0.0894)	(0.165)	(0.164)
N	4971	4971	4971	4971	4971	4971	2699	2867

Table 2: First Choice Department

	Tech. & Eng.	Education	Medicine	Mang. & Econ	Med. Service	Science	Distance Edc.	Agriculture
Male	0.819***	-0.493***	-0.273***	0.0574	-0.982***	-0.129**	0.22	0.0656
	(0.0366)	(0.0411)	(0.0388)	(0.0545)	(0.0832)	(0.0486)	(0.136)	(0.131)
Income								
<250 TL (base)								
$[250TL, 500TL]$	0.183***	-0.245***	-0.0431	0.206*	-0.0258	0.146*	0.019	0.237
	(0.0516)	(0.0516)	(0.0547)	(0.0916)	(0.0855)	(0.0698)	(0.187)	(0.195)
>500TL	0.483***	-0.705***	-0.182**	0.537***	-0.482***	0.235**	0.239	0.217
	(0.0576)	(0.0647)	(0.0629)	(0.095)	(0.114)	(0.0781)	(0.199)	(0.214)
Y-OSS-SAY								
<165 (base)								
[165, 180)	0.141*	0.551***	-0.0491	0.0698	-0.251**	-0.367***	-1.183***	-0.442**
	(0.0664)	(0.0669)	(0.0862)	(0.0898)	(0.0855)	(0.0686)	(0.191)	(0.142)
[180, 195)	0.323***	0.496***	0.533***	-0.141	-0.677***	-0.689***	-1.519***	-1.055***
	(0.0621)	(0.0645)	(0.0761)	(0.0884)	(0.0932)	(0.0692)	(0.236)	(0.204)
$\geq$ 195	0.477***	-0.553***	1.257***	-0.260**	-1.635***	-1.037***		
	(0.0602)	(0.0708)	(0.0732)	(0.087)	(0.168)	(0.074)		
Mother's Education Level								
Primary or less (base)								
Middle/High school	0.372***	-0.501***	-0.129*	0.273***	0.0142	0.111	0.0213	0.400*
	(0.0479)	(0.054)	(0.052)	(0.0767)	(0.0895)	(0.0659)	(0.178)	(0.178)
2-year higher education	0.165*	-0.247*	-0.0413	0.289*	-0.156	0.143		0.344
	(0.0789)	(0.0965)	(0.0839)	(0.116)	(0.205)	(0.114)		(0.334)
College/Master/Phd	0.413***	-0.668***	-0.229**	0.236*	0.0571	0.243*	-0.196	0.461
	(0.069)	(0.0993)	(0.0758)	(0.104)	(0.174)	(0.0994)	(0.278)	(0.272)
Missing	0.264	0.0145	-0.194	0.145	-0.307	0.131	-0.527	-0.0448
	(0.135)	(0.141)	(0.152)	(0.219)	(0.237)	(0.182)	(0.451)	(0.388)
Father's Education Level								
Primary or less (base)								
Middle/High school	0.0394	0.00423	0.153**	-0.0449	-0.157	-0.0858	-0.264	-0.183
	(0.0544)	(0.0558)	(0.0591)	(0.0903)	(0.0928)	(0.0716)	(0.195)	(0.196)
2-year higher education	-0.167*	0.15	0.317***	-0.146	-0.163	-0.211	-0.0389	-0.226
	(0.0801)	(0.0867)	(0.085)	(0.129)	(0.155)	(0.111)	(0.307)	(0.299)
College/Master/Phd	0.0712	-0.0927	0.251***	-0.0269	-0.339**	-0.301***	0.0953	-0.313
	(0.0632)	(0.0719)	(0.0691)	(0.1)	(0.129)	(0.089)	(0.225)	(0.239)
Missing	0.006	-0.147	0.143	0.0303	0.0257	-0.139	0.218	0.321
	(0.125)	(0.129)	(0.139)	(0.206)	(0.205)	(0.169)	(0.358)	(0.345)
Constant	-1.658***	-0.143*	-1.301***	-1.961***	-0.560***	-0.826***	-1.716***	-2.188***
	(0.0734)	(0.0687)	(0.0825)	(0.111)	(0.0907)	(0.0776)	(0.172)	(0.199)
N	6044	6044	6044	6044	6044	6044	3563	3789

Given that income is an important predictor of college program choice, it is important to note that the income distribution of female and male students are very similar (see figure **??**). If there is any difference, it is in favor of female students. So, differences in major choice cannot be explained by differences in income distributions. In the next section, we will investigate whether differences in the score distributions between male and female students are enough to explain gender differences in college program choice.

5 Do Score Difference Explain the Gender Gap in Major Choice?

In this section, we investigate whether the differences in the score distribution of females and males explain the gender differences in major choice. As shown in the previous section, males outperform females in math and science fields while females perform better in Turkish. In addition, males are more likely to be placed in an engineering program. So, it is natural to wonder whether the performance difference between males and female is enough to explain the difference in college major choice. In the next section, by using an extended version of Oaxaca-Blinder decomposition, we identify the effect of students' exam performance on his college program choice decision.

5.1 The Model

We assume that student i's utility from program j is given by

$$U_{ij} = \beta_j X_i + \varepsilon_{ij}$$

where X_i is the characteristics of the student i, β_j is the weight attached to these characteristics in program j., and ε_{ij} is the unobserved preference shock drawn from a Type I extreme value distribution. Under these assumptions, student i's probability of choosing major j is given as

$$P_{ij} = F(X_i, \beta_j) = \frac{\exp(\beta_j X_i))}{\sum\limits_{k \in J} \exp(\beta_k X_i)}$$

where J is the set of majors. In this specification, we use students' performance in each section of the test- Turkish, math, science, and social science- and their high school performance adjusted for school quality as explanatory variables.

We use the Blinder-Oaxaca decomposition method (Blinder [1973], Oaxaca [1973]) extended by Fairlie [2006] to decompose the effect of score differences. The average gender gap in majors can be decomposed as follows:

$$\begin{aligned}\bar{P}_j^F - \bar{P}_j^M &= \left[\sum_{i=1}^{N_F} \frac{F(X_i^F, \hat{\beta}_j^F)}{N_F} - \sum_{i=1}^{N_M} \frac{F(X_i^M, \hat{\beta}_j^F)}{N_M}\right] \\ &\quad + \left[\sum_{i=1}^{N_M} \frac{F(X_i^M, \hat{\beta}_j^F)}{N_M} - \sum_{i=1}^{N_M} \frac{F(X_i^M, \hat{\beta}_j^M)}{N_M}\right]\end{aligned}$$

The first part of the expression represents the differences due to observed score differences, and the second part represents the differences due to differences in the preferences and the factors other than performance of students. Alternatively, the decomposition could be constructed by taking male students as reference group, however this may provide different estimates. To prevent this, so-called index problem, we used the pooled sample of both groups as reference group as suggested in Oaxaca and Ransom [1994].

In addition to the overall effect of score distributions, we want to understand the individual contributions of specific variables. In a logit model, the individual effect of any variable depends on the values of all other variables. So identifying the effect of a variable is not straightforward. Fairlie [2006] suggest a method to identify the effects of a specific variable. Firstly, for the groups of interest, a sample of size equal to the size of the smaller group is drawn from the larger group. Based on the predicted probabilities estimated by using the both group, they are matched according to their rank. Then, for each sub-sample the effect of a variable is given by

$$\frac{1}{N_F} \sum_{i=1}^{N_F} F(\hat{\beta}_0^* + \hat{\beta}_1^* X_{1i}^F + \hat{\beta}_2^* X_{2i}^F) - F(\hat{\beta}_0^* + \hat{\beta}_1^* X_{1i}^M + \hat{\beta}_2^* X_{2i}^F)$$

where $\hat{\beta}^*$ is the vector of coefficients observed from the pooled data.

This process is repeated several times, and the average effect is found over the all sub-samples. For a large number of sample draws, the results will approximate the average contribution of variables. The next section presents the results.

5.2 The Results

We assume that each student has ten alternative to choose from, including the retaking and quitting options. We group college programs into eight groups as presented in Appendix C. We also assume that students have the option to choose to quit the system or to retake the exam.

We use the data from the survey question that asks about the plans of students regarding retaking and quitting to define these decisions. We assume that the student decided to retake the exam if i) the score of the student is above the threshold needed to submit a preference list, i.e., above 105; ii) Either the student did not submit a preference list, or he did not submit a feasible preference list, so he was not assigned to any program, iii) In a survey question that asks about the plans of the student, he mentioned that he will retake the exam if he cannot make it this time. Similarly, we assume the students who satisfy the first two conditions, and who answered to the survey question as they will not retake again, want to quit.

We draw 500 sub-samples to estimate the individual effects of variables. Table 9 displays the estimation results for males, females, and both groups. The results show that students' performance on math and science tests is particularly important in their college program choice decision. Estimation results suggest that a unit increase in the math (science) score increases the odds of choosing an engineering major by 6.2% (9.5%) and 8.2% (7.3%) for females and males, respectively. However, the same change increases odds of choosing an education program by 8.4% (2.9%) and 8.2% (-0.2%) for female and male students, respectively. Female and male students' probability of choosing a medicine program increases by 15% and 4.2% for a unit increase in the math score and 16.5% and 12.2% for a unit increase in the science score, respectively. These results suggest that the performance on science and math tests are the most important predictor of college major choice.

Table 3: Decomposition of Gender Differences due to Differences in Score Distribution

	Tech.& Eng.	Educ.	Medicine	Med. Serv.	Science	Mang. & Econ	Agriculture	Dist. Educ.	Retake	Quit
Difference	-0.092	0.051	0.006	0.035	0.016	0.006	-0.001	-0.001	-0.015	-0.005
Explained	-0.049	0.014	-0.012	0.006	-0.004	0.014	-0.001	0.000	0.018	-0.002
Unexplained	-0.043	0.037	0.018	0.029	0.021	-0.008	0.001	-0.002	-0.033	-0.004
Raw Math	-0.0174***	-0.00993***	-0.00399***	0.000	-0.00667***	-0.0166***	0.000	0.000	0.0262***	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.000)	(0.002)	(0.000)
Raw Science	-0.0230***	-0.001	-0.0105***	0.000	-0.001	0.0179***	0.000	0.000783**	0.0250***	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.000)	(0.000)	(0.002)	(0.000)
Raw Turkish	0.00309**	0.00751***	0.000	0.000	0.001	0.0169***	0.001	0.000	-0.0331***	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)
Raw Social Sci.	-0.00288*	0.00758***	-0.00293***	0.003	0.00260**	-0.002	0.000	0.000	0.00847***	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Adjusted GPA	-0.00839***	0.00979***	0.00476***	0.00405***	-0.001	-0.002	-0.00173**	-0.000697*	-0.00879***	-0.000816*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)	(0.000)

Note-1: Standard errors are reported in parentheses. *, **, *** indicate significance at the .90, .95 and .99 levels, respectively.
Note-2: Estimates are rounded due to presentation purpose, co very small standard errors appear as zero.

Table 3 shows the actual gender differences in majors, the differences due to different score distributions and the differences due to preferences and other factors.[6] Table 3 shows that the largest gap exists in engineering, education and medical service majors. While 18.4% of male students major in Engineering, 9.2% of females do so. According to the decomposition analysis, the differences in score distribution explain 53.1% of this difference. In a similar study that uses data from the US, Turner and Bowen [1999] find that 32.1% of the 8.4% gender gap in Engineering field is explained by differences in SAT score distribution. In the US, students choose their majors after they start college and experiment with different major related courses, however in Turkey they choose major and college together based on their performance in the exam. Stinebrickner and Stinebrickner [2014] find that students enter college quite optimistic about obtaining a science degree, however as their expectations regarding their performance in science field change, they leave the science major. The fact that the students' performance after they get into college plays an important role in their decision in the US system may explain the differences in our findings.

The differences in score distribution can explain only 28.1% of the 5.1% percent differences in education major and account for only 18.3% of 3.5% gender gap in the medical service programs. As shown before, demand for medical service departments come from lower performing female students, and male students are very unlikely to list these programs in their preference list[7]. Given the stereotypes regarding the nursing profession, it is not surprising to see the majority of the gender gap in medical service programs driven by preferences.

According to differences in score distributions, we expect to see more male students than female students in medicine and science programs. However, female's preferences, unexplained by score distribution, for these departments are so strong relative to male students that female students are slightly over represented in these majors.

In this section, we show that the performance, especially, in math and science fields is an important factor that affects students' college program choice decisions. This suggests

[6]Table 10 in the appendix presents the results where male students are taken as a reference group.

[7]In our sample, medical service programs appear 7290 times in the preference list of students, and only 756 of them in male students' preference list.

that policies that increase female students' performances in the math and science fields could help to reduce the gender gap in engineering and technical science programs dramatically. Our results also point to preference differences across gender. This may well be due to social pressures: for example, being a school teacher may be seen as an appropriate job for a woman if it leaves more time for family duties.

The results presented in Table 3 shows that given their score distributions, female students should be more likely to retake. However, in reality male students are more likely to retake the exam. In the next section, we investigate the factors that affect students' retake decision.

6 Decision to Retake

A large number of students retake the university entrance exam in Turkey. In 2002, two thirds of the students, or around one million students, took the university exam at least twice. As retaking is an important part of the system, it is important to understand which characteristics of students affect their decision to retake. In the previous section, we showed that male students are more likely to retake. This difference could be the effect of gender related restrictions: female students may not be allowed to retake the exam. It may also be that students' willingness to take the risk associated with retaking differs by gender. In the Turkish system there is no punishment for retaking, unless they are already placed in a program. If male students can avail of this retaking option but female students can less easily do so, retaking may increase the performance gap between female and male students. As a result, it will have an adverse effect on the gender gap in college programs. The gender gap in engineering programs is 10.7% among the second time taker students, and 66.4% of this difference can be explained by the differences in the score distributions. (see table 11)

We use the probit model as in 4, where the outcome variable, y_i, is a dummy variable that shows whether student i decided to retake the exam or not.

Table 4 shows the estimation results. Students are less likely to retake the exam as their score increases. This is expected given that as the score of a student improves, the quality of the programs in the feasible set of students[8] increases, which makes it hard to rationalize retaking. In our estimation, we included the differences between students' score and their score predicted by using their schools quality adjusted high school performance, income, gender, school type, preparation expenditure, number of siblings, education level of parents and experience in the exam. The coefficients of these variables are all positive and significant, which suggests that students are more likely to retake when they perform worse than expected. These findings are similar to the findings of Arcidiacono [2004] that students who perform worse than expected are more likely to switch to another majors or dropout of the college. However, in the Turkish system, retaking the exam is also an opportunity to improve: Frisancho et al. [2013] find large learning gains over attempts and more so for the less advantaged and may be a factor that levels the playing field.

Finally, our results suggest that male students are more likely to retake the exam while controlling for their allocation score and a rich set of other characteristics. In the literature, it is found that conditional on performance, males are more willing to compete, and females are more risk-averse relative to their male counterparts (see Niederle and Vesterlund [2011], Niederle and Vesterlund [2007], Eckel and Grossman [2008]). So male students' willingness to retake the exam might be due to their attitudes towards risk and competition. Our results also show that students living in less developed areas and with less educated mothers are more likely to retake. Although, the coefficient of interaction of gender and the development level of the city students live is positive for the less developed areas, which suggest male students in these areas more likely to retake relative to female students, the coefficient is not significant. The results suggest that the retaking option is perceived as an opportunity, especially by students coming from disadvantaged backgrounds.

Table 4: Probability to Retake

Variable	1st time takers	2nd time takers
Male	0.239***	0.201***
	(0.046)	(0.053)

(continued on next page)

[8]Feasible set of a student can be defined as the programs whose minimum admission cutoff scores are less than student's score.

Variable	1st time takers	2nd time takers
Y-OSS-SAY	-0.0452***	-0.0318***
	(0.001)	(0.001)
(Predicted Raw Math-Raw Math)[a]	0.00623*	0.0105***
	(0.003)	(0.003)
(Predicted Raw Science-Raw Science)[a]	0.00609*	0.005
	(0.003)	(0.003)
(Predicted Raw Turkish-Raw Turkish)[a]	0.0184***	0.0153***
	(0.002)	(0.002)
(Predicted Raw Social-Raw Social)[a]	0.00787***	0.00511**
	(0.002)	(0.002)
Income (base: <250 TL)		
$[250TL, 500TL]$	0.031	-0.035
	(0.038)	(0.040)
>500TL	0.045	-0.089
	(0.047)	(0.052)
Education of Dad(base: Primary or less)		
Middle/High school	0.028	0.042
	(0.040)	(0.042)
2-year higher education	0.106	0.034
	(0.063)	(0.079)
College/Master/Phd	0.051	0.139*
	(0.051)	(0.060)
Missing	-0.095	0.102
	(0.091)	(0.100)
Education of Mom (base: Primary or less)		
Middle/High school	-0.136***	-0.082
	(0.038)	(0.046)
2-year higher education	0.042	-0.041

(continued on next page)

Variable	1st time takers	2nd time takers
	(0.068)	(0.093)
College/Master/Phd	-0.151*	-0.147
	(0.063)	(0.087)
Missing	0.058	0.006
	(0.100)	(0.110)
$\mid DevelopmentIndex > 0 \mid$ [b]	-0.0366**	-0.0404*
	(0.014)	(0.016)
$\mid DevelopmentIndex < 0 \mid$ [b]	0.254*	0.261*
	(0.117)	(0.114)
Male*$\mid DevelopmentIndex > 0 \mid$	-0.030	-0.013
	(0.017)	(0.021)
Male*$\mid DevelopmentIndex < 0 \mid$	0.165	0.153
	(0.144)	(0.137)
Preparation Expenditure (base: No prep school)		
Scholarship	0.011	0.215*
	(0.093)	(0.099)
<1000TL	-0.066	-0.089
	(0.067)	(0.054)
$[1000TL, 2000TL]$	-0.146*	-0.219***
	(0.071)	(0.059)
>2000TL	-0.326***	-0.275***
	(0.080)	(0.072)
Missing	-0.206*	-0.087
	(0.085)	(0.062)
Constant	8.272***	5.206***
	(0.167)	(0.182)
School Type Control	Yes	Yes
N	13526	7870

(continued on next page)

Variable	1st time takers	2nd time takers
Note: Standard errors are reported in parentheses. *, **, *** indicate significance at the .90, .95 and .99 levels, respectively.		

[a] Difference between students' predicted raw scores and their real raw scores. Raw scores are predicted as follows: first we adjust students' high school GPA according to the quality of the school they attended in the following way:

$$A\_GPA_{ij} = \mu_{OSS-SAY,j} + \sigma_{OSS-SAY,j} \frac{GPA_i - \mu_{gpa,j}}{\sigma_{gpa,j}}$$

. Then by using adjusted GPA, gender, income, school type, preparation expenditure, number of siblings and education level of parents, we predict raw score.

[b] Development index of the city student lives

7 Conclusion

Gender differences in college major choice have significant implications in terms of the labor market outcomes of women and men. In this paper, by using data from a centralized system in which students are assigned to a college and a major simultaneously only based on their score in the university entrance exam, and their preference lists, we investigate the effect of score differences on the college program choice decision. In addition, we investigate factors affecting students' decision to retake. The retake decision has important implications for all students, since every year around one million students took the exam as repeat takers.

We firstly show that there are substantial gender differences in college major choice. Male students are overrepresented in engineering and technical science programs, while females are overrepresented in education programs. We also show that male students are performing better in math and science, while female students perform better on the Turkish part of the test. The differences in the score distribution in math and science field explains 50% of the gender gap in the engineering and technical science programs. These findings suggest that the policies that could help improve female students' achievements in science and math field could increase the number of women in engineering and technical fields.

We also show that although conditional on their scores, one would expect male students to be less likely to retake, they are more likely to choose retaking option. We provide evidence that this behavior is related to male students' willingness to take the risk and compete. These differences in willingness to retake lead to a higher gender gap in engineering programs among the second-time-taker students. On the other hand, we find that students coming from less advantageous backgrounds are more likely to retake.

References

Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004.

Alan S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):pp. 436–455, 1973.

Asena Caner and Cagla Okten. Risk and career choice: Evidence from Turkey. *Economics of Education Review*, 29(6):1060–1075, December 2010.

C. C. Eckel and P. J. Grossman. Men, women, and risk aversion: Experimental evidence. *Handbook of Experimental Economics*, 1(113):1061–1073, 2008.

Robert W. Fairlie. An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models. IZA Discussion Papers 1917, Institute for the Study of Labor (IZA), January 2006.

Verónica Frisancho, Kala Krishna, Sergey Lychagin, and Cemile Yavas. Better luck next time: Learning through retaking. Working Paper 19663, National Bureau of Economic Research, November 2013.

Ahu Gemici and Matthew Wiswall. Evolution of gender differences in post-secondary human capital investments: College majors. *International Economic Review*, 55(1):23–56, 2014. ISSN 1468-2354.

Linda Datcher Loury. The gender earnings gap among college-educated workers. *Industrial and Labor Relations Review*, 50(4):pp. 580–593, 1997.

Linda Datcher Loury and David Garman. College selectivity and earnings. *Journal of Labor Economics*, 13(2):pp. 289–308, 1995.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

Muriel Niederle and Lise Vesterlund. Gender and competition. *Annu. Rev. Econ.*, 3(1): 601–630, 2011.

Ronald Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709, October 1973.

Ronald L. Oaxaca and Michael R. Ransom. On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1):5–21, March 1994.

Ralph Stinebrickner and Todd R. Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, page 426–472, 2014.

Sarah E. Turner and William G. Bowen. Choice of major: The changing (unchanging) gender gap. *Industrial and Labor Relations Review*, 52(2):pp. 289–313, 1999.

Basit Zafar. College major choice and the gender gap. *Journal of Human Resources*, Forthcoming.

A Appendix

A.1 Tables and Figures

Table 5: Summary Statistics

	1st time takers			2nd time takers		
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Male	16252	0.568		8986	0.597	
Raw Math Score	16252	23.489	12.852	8986	23.127	12.187
Raw Science Score	16252	16.735	12.570	8986	15.497	11.583
Raw Turkish Score	16252	20.957	11.513	8986	19.919	10.840
Raw Social Science Score	16252	11.350	11.392	8986	9.635	10.155
OSS-SAY	16252	132.260	23.574	8986	130.374	20.885

(continued on next page)

	1st time takers			2nd time takers		
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
OSS-EA	16252	123.351	20.036	8986	121.521	17.088
Y-OSS-SAY	16252	164.002	27.104	8986	160.697	23.534
Y-OSS-EA	16252	155.023	23.513	8986	151.830	19.751
Weighted GPA-SAY (AOBP-SAY)	16252	63.483	10.088	8986	60.646	9.550
Weighted GPA-EA (AOBP-EA)	16252	63.345	10.084	8986	60.617	9.518
Normalized GPA	16252	52.779	10.242	8986	51.735	10.380
Students who submit preference list	16252	0.412		8986	0.529	
Mother's Education Level						
Primary or less	16252	0.564		8986	0.645	
Middle/High school	16252	0.246		8986	0.208	
2-year higher education	16252	0.048		8986	0.038	
College/Master/Phd	16252	0.075		8986	0.052	
Missing	16252	0.067		8986	0.057	
Father's Education Level						
Primary or less	16252	0.330		8986	0.380	
Middle/High school	16252	0.312		8986	0.317	
2-year higher education	16252	0.067		8986	0.057	
College/Master/Phd	16252	0.204		8986	0.168	
Missing	16252	0.086		8986	0.078	
Income						
<250 TL	16060	0.298		8862	0.348	
$[250TL, 500TL]$	16060	0.414		8862	0.405	
>500TL	16060	0.288		8862	0.247	
Reason to attend a University						
Have an occupation that fits my interest and abilities	16252	0.013		8986	0.045	
Improve my knowledge	16252	0.007		8986	0.030	
Obtain a well-paying job	16252	0.013		8986	0.040	

(continued on next page)

Variable	1st time takers Obs	Mean	Std. Dev.	2nd time takers Obs	Mean	Std. Dev.
To get the prestige of a diploma	16252	0.004		8986	0.013	
To satisfy family demands	16252	0.002		8986	0.006	
To make it easier my promotion at the workplace	16252	0.000		8986	0.012	
To postpone my military service	16252	0.000		8986	0.001	
To keep up with my peers	16252	0.000		8986	0.003	
Plan if not placed						
Quit	16039	0.013		8749	0.135	
Re-try after attending prep school next year	16039	0.615		8749	0.228	
Retry by studying on my own.	16039	0.370		8749	0.603	
Retry without preparation	16039	0.002		8749	0.034	

Table 6: College Programs open to Science Track Students: Summary Statistics

Science Score					
Variable	Obs	Mean	Std. Dev.	Min	Max
Minimum Cutoff in 2000	1448	181.981	18.513	126.635	226.133
Maximum Cutoff in 2000	1448	192.218	16.431	139.696	239.692
Number of Seats in 2000	1448	49.710	31.079	1	319
Minimum Cutoff in 2001	1549	178.140	22.320	125.512	225.541
Maximum Cutoff in 2001	1549	190.640	17.899	133.685	238.576
Number of Seats in 2001	1549	49.231	31.026	1	319
Minimum Cutoff in 2002	1639	176.563	21.905	110.894	223.208
Maximum Cutoff in 2002	1639	188.611	17.887	134.016	236.02
Number of Seats in 2002	1639	52.723	67.918	1	2521

(continued on next page)

Science Score					
Variable	Obs	Mean	Std. Dev.	Min	Max
Tuition Cost					
Public University	1391	283.433	221.049	35	1135
Private University	248	5124.667	5355.643	0	18712
Scholarship	1639	25.694	209.341	0	2196.628
If given	100	421.121	746.240	0	2196.628
Private University	1639	0.151		0	1
Technical Sci. and Engineering	1639	0.392		0	1
Education	1639	0.181		0	1
Medical Service	1639	0.085		0	1
Medicine	1639	0.058		0	1
Agriculture	1639	0.049		0	1
Science	1639	0.235		0	1
Distance Education	1639	0.001		0	1
Weighted according to quota					
Technical Sci. and Engineering	1639	19.412	30.495	0	217
Education	1639	9.262	21.830	0	155
Medical Service	1639	3.937	14.292	0	155
Medicine	1639	4.754	24.384	0	319
Agriculture	1639	2.698	12.656	0	104
Science	1639	11.121	22.730	0	186
Distance Education	1639	1.538	62.271	0	2521

Table 7: College Programs open to Science Track Students: Summary Statistics

Turkish-Math Score

Variable	Obs	Mean	Std. Dev.	Min	Max
Minimum Cutoff in 2000	374	169.680	15.026	109.766	207.69
Maximum Cutoff in 2000	374	178.934	12.022	135.982	211.394
Number of Seats in 2000	374	212.896	1879.452	1	34048
Minimum Cutoff in 2001	416	167.021	15.191	109.127	206.981
Maximum Cutoff in 2001	416	177.366	12.098	133.492	210.377
Number of Seats in 2001	416	209.921	2107.124	1	41359
Minimum Cutoff in 2002	449	164.855	14.817	109.447	206.102
Maximum Cutoff in 2002	449	175.732	12.695	131.82	209.683
Number of Seats in 2002	449	201.864	2123.482	2	44007
Tuition Cost					
Public University	344	311.875	208.059	35	615
Private University	105	5058.187	5000.327	0	17898.45
Scholarship	449	35.310	236.610	0	2033.915
If given	44	360.326	680.766	0	2033.915
Private University	449	0.234		0	1
Management and Economics	449	0.633		0	1
Education	449	0.301		0	1
Distance Education	449	0.004		0	1
Weighted according to quota					
Management and Economics	449	49.04232	56.19307	0	309
Education	449	30.36526	56.82103	0	258
Distance Education	449	119.9889	2127.377	0	44007

Table 8: Percentage of Male and Female Students w.r.t Rank of the Placed Department

	1st time takers		2nd time takers	
	Female	Male	Female	Male
Submit preference list	49.31712	46.0742	61.41338	57.54992
Not placed	20.331	19.7164	16.0387	16.5884
Placed Rank				
1	10.1993	9.84484	12.5255	11.6371
2	9.45626	8.72124	6.97556	7.44489
3	9.25363	10.1926	6.92464	6.93892
4	8.20669	8.66774	7.28106	6.8305
5	6.75448	7.6779	7.38289	6.36068
6	6.24789	5.51097	4.73523	5.52945
7	5.47113	5.21669	5.2444	5.20419
8	4.59304	4.92242	5.34623	4.55367
9	3.5461	3.42429	4.17515	4.11999
10	3.10706	2.56822	3.76782	3.21648
11	2.06012	2.11343	3.05499	3.43332
12	2.12766	1.84591	3.10591	2.85508
13	1.65485	1.89941	2.18941	2.20455
14	1.45221	1.65864	1.78208	2.45754
15	0.979399	1.07009	1.93483	1.91543
16	1.08072	1.15035	1.78208	1.95157
17	0.810537	0.722311	1.17108	1.44561
18	0.540358	0.882825	0.86558	1.44561
19	0.472813	0.454789	1.06925	0.795085
20	0.472813	0.535046	0.610998	0.758945
21	0.270179	0.34778	0.458248	0.614384
22	0.337724	0.107009	0.610998	0.758945
23	0.371496	0.294275	0.560081	0.361402

(continued on next page)

	1st time takers		2nd time takers	
	Female	Male	Female	Male
24	0.202634	0.454789	0.407332	0.578244

Table 9: Coefficient Estimates for Female and Male Students

	1st time takers					
	Male		**Female**		**Pooled**	
	Coef.	**Std. Err.**	**Coef.**	**Std. Err.**	**Coef.**	**Std. Err.**
Major						
Engineering & Technical Science						
Raw Math Score	0.079	0.008	0.060	0.012	0.080	0.007
Raw Science Score	0.071	0.007	0.091	0.010	0.081	0.005
Raw Turkish Score	0.020	0.004	0.020	0.008	0.009	0.003
Raw Social Science	0.005	0.004	-0.007	0.006	0.007	0.003
Normalized GPA	-0.019	0.005	-0.010	0.009	-0.028	0.004
Constant	-5.139	0.300	-6.084	0.559	-4.789	0.253
Education						
Raw Math Score	0.081	0.013	0.078	0.010	0.070	0.008
Raw Science Score	0.028	0.011	-0.002	0.008	0.003	0.006
Raw Turkish Score	-0.006	0.007	0.037	0.007	0.030	0.005
Raw Social Science	-0.029	0.006	-0.030	0.005	-0.038	0.004
Normalized GPA	0.042	0.008	0.039	0.008	0.056	0.005
Constant	-8.547	0.535	-7.826	0.514	-8.842	0.355
Medicine						
Raw Math Score	0.041	0.020	0.140	0.028	0.069	0.016
Raw Science Score	0.115	0.015	0.153	0.019	0.120	0.012
Raw Turkish Score	0.002	0.009	0.003	0.015	0.007	0.007
Raw Social Science	0.019	0.008	0.030	0.009	0.020	0.006
Normalized GPA	0.052	0.011	0.088	0.019	0.071	0.009
Constant	-11.984	0.784	-19.646	1.552	-14.648	0.706
Medical Service						
Raw Math Score	0.150	0.060	0.026	0.013	0.007	0.012
Raw Science Score	-0.017	0.044	0.018	0.012	0.002	0.011
Raw Turkish Score	-0.003	0.029	-0.049	0.009	0.003	0.008
Raw Social Science	-0.095	0.032	-0.030	0.009	-0.061	0.008
Normalized GPA	0.063	0.038	0.039	0.011	0.076	0.009
Constant	-13.725	2.598	-5.450	0.650	-8.805	0.595
Science						
Raw Math Score	0.062	0.012	0.085	0.012	0.067	0.008
Raw Science Score	0.026	0.011	0.002	0.010	0.009	0.007
Raw Turkish Score	-0.012	0.007	0.005	0.008	0.007	0.005
Raw Social Science	-0.009	0.007	-0.022	0.007	-0.023	0.005
Normalized GPA	-0.025	0.008	-0.008	0.009	-0.005	0.006
Constant	-3.742	0.472	-4.606	0.563	-4.757	0.347
Management & Econ						
Raw Math Score	0.085	0.013	0.092	0.014	0.094	0.010

(continued on next page)

	1st time takers					
	Male		Female		Pooled	
	Coef.	**Std. Err.**	**Coef.**	**Std. Err.**	**Coef.**	**Std. Err.**
Raw Science Score	-0.111	0.011	-0.094	0.012	-0.103	0.008
Raw Turkish Score	0.194	0.012	0.177	0.015	0.180	0.009
Raw Social Science	0.004	0.007	0.014	0.008	0.012	0.005
Normalized GPA	-0.011	0.009	0.001	0.012	-0.014	0.007
Constant	-8.207	0.591	-9.644	0.784	-8.368	0.445
Agriculture						
Raw Math Score	0.028	0.024	0.037	0.031	0.029	0.019
Raw Science Score	-0.011	0.024	0.023	0.030	0.002	0.019
Raw Turkish Score	0.005	0.016	0.043	0.026	0.017	0.013
Raw Social Science	-0.039	0.019	-0.017	0.022	-0.031	0.014
Normalized GPA	-0.064	0.020	-0.111	0.025	-0.077	0.015
Constant	-1.439	1.025	-0.276	1.319	-1.142	0.770
Distance Education						
Raw Math Score	0.011	0.024	-0.126	0.037	-0.028	0.019
Raw Science Score	-0.097	0.027	-0.023	0.042	-0.071	0.022
Raw Turkish Score	0.001	0.018	0.028	0.034	-0.002	0.014
Raw Social Science	-0.006	0.021	0.027	0.030	0.014	0.017
Normalized GPA	-0.026	0.020	-0.047	0.027	-0.040	0.016
Constant	-2.469	1.085	-1.154	1.578	-1.428	0.849
Retake						
Raw Math Score	-0.073	0.006	-0.077	0.007	-0.071	0.004
Raw Science Score	-0.062	0.005	-0.068	0.006	-0.062	0.004
Raw Turkish Score	-0.049	0.004	-0.041	0.005	-0.052	0.003
Raw Social Science	-0.023	0.003	-0.031	0.005	-0.021	0.003
Normalized GPA	-0.004	0.004	-0.027	0.006	-0.019	0.003
Constant	5.718	0.256	7.028	0.353	6.540	0.201
Quit						
Raw Math Score	0.008	0.021	-0.106	0.040	-0.010	0.018
Raw Science Score	-0.016	0.021	0.025	0.043	-0.002	0.019
Raw Turkish Score	-0.002	0.014	-0.027	0.032	-0.020	0.012
Raw Social Science	0.010	0.015	0.014	0.034	0.021	0.013
Normalized GPA	-0.041	0.017	0.024	0.031	-0.039	0.014
Constant	-2.253	0.877	-5.049	1.859	-2.177	0.759

Table 10: Decomposition of Gender Differences due to Differences in Score Distribution (Reference Group Male Students)

	Tech.& Eng.	**Educ.**	**Medicine**	**Med. Serv.**	**Science**	**Mang. & Econ**	**Agriculture**	**Dist. Educ.**	**Retake**	**Quit**
Difference	-0.092	0.051	0.006	0.035	0.016	0.006	-0.001	-0.001	-0.015	-0.005
Explained	-0.043	-0.003	-0.011	0.000	-0.010	0.022	-0.001	0.001	0.029	-0.001
Unexplained	-0.049	0.054	0.016	0.035	0.027	-0.016	0.001	-0.002	-0.044	-0.004
Raw Math Score	-0.0193***	-0.00784***	-0.00235*	-0.001	-0.00466***	-0.010	0.000	0.000	0.0266***	0.000
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.005)	(0.001)	(0.000)	(0.002)	(0.000)
Raw Science Score	-0.0224***	-0.00336*	-0.00941***	0.000	-0.00223*	0.0137*	0.000	0.00126**	0.0256***	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)	(0.000)	(0.000)	(0.002)	(0.001)
Raw Turkish Score	0.00731***	-0.001	0.000	0.000	-0.002	0.0204***	0.000	0.000	-0.0307***	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)
Raw Social Science Score	-0.002	0.00447**	-0.00261*	0.001	0.001	-0.001	0.000	0.000	0.00963***	0.000
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.000)	(0.000)	(0.002)	(0.000)
Adjusted GPA	-0.00607***	0.00516***	0.00355***	0.001	-0.00273*	-0.001	-0.00133*	-0.001	-0.002	-0.00112*
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.002)	(0.000)

Note-1: Standard errors are reported in parentheses. *, **, *** indicate significance at the .90, .95 and .99 levels, respectively.
Note-2: Estimates are rounded due to presentation purpose, co very small standard errors appear as zero.

Table 11: Decomposition of Gender Differences due to Differences in Score Distribution (2^{nd} time taker students)

	Tech.& Eng.	**Educ.**	**Medicine**	**Med. Serv.**	**Science**	**Mang. & Econ**	**Agriculture**	**Dist. Educ.**	**Retake**	**Quit**
Difference	-0.107	0.034	-0.016	0.082	0.035	0.005	-0.008	0.006	-0.024	-0.006
Explained	-0.071	-0.007	-0.017	0.009	-0.019	0.011	-0.002	0.013	0.044	0.015
Unexplained	-0.036	0.041	0.002	0.073	0.053	-0.006	-0.006	-0.007	-0.067	-0.022
Raw Math Score	-0.0263***	-0.0181***	-0.00465***	0.000	-0.0177***	-0.0167***	-0.001	0.00576***	0.0478***	-0.003
	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.004)	(0.001)	(0.001)	(0.003)	(0.003)
Raw Science Score	-0.0329***	-0.00686***	-0.0104***	-0.001	-0.00703**	0.0175***	0.001	0.00677***	0.0370***	0.0143***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.001)	(0.001)	(0.003)	(0.003)
Raw Turkish Score	-0.001	0.00912***	0.000	0.00303*	0.002	0.0148***	0.000	0.002	-0.0412***	-0.00465*
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
Raw Social Science Score	-0.00469**	0.00478***	-0.00436***	0.00259**	0.00885***	-0.00387***	0.000	0.000	0.00284***	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Adjusted GPA	-0.00637***	0.00420***	0.00187**	0.00367***	-0.00463***	-0.001	-0.00213***	-0.001	-0.00308*	0.00891***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)

Note-1: Standard errors are reported in parentheses. *, **, *** indicate significance at the .90, .95 and .99 levels, respectively.
Note-2: Estimates are rounded due to presentation purpose, co very small standard errors appear as zero.

Figure 1: Gender Differences in Major Choice

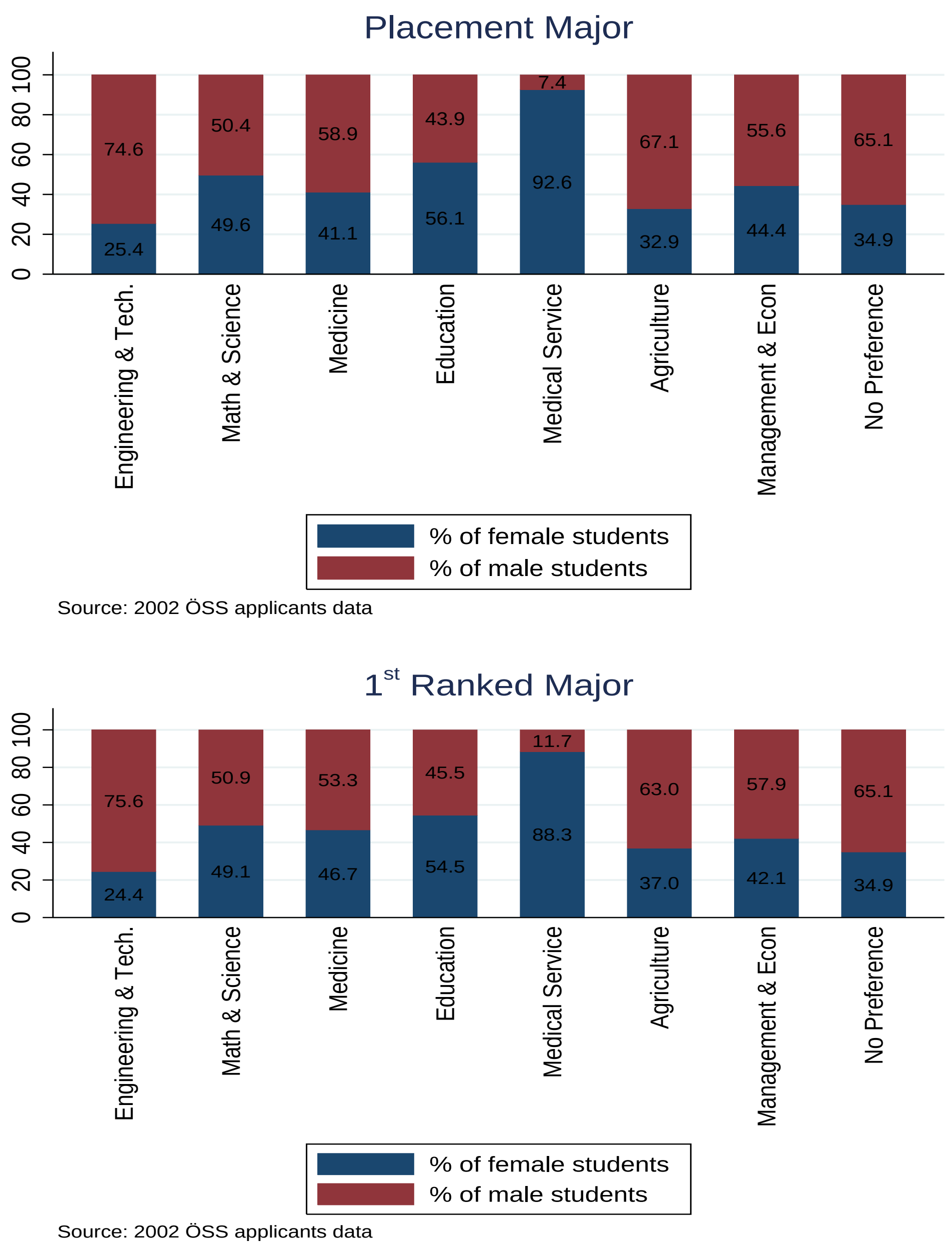


Figure 2: Gender Differences within Education Majors

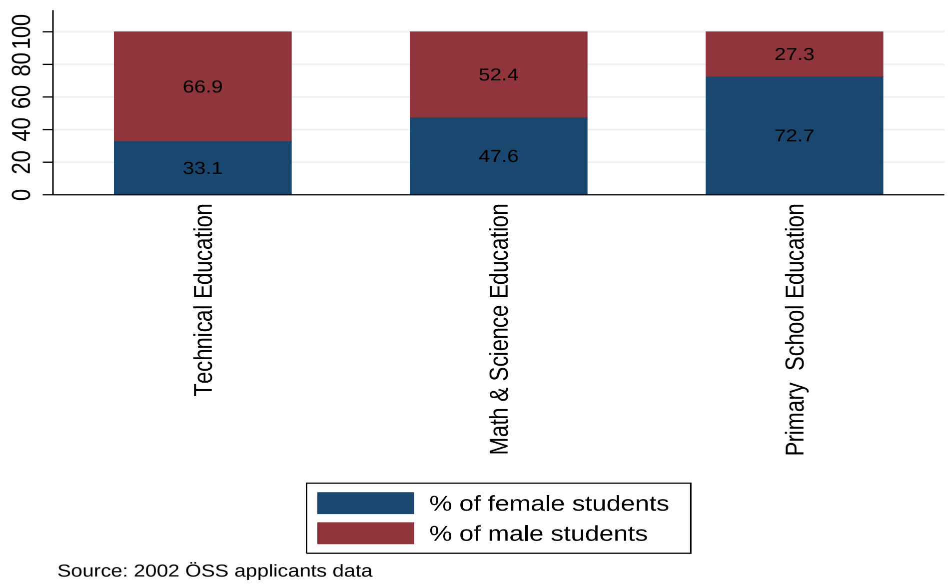


Figure 3: Allocation Score (Y-OSS-X) Distributions

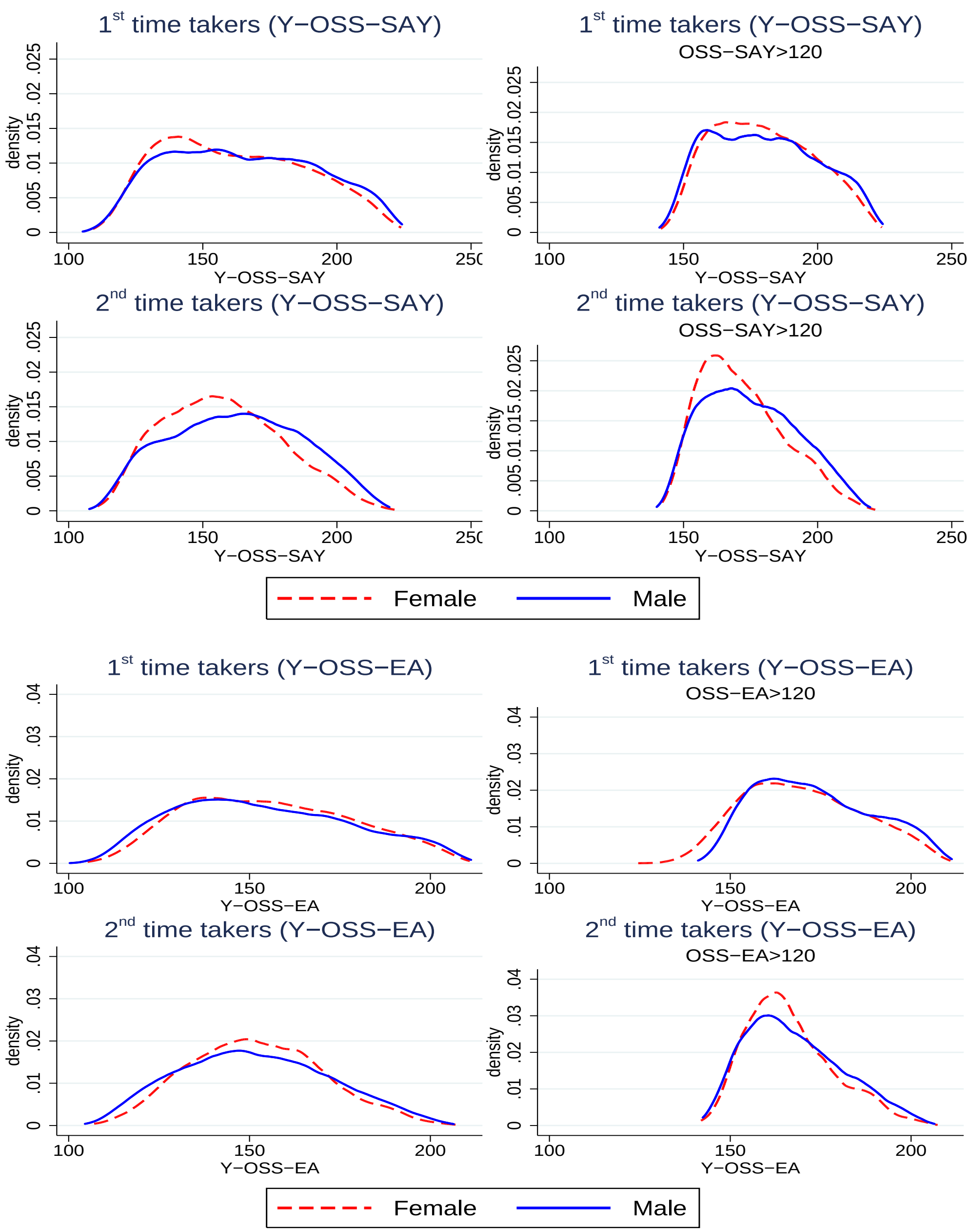


Figure 4: Raw Score Distributions

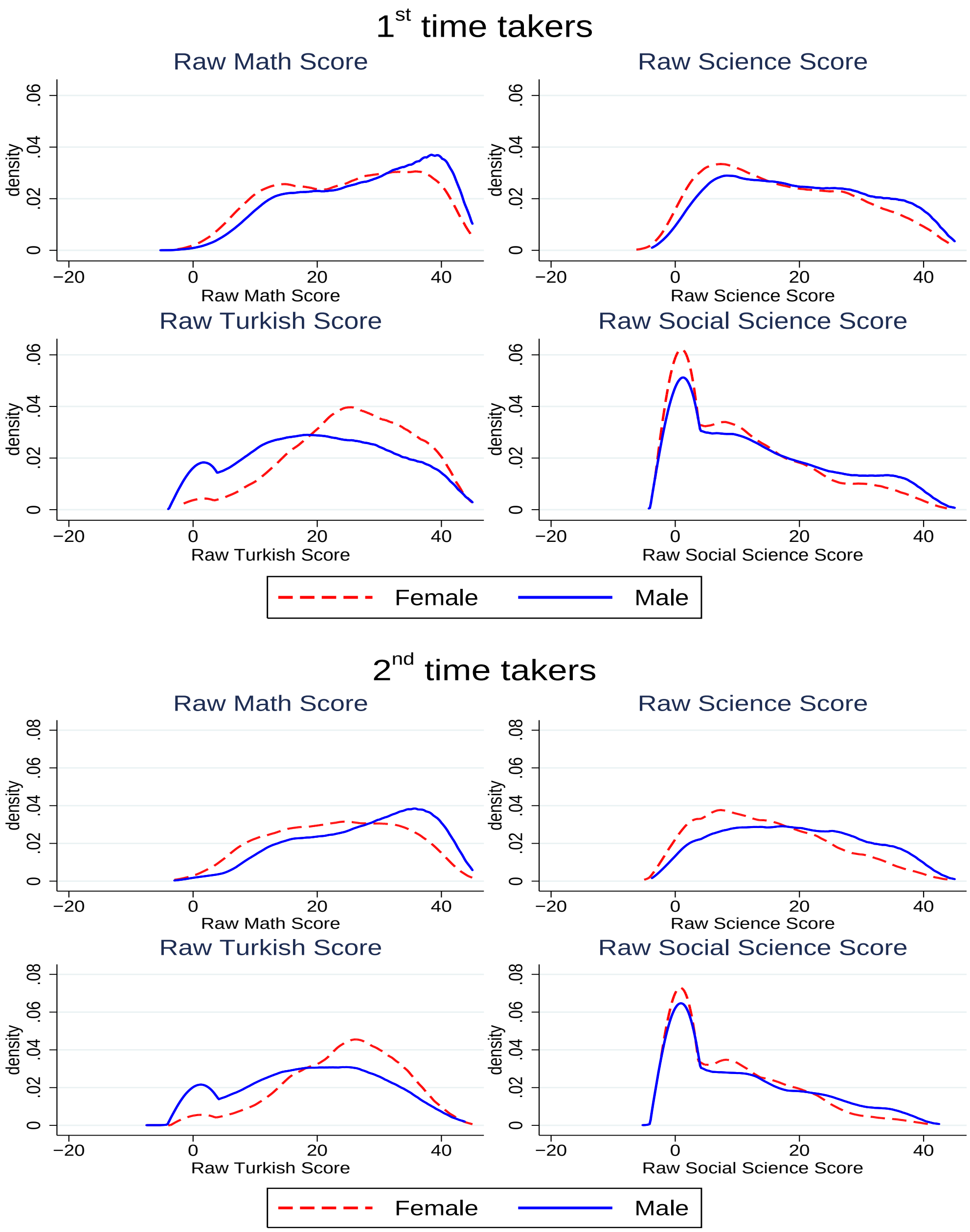


Figure 5: Adjusted GPA

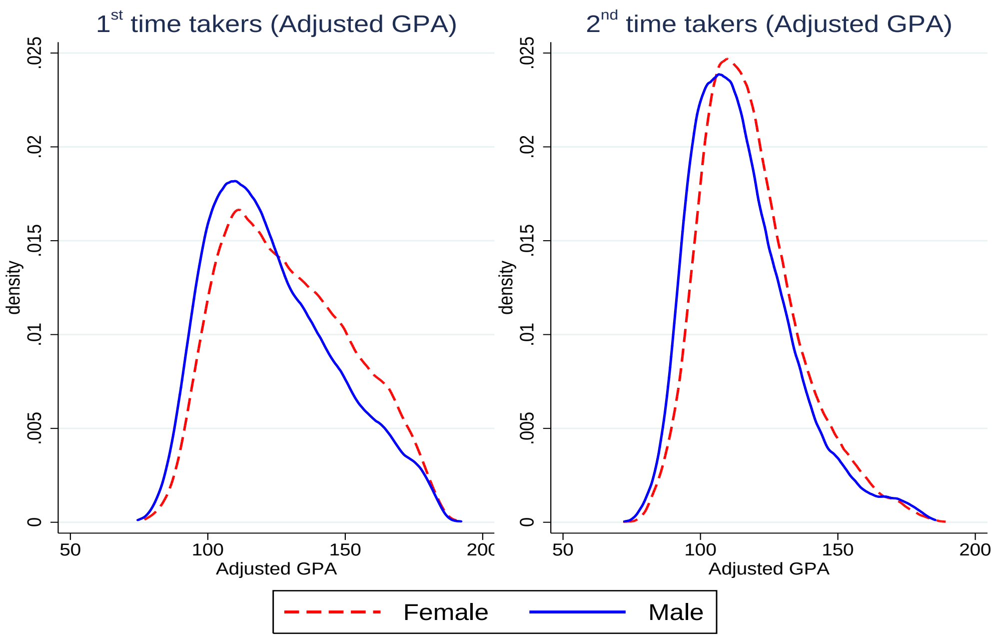


Figure 6: Preferred Departments

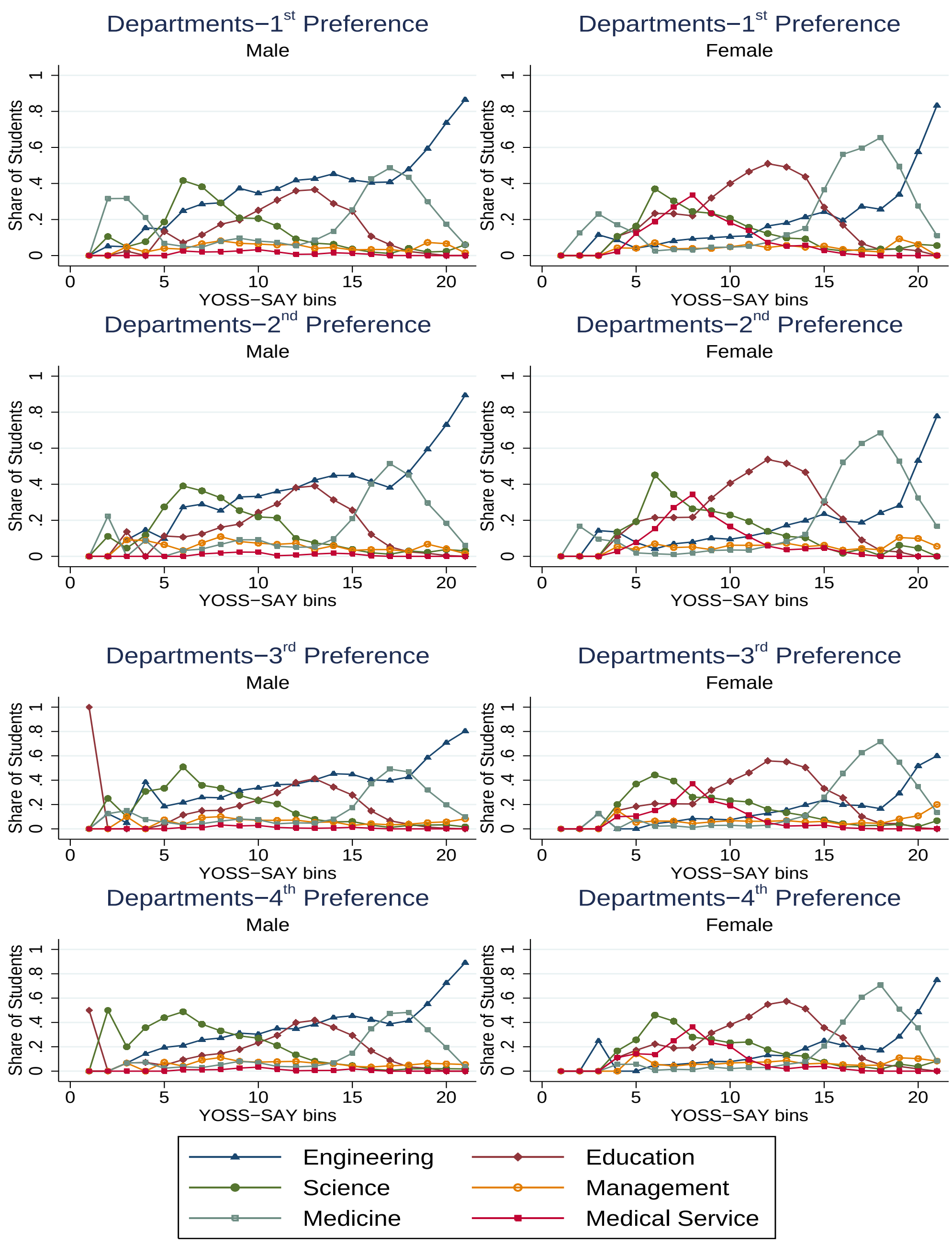
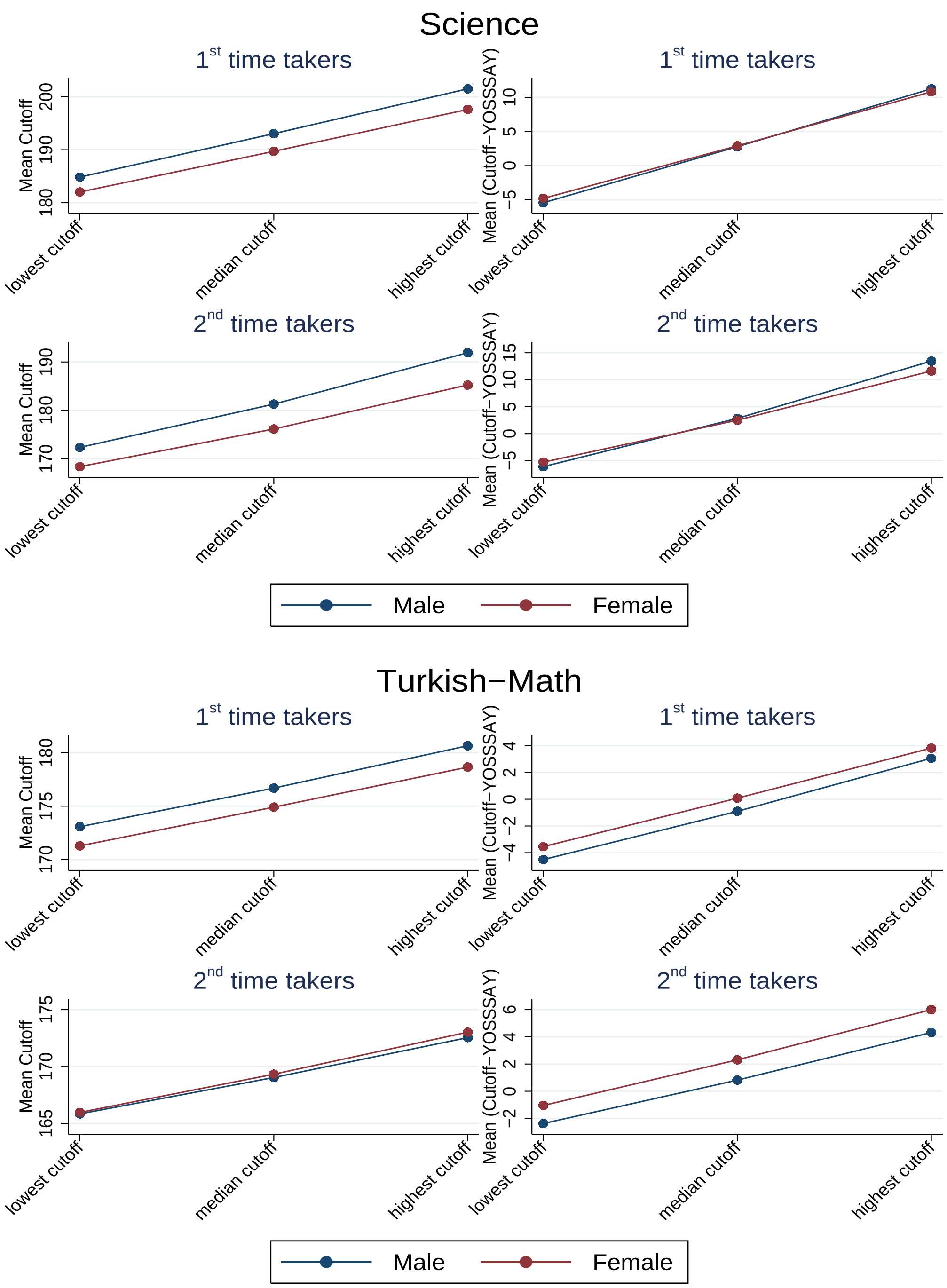


Figure 7: Mean Selectivity (Cutoff) of Preferences w.r.t Rank



B Allocation Score (Y-ÖSS)

University Entrance Exam allocation score (Y-ÖSS) is a function of students' ÖSS scores and their weighted normalized high school grade points (AOBP).

$$\text{Y-ÖSS-X}_i\text{=ÖSS-X}_i + \alpha \text{AOBP-X}_i$$

where X$\in${SAY,SOZ,EA,DIL}, and α is a pre-determined constant, changes according to students' track, preferred department and whether student placed in a regular program in the previous year or not. ÖSYM publishes the lists of departments open to students according to their tracks. When students choose a program from this list, α equals to 0.5. If it is outside the open list, α equals to 0.2. If student graduated from a vocational high school and prefer a department that is compatible to his high school field, α equals to 0.65. If student placed in a regular university program in the previous year, student punished and α equals to 0.25, 0.1, and 0.375, respectively, that is, for those students α coefficient equal to half of the regular α.

AOBP score is a function of normalized high school GPA (OBP), minimum and maximum normalized high school GPA in the school, student graduated from, and the mean ÖSS score in his school in the year of his graduation. Students keep their AOBP over the attempts. We don't observe students' AOBP score in our data set. However, we observe normalized high school GPA (OBP) for all students. We also got the school level mean ÖSS scores in each field for the 2002 high school graduates from ÖSYM website.

$$AOBP_{ij} = F[OBP_i, \mu_j, \sigma_j, \min_{i \in j}(OBP_i), \max_{i \in j}(OBP_i)]$$

We observe students' OBP, schools' μ_j, and σ_j. However, we don't observe minimum and maximum normalized high school GPA (OBP) in a school, but we have sample of students from each school. We need a good estimate of minimum and maximum OBP to calculate AOBP.

ÖSYM calculate OBP as follows:

Let $\mu_{gpa,j}$ =Average GPA in school j and $\sigma_{gpa,j}$ =Standard deviation of GPA within School j

Student i's, graduating from school j, OBP score is given as

$$OBP_{ij} = 10\frac{gpa_i - \mu_{gpa,j}}{\sigma_{gpa,j}} + 50$$

OBP is a function of students own GPA (observed in the data), and mean and standard deviation of GPA in their school. If we can observe at least two students with OBP and GPA available in our data set from a school, we can solve for μ_{gpa} and σ_{gpa} for that school.

- OBP is between 30 and 80 (This is a rule set of ÖSYM, if OBP formula gives less than 30, it is set to 30 and if it is more than 80, it is set to 80)
- OBP formula suggests that average student in each school gets 50 as OBP. Therefore, maximum OBP cannot be less than 50 and minimum OBP cannot be more than 50.

To get some idea about maximum OBP in a school, we look at the schools that we observe their first ranking students

- In the data set we can observe 445 first ranked students. Summary statistics of OBP of these students are as follows:

	# of Obs	Mean	Std	Min	Max
OBP	445	71.03213	5.557276	55.538	80

This means that on average first rank students are two standard deviations away from mean GPA. However, we should also note that since GPA is bounded from above with 5, depending on mean GPA, maximum OBP also bounded from above. That is, if mean GPA in a school is very high, maximum OBP is smaller. To find maximum possible OBP in a school, we calculate the OBP of the student with GPA 5, $\max_{obp,j}$, in schools where we don't observe first ranking student. Notice that we don't know whether there exist a student with GPA 5, but we know that max OBP cannot be higher than calculated OBP for this hypothetical student.

In the next step, we assume that OBP scores in each school has beta distribution with mean 50, standard deviation 10, and supports $[30, max_{obp,j}]$

Firstly, for each school we find the parameters of distribution given mean, standard deviation and support of the distribution for this school. Since mean and standard deviation are same in all schools, parameters differ in each school only because of the different support of the distribution.

In second step we draw from the beta distribution with the parameters estimated in the first step S times, and find average minimum and maximum OBP over samples as

$$\min_{i\in j} OBP_i = \frac{1}{S}\sum_{k=1}^{S} \min_{i\in j} OBP_i^k$$

$$\max_{i\in j} OBP_i = \frac{1}{S}\sum_{k=1}^{S} \max_{i\in j} OBP_i^k$$

Finally, we match estimated minimum and maximum OBP scores with our data set. If we observe a lower bound for OBP in our data set than what simulated, we use it as min OBP for this school, or if we observe higher maximum OBP, we use it as max OBP for this school. Otherwise, we use simulated minimum and maximum OBP scores.

Next section explains and illustrates how to compute AOBP score.

B.0.1 AOBP

$AOBP$ =Weighted Normalized high school gpa

OBP =Normalized high school gpa of the student

B =Highest OBP in the school

C =Lowest OBP in the school

$\ddot{O}SS = A$ =Mean OSS score in the school

$$D = \frac{80 - [(\frac{\ddot{O}SS}{80} \times C) - (\frac{\ddot{O}SS-80}{10})]}{(\frac{\ddot{O}SS}{80} \times B) - (\frac{\ddot{O}SS}{80} \times C)}$$

$$\begin{aligned} (\frac{\ddot{O}SS-80}{10}) &\geq 0 \\ B &> C \end{aligned}$$

$$AOBP = [(\frac{\ddot{O}SS}{80} \times C) - (\frac{\ddot{O}SS - 80}{10})] + [(OBP \times \frac{\ddot{O}SS}{80}) - (\frac{\ddot{O}SS}{80} \times C)] \times D$$

$OBP = 56.862$

$B = 74.184$

$C = 32.646$

$2002 - OSS - SOZ = 122.440$

$$\begin{aligned} AOBP - SOZ &= [((A/(80)) \times C) - (((A - 80)/(10)))] \\ &\quad +[(OBP \times (A/(80))) - ((A/(80)) \times C)] \times D \end{aligned}$$

$X = \frac{A}{80} = \frac{122.440}{80} = 1.530\,5$

$Y = \frac{A-80}{10} = \frac{122.440-80}{10} = 4.244$

$D = \frac{80-XC+Y}{X(B-C)} = \frac{80-1.5305*32.646+4.244}{1.5305*(74.184-32.646)} = 0.539\,2$

$S_1 = XC - Y - XCD$

$S_2 = XD$

$$\begin{aligned} S_1 &= XC - Y - XCD \\ &= 1.5305 * 32.646 - 4.244 - 1.5305 * 32.646 * 0.5392 = 18.780 \end{aligned}$$

$$S_2 = XD = 1.5305 * 0.5392 = 0.82525$$

$$\begin{aligned} AOBP - SOZ &= (1.5305 * 32.646 - 4.244) + 1.5305 * (56.862 - 32.646) * 0.5392 \\ &= 65.705 \end{aligned}$$

C Department Classification

Language and Literature:

Turkish Language and Literature

Western Language and Literature

Eastern Language and Literature

Ancient Language and Literature

Linguistics

Interpreting and Translating

Comparative literature

Math and Science:

Mathematics

Physics

Chemistry

Biology

Molecular Biology and Genetics

Statistics

Medical Service:

Nursing

Midwifery

Phy& Rehabilitation

Social Work

Health Visitor

Social Science:

Anthropology

Archeology

Geography

Theology

Philosophy

Psychology

History of Art

Sociology

History

Engineering & Technical Science:

Engineering

Architecture

Interior Design

City and Regional Planning

Industrial Production Design

Medicine:

Medicine

Veterinary Sciences

Dentistry

Pharmacy

Education:

Vocational Education

Technical Education

Math & Science Education

Social Science Education

Language Education

Public Relations:

Journalism & Public Relations

Advertising

Radio, Cinema, Television

Communication Design

Agriculture:

Plant Production

Animal Production

Landscape Architecture

Fish and Fisheries

Agricultural Equipment

Management & Economics:

Banking and Insurance

Public Finance

Accounting

Management

Economics

Marketing

International Trade

Econometrics

Public Administration and Law:

Political Science

Law

Public Administration

International Relations