Gender Differences in Willingness to Guess on High-Stakes Standardized Tests *

Ann Atwater†
Perihan O. Saygin‡

February 8, 2020

Abstract

Around the world, multiple-choice tests are widely used as part of high-stakes examinations, such as college-admissions tests. To counteract the ease of guessing, many of them have instituted a penalty for wrong answers. This results in the potential for takers with the same level of knowledge to have different scores due to their different willingnesses to guess. In this paper, we use administrative data from the Turkish college admissions test to study gender differences in willingness to guess and the heterogeneity in these differences across subjects, difficulty levels, and stakes. By exploiting the tracking system and using the resulting variation across different test sections, we find that female test-takers skip significantly more questions than male test-takers in quantitative tracks while we do not find a significant difference in other tracks. The gender gap is larger especially in Math among quantitative track students and when questions are more difficult. Our results suggest that the gender gap in willingness to guess is not only driven by risk aversion but may also differ across subject and difficulty levels. We also find that males are more likely to report that they are good at Math, Science, and Social Sciences conditional on the correct answers in the corresponding subjects of the test, and in all subjects, self-assessment is related to guessing behavior and gender differences in guessing shrink when controlling for self-assessment.

JEL Classification: J16, I23, D80
Keywords: Multiple-choice test, guessing, gender gap.

---

*This is a preliminary draft. Please do not quote or distribute without permission. Perihan Saygin would like to thank to David Card for valuable advice. This paper was partly written during her PhD and the earlier version of this paper was a working paper titled "Gender Differences in Guessing on Multiple-Choice Tests". Saygin is grateful to the Student Selection and Placement Center (OSYM in Turkish) in Turkey for sharing data. We would like to thank also Pamela Campa, Scott Kostyshak, Dana Sisak, Brenden Timpe, and seminar participants at SITE - Stockholm School of Economics for thoughtful comments. All errors are our own. Corresponding author's email: psaygin@ufl.edu

†Department of Economics, University of Florida.

‡Department of Economics, Robert F. Lanzillotti Public Policy Research Center, and affiliate faculty, Education Policy Research Center, University of Florida.

1 Introduction

Each year, millions of people take multiple-choice standardized exams. These exams are a key component of the college admissions process in countries around the world. As access to college is associated with a potential increase in earnings and certain colleges stand to confer larger income improvements than others, these exams have the potential to change students' lifetime earnings through their impact on participant college access.

While multiple-choice exams benefit from being easy to grade for large numbers of applicants, their accuracy is vulnerable to non-test related cognitive and non-cognitive issues[1]. Furthermore, to counteract the ease of guessing on these multiple-choice standardized exams, many subtract points for incorrect answers. Traditionally, these penalties are calculated to be risk neutral for students. While a risk-neutral penalty for guessing may cause an expected gain of zero points should a student guess without being able to eliminate any answers, the expected gain is greater than zero points should a student eliminate one or more answers. As such, if some students are less willing to guess even after eliminating these answers, then their expected performance will be less than their counterparts.

Prior research has found that male students are more willing to guess on standardized exams. Studies with observational data [Pekkarinen, 2015, Akyol et al., 2016], with field experiments [Gershon and Yakov, Espinosa and Gardeazabal, 2010, 2013, Riener and Wagner, 2018, Funk and Perrone, 2016, Iriberri and Rey-Biel, 2019, Coffman and Klinowski, 2020], and using lab experiments [Baldiga, 2014] have all replicated this finding.

Despite this, the literature is split on if gender differences in willingness to

[1]Non-test related issues are associated with long-run effects on wages. Ebenstein et al. [2016] found that Israeli test takers who were exposed to higher levels of air pollution on the day that they took the Bagrut standardized pre-college exam had statistically significant diminished monthly earnings a decade out from the exam.

guess represent a possible cause of the observed differences in male and female standardized test scores. Akyol et al. [2016] and Funk and Perrone [2016] concluded that differences in risk aversion do not explain overall gender gaps on standardized exams. In a large natural field experiment, Iriberri and Rey-Biel [2019] found significant gender differences in willingness to guess and a significant negative effect on females' performance resulting from this[2].

Previous research has suggested possible sources that may be responsible for this difference in willingness to guess between male and female test takers such as differences in knowledge, risk aversion and self-confidence. If students differ in terms of risk aversion or self confidence, this may lead to the observed gender differences in skipped questions and test outcomes. An extensive body of literature has provided evidence on gender differences in risk aversion [Eckel and Grossman, 2008, Croson and Gneezy, 2009, Filippin and Crosetto, 2016] and in self-confidence [Beyer, 1999, Barber and Odean, 2001]. Gender differences resulting from the high-stakes nature of standardized exams is also documented by previous literature [Niederle and Vesterlund, 2007, 2010, Azmat et al., 2016, Ors et al., 2013, Jurajda and Munich, 2011].

Gender differences in guessing behavior may also depend on other factors. For example, stereotype threat[3] may result in different answering patterns on more difficult problems, namely that girls are less likely to answer them. In an experimental study that evaluated the willingness to guess the answers to mathematics questions of elementary school students, Riener and Wagner [2018] provided results that are consistent with stereotype threat causing the observed gender difference in willingness to guess the answers to more difficult problems. Iriberri and Rey-Biel [2019] also provide evidence for heterogeneity in

[2] An important difference of this study is that the differential scoring rule rewards omitted questions rather than penalizing wrong answers. This may imply lower omitted questions as opposed to penalty in the case of risk aversion. Thus, it is reasonable to expect that this is a lower bound of the gender gap and its effect on test scores.

[3] The idea that prejudicial social stereotypes cause worse individual performance.

gender differences in guessing by showing that high-ability female participants of a mathematics competition were less willing to guess than low-ability female participants.

In this paper, we seek to quantify heterogeneity in differential guessing behavior across different subjects, levels of difficulty, and stakes. This is the first study in the literature to look at heterogeneity in guessing behavior along these axes using solely administrative data from a high-stakes exam rather than experimental data.

The use of administrative data is valuable in studying this topic as it allows for us to have larger sample sizes than would be feasible using experiments. Furthermore, our results do not have the problems with incentives that are faced by experimental studies. As most experimental studies reward correct answers with some sort of reward (typically money), the motivation for a correct answer is different than on an actual exam.

The effect of self-assessment or self-efficacy which measures students' beliefs about their performance on their educational outcomes and choices has also attracted attention in the literature. Murphy and Alexander [2000] and Perez-Felkner et al. [2017], and Saltiel [2019] find that math self-efficacy explains part of the gender gap in STEM fields. These studies show significant differences in math self-efficacy, where even among students in the top math test score quintile, men's observed self-efficacy exceeds that of women. In line with this literature, we explore the gender differences in self-assessment conditional on their performance on the test. We also explore the role of self-assessment in explaining the gender gap in willingness to guess.

We analyze the results from the college entrance exam, known as the ÖSS, used by Turkey as part of its college admissions process. This data set has a few key features to it that are conducive to our research. Unlike many other college systems, Turkish admissions decisions are decided purely by algorithm. This is in contrast to the college admissions process used in countries such as the United States where the college admissions process uses non-numerical information, such as essays and club participation, as part of the college admissions process. As such, the importance of questions on the ÖSS is high.

Furthermore, in the Turkish school system, students choose a course of study known as a "track" at the beginning of high school. This track determines the college majors that students are able to apply to and the weighting of questions on the exam. As a result of the different weights applied to sections of the test, guessing is riskier on certain subjects than on others[4]. As such, it allows us to vary the importance of the questions for students and see if the gender difference in willingness to guess changes with the stakes involved. Furthermore, as some sections of the test have the same stakes and different difficulty levels, this allows us to analyze the gender differences in guessing *controlling for stakes or the risk involved in guessing.*

Our results show that female test-takers who specialize in quantitative subjects skip significantly more questions than male test-takers. We do not find a significant gap for students specializing in qualitative subjects. For quantitative track test takers, the gender gap is significant in math, science, and social science sections and is larger when questions are more difficult. Interestingly, we find very similar gender differences in willingness to guess for quantitative track test takers in the science and social science sections even though the stakes of guessing are much lower in the social science section. Additionally, we find a remaining significant gender gap in skipped questions in the more difficult math section even after controlling for the skipped questions in the easier math section.

[4] A quantitative track student would have it be riskier to guess on math or science questions than a literature question as those questions are worth more to the student's overall score.

We also use certain survey questions relating to self-assessment to proxy self-confidence. Applicants are asked whether they believe themselves to be good in math, science, literature, and social science. We find that in all subjects, self-assessment is significantly related to guessing behavior in the corresponding subject. Moreover, applicants who find themselves good at math, tend to guess less in all sections except for literature. That is, perceived self-efficacy in math is related to guessing behavior in other subjects as well. We also find that gender differences in guessing shrink when controlling for self-assessment in the corresponding subject. Finally, we also provide evidence on gender differences in self-assessment conditional on achievement on the test. We find that males are more likely to report that they are good at math, science, and social science conditional on the correct answers in the corresponding subjects of the test while females are more likely to report that they are good at literature conditional on the correct answers in the literature sections.

The rest of the paper's organization is as follows: sections 2 and 3 detail the Turkish educational system's structure as well as our data sources. Section 4 describes our measure used as a proxy for guessing behavior which we call "Unwillingness to Guess". Finally, Section 5.1 details the results on gender differences in willingness to guess and Section 5.2 elaborates on the impact of gender differences in self-assessment. We conclude in Section 6.

2 Data and Institutional Setting

Turkey uses a centralized college admissions process primarily based on a nationally administered standardized exam. This exam is called the "Student Selection Exam" or ÖSS in Turkish. It is administered by the Student Selection and Placement Center or ÖSYM in Turkish. Anyone who has graduated from high school or is able to graduate from high-school is eligible to take the exam and apply to Turkish universities. The only non-test component of admissions decisions is the high school GPA of applicants, which is submitted by high schools across the country to the ÖSYM. However, admissions results are primarily derived from exam performance.

Meanwhile, universities in Turkey have a limited number of applicants that they can accept each year decided by the Higher Education Council. Students choose which universities to apply to after receiving their placement scores from the OSS. The cutoff scores from previous years are available for students to use in deciding which universities and majors to apply to. It is possible for an applicant to retake the examination in subsequent years should the exam not yield their desired result. As there are much fewer spots than the number of applicants, the number of applicants retaking the test grows each year. Due to this fierce competition, every single point matters as it may cause a substantial change in the ranking of applicants[5].

The Turkish high school and college admission system provides an interesting setting to study the gender differences in willingness to guess for three key reasons. First, as we described above, the centralized system provides us with administrative data on high school performance, college applications, and test scores. Second, the fierce competition allows for a real high-stakes setting to study potential gender differences in test-taking behavior rather than a lab setting. Last but not least, the tracking system and the structure of the test provides us with variation in the stakes and difficulty across sections of the test.

The Turkish high school curriculum features high school students being split into tracks of their choosing in their second year. The tracks are Science-Mathematics (Quantitative), Turkish-Mathematics (Equally Weighted), Social Sciences (Qualitative), Foreign Languages, and the Arts. These fields line up

[5]For example, at the 95th percentile, 1 point change in test score leads to a change of 1689 in the rankings while it causes a change of 13,390 in the rankings at the 10th percentile. Please see Table 1.

with the topics covered on the ÖSS. It has a total of four core sections on it: social science (which covers history, geography, and philosophy), science (biology, chemistry, and physics), mathematics, and literature[6]. Each subject is comprised of a lesser difficulty section and a higher difficulty one. Thus, there are 8 sections in total, and in our notation, we will use Math 1 to refer to the less sophisticated math section and Math 2 to refer to the more sophisticated math section. Students have a set amount of time to work on all sections and may move freely between them during the allocated time.

The format of all exam sections is the same. They are all multiple-choice, with each question having five answer choices. A full point is given for a correct answer while an incorrect answer results in the loss of a quarter of a point. Skipping a question results in a student receiving no points from it.

To generate the test score, a total is calculated for each of the eight sections. Then, students scoring above a certain amount have six different placement scores calculated using the raw section scores[7]: Quantitative 1, Quantitative 2, Equally Weighted 1, Equally Weighted 2, Qualitative 1, and Qualitative 2. These scores are calculated by giving differing weights to the raw scores obtained from each section. As can be seen in Table 2, the Quantitative 1 test score is calculated based on the raw scores of the Mathematics 1, Science 1, Social Sciences 1, and Literature 1 sections with the higher weights given to the math and science sections while the Quantitative 2 test score is calculated by including more difficult sections (Mathematics 2 and Science 2) in addition to the 4 lower difficulty sections. While the math sections always have the highest weight, more difficult sections have the same weights as the less difficult ones.

These weights result in significant heterogeneity in the value of the various sections to applicants. While a student on the quantitative track would obtain

[6]The Foreign Language section is an additional test and is not taken by all applicants.
[7]The foreign language section has its own seventh score.

the highest points by answering questions in the Math 1 and Math 2 sections, they will receive no points from correct answers in the Social Science 2 section and fewer points from the Social Science 1 section than a student on the qualitative track[8].

As a result of the different weights applied to section scores, certain sections are riskier to guess on than others. A student on the quantitative track would have it be more risky to guess on a math or science question than a social science or literature question as those sections are worth more to the student's overall score.

This setting allow us to study the guessing behavior under differing stakes through the various sections. Moreover, it also enables an analysis of gender difference in willingness to guess on certain sections conditional on the willingness to guess on other sections that are relatively less important. Finally, we are also able to make an assessment whether there is still a remaining gender gap in more difficult sections when it is estimated conditional on the willingness to guess on the easier section of the same subject with the same importance to test score calculation.

In this paper, we use two linked data sources to draw our conclusions. The first of these is the 2008 ÖSS Administrative Dataset which contains the number of correct and incorrect answers in each section, overall test scores, high school type, student track, and high school GPA's for 10,000 randomly selected college applicants. It should be noted that due to the use of an algorithm to determine admissions decisions, this data set comprises all relevant factors for college admissions for these 10,000 applicants. As such, there are no omitted variables that are used in determining admissions decisions.

The second source of data is the 2008 Survey of ÖSS Applicants. This details the applicant's socioeconomic characteristics and if they received private

[8]Similarly, a student on qualitative track obtains no points from Math 2 section.

tutoring, among other student traits. While this data is not used in determining if a student is admitted or not to a university, it does allow us to control for individual characteristics that could potentially impact a student's unwillingness to guess.

3 Sample Selection and Summary Statistics

In Turkey, female students have been catching up and even outperforming in educational outcomes. This trend is identical to that of many other countries. However, as found in the companion paper [Saygin, 2019], female students do worse on multiple choice exams than they do on essay type exams when compared to their male counterparts. This finding is consistent with examination contexts in other countries as suggested by several studies [Hirschfeld et al., 1995, Walstad and Robson, 1997, Duckworth and Seligman, 2005, Breda and Ly, 2015, Montolio and Taberner, 2018].

From Table 3, it is apparent that female students in Turkey have higher average high school GPAs and test scores than their male peers and retake the ÖSS at a lower rate than their male peers. However, at the upper end of the scoring distribution of the ÖSS, this higher average female performance disappears once conditioned on high school GPA. As a result of this, female applicants' college placement is not as strong as it otherwise would be [Saygin, 2019].

As participants self-select into retaking the exam, we focus solely on first-time test takers. Retakers are systematically different than the first-time takers[9] and omitting them from the sample will reduce confounding factors. Panel B of Table 3 shows summary statistics for high school GPA, test scores and share

[9] For example, the gender ratio of first-time test takers is different than that of non-first-time test takers.

of high school tracks by gender. Females are less likely to be on quantitative track and more likely to be on the equally-weighted track among the first time takers. Furthermore, to control for the possibility of different willingnesses to guess between tracks, we conduct our analysis separately on each track.

4 Unwillingness to Guess

We will first define a non-correct question as one that is either skipped or incorrect. To assess test taker guessing behavior, we will analyze the proportion of skipped questions to the total number of non-correct questions. We will call this variable of interest "unwillingness to guess" and it is calculated by the formula $\frac{Skipped}{Skipped+Incorrect}$.

As such, if an applicant has 0 incorrect answers, this ratio would be equal to 1, and we consider the applicant to be 100% unwilling to guess. For any given number of correct answers, a higher number of incorrect answers leads to lower unwillingness to guess. Any test takers that answered all questions correctly are dropped from our analysis as their unwillingness to guess is undefined.

Our critical assumption is that conditional on having guessed, all test takers have the same probability of answering a question correctly. If girls had better guessing skills than boys and this lead to a higher number of correct answers, we would falsely identify female applicants as less willing to guess. We acknowledge that our measure could be capturing these potential differences.

The most important reason we chose this ratio is a limitation of the data. The data set contains only the number of total questions, number of correct and incorrect answers in each section for each applicant. Due to this, the number of correct and incorrect answers for a given test taker not only reflects her willingness to guess but also her knowledge. In other words, the number of correct and incorrect answers includes answers that were guessed. If we were to focus solely on the number of skipped questions, any observed gender gap could reflect the differences in knowledge and willingness to guess. One potential solution would be to analyze the gender gap in the number of skipped questions conditional on knowledge proxied by the number of correct answers. However, since the number of correct answers could reflect the willingness to guess of a test taker, conditioning on this confounds the analysis rather than helps the identification of gender differences in guessing.

We acknowledge that applicants might have different guessing skills based on their ability to rule out some incorrect answers. Our heterogeneity analysis assumes that the guessing skills might be different across genders but that gender difference in guessing skills are identical across different sections. With this assumption, we estimate overall student guessing aversion differentials across genders by way of sections' varying effective guessing penalty and difficulty levels.

Using the "unwillingness to guess" measure has a key advantage and disadvantage when compared to other possible measures. Given the limitations of the data, we consider the number of correct answers to be a function of test taker knowledge and willingness to guess. As such, a student with partial knowledge on a question's topic is more likely to guess on that question than a student that does not possess prior knowledge. Therefore, it is beneficial to use a measurement that omits correct answers as we wish to assess test taker willingness to guess and not knowledge. However, the share of skipped questions in skipped and incorrect answers captures time management abilities and simple mistakes in arithmetic and reading. As such, if female test takers are worse at time management than male test takers, then that would be captured here. Similarly, if male test takers are more likely to make mistakes, then the results would incorrectly suggest differences in willingness to guess as the source of differentials even if they are in fact as unwilling to guess as their female peers. In the results section, we provide some robustness checks to discuss these possibilities.

Our empirical model is a reduced form model that estimates the overall gender difference in willingness to guess. For applicant i from school type h in city c the model is given by:

$$N_{ihc} = \alpha + \delta F_i + x_i'\beta + \mu_h + \mu_C + \epsilon_{ihc}$$

where N_{ihc} is the share of skipped questions out of all non-correct answers in a given section of the test and ϵ_{ihc} is a random error term. Furthermore, F_i is a dummy variable for gender and X_i' is a vector containing individual student characteristics such as high school GPA, parental education, and if they received private tutoring in preparation for the test. The hypothesis to be tested is $\delta > 0$ and whether it varies across sections of the test.

5 Results

5.1 Willingness to Guess

We first analyze gender differences in unwillingness to guess across all of an applicant's relevant sections. To do so, we take sub-samples of the first time test takers from the equally-weighted, qualitative, and quantitative high school tracks and analyze them separately. We calculate the share of skipped questions out of non-correct ones using the four most relevant sections of the ÖSS for each track. For example, we calculate this value for the applicants of the equally-weighted track using the Math 1, Math 2, Literature 1, and Literature 2 sections. For quantitative track students, 0.41 of the non-correct questions were skipped, with a standard deviation of 0.24. In the qualitative and equally-weighted tracks the averages are 0.21 and 0.58 with standard deviations of 0.21

and 0.22 respectively. We provide both the summary statistics for the share of skipped questions within each section and as a total of the relevant sections with a gender comparison in Table 4.

Table 5 shows the results of our estimations on the sub-samples of all three tracks. While we do not find a statistically significant gender difference for equally-weighted and qualitative track students in the share of skipped questions in the most relevant sections, we find that female applicants in the quantitative track skip 6.5 percentage points more questions in the most relevant sections[10] of the exam compared to their male peers. We obtained these estimates while controlling for high school GPA, the receipt of private tutoring, parental education, and employment status, in addition to fixed effects for high school type and city. This set of control variables is used for the following specifications throughout the paper.

For the rest of the detailed analysis, we restrict the sample to only those on the quantitative high school track[11]. This particular track is valuable for our analysis as it appears to have the highest gender gap in overall willingness to guess. This is significant for two reasons. The gender scoring gap is traditionally largest on mathematics exams, and as such this particular track has implications to the larger literature. Furthermore, as female students are less likely to select into the quantitative track, we expect a positive selection of females. As such, finding a gender gap despite this positive selection makes the results stronger.

Table 6 provides the estimates for the gender coefficients in the estimation of the share of skipped questions for all sections separately for the sub-sample of first-time taker quantitative track students. As is reported in columns 1 and 5, the gender gap seems to be largest in the math sections. Notably, the gap is larger on the harder math section. The gender gap is statistically significant

[10]These sections are the Math 1, Math 2, Science 1, and Science 2 sections.

[11]While the gender gap is mostly insignificant for the equally-weighted and qualitative tracks, the results on the sub-samples of these tracks are available upon request.

and positive in all sections except for the Literature 1 section, where the gender gap is insignificant and negative. That is, female test takers skip less than male test takers on this section, but to a statistically insignificant extent. While the Social Science 1 section has the lowest weight for quantitative track score calculation, the gender gap on it is similar to the other non-Literature sections. While the math sections have the same weight, female applicants tend to skip more on the Math 2 section which contains more complex problems than the Math 1 section. These results suggests that gender gaps in willingness to guess may not only results from risk aversion but also test subject and difficulty level.

In order to look at the heterogeneity in gender differences in guessing across sections more closely, we study the gender gap in each section separately conditional on guessing behavior in other sections. We first start with the unwillingness to guess in the Math 1 section. This section is one of the most important sections for quantitative track test takers. The first column of Table 7 shows that the share of skipped questions is 6.5 percentage points higher for female test takers in the Math 1 section controlling for high school GPA, the receipt of private tutoring, parental education, and employment status, as well as fixed effects for high school type and city. In the second column, we add self-assessment measures into the estimations and this seems to lower the gender gap to 5.1 percentage points. We find that students who find themselves being good at math tend to skip less questions on this section with the self-assessments for other subjects not affecting the unwillingness to guess in the Math 1 section. In the third column, we include the share of skipped questions in the Social Science 1 and Literature 1 sections. These sections are less important to quantitative track scores but are still relevant. This column shows that guessing behavior is correlated between sections. That is, the share of skipped questions in the Literature 1 and Social Science 1 sections are positively related with the share of skipped questions in the Math 1 section. However, controlling for guessing behavior in the less important sections does not seem to affect the gender gap in unwillingness to guess in Math 1 section. In the fourth column, we also add the share of skipped questions in the Science 1 section which is a relatively more important section than the Literature 1 and Social Science 1 sections but slightly less important than the Math 1 section. We can see that the correlation between guessing behavior in the Science 1 and Math 1 sections seems to be stronger than the correlation between the Math 1 section and the Social Sciences 1 and Literature 1 sections. Controlling for the share of skipped questions in the Science 1 section lowers our estimate of the gender gap in share of skipped questions in Math 1 to 4.7 percentage points. It seems that even after controlling for the guessing behavior in other sections, a large part of the gender gap remains in Math 1 section. In the last column, the gender gap becomes insignificant when controlling for the guessing behavior in the Math 2 section which is equally important to the Math 1 section but covers more difficult questions.

In Table 8, we run similar regressions for the Science 1 section as we did for the Math 1 section. Column 1 shows that female applicants skip 5.2 percentage points more in the Science 1 section and that this gender gap goes down to 3.6 percentage points while controlling for self-assessment. The self-assessments in science and math seem to be negatively associated with the share of skipped questions while the self-assessments in social science and literature are not associated with the share of skipped questions in Science 1 section. Interestingly, the self-assessment in math seems to have a slightly stronger impact on skipped questions in the Science 1 section than the self-assessment in science. Similarly, as seen in column 3, guessing behavior in the Literature 1 and Social Science 1 sections are positively associated with the guessing behavior in the Science 1 section. However, the gender gap remains largely unaffected while controlling for them. The gender gap goes down to 2.9 percentage points when we control for the skipped questions in Math 1 section. This control scheme is interesting to us as the Math 1 section has the highest importance for the quantitative score calculations. As such, that an unexplained gender difference in the share of skipped questions in the Science 1 section which remains after controlling for guessing behavior in the Math 1 section is noteworthy as this provides evidence that the gender gap in guessing behavior is not solely driven by differences in risk aversion. The gender gap disappears when we further control for the guessing behavior in the Science 2 section.

Table 9 provides us with the results for the gender gap in unwillingness to guess on the Social Science 1 section across various specifications. Column 1 shows that female test takers skip 5.9 percentage points more in the Social Science 1 section than their male peers. This gender gap is reduced to 4.5 percentage points when we control for self-assessment variables. It appears that test takers considering themselves as being successful in the social sciences is negatively associated with the share of skipped questions. Notably, self-assessment in math also has an impact on the guessing behavior in the Social Science 1 section while the self-assessments in science and literature do not appear to be related. Our results reported in the third column show that controlling for the share of skipped questions in the Literature 1 and Science 1 sections, which are significantly more important in quantitative score calculations than the Social Science 1 section, do not seem to affect the gender gap in unwillingness to guess in the Social Science 1 section. In other words, even when we control for the guessing behavior in these higher-stakes sections, female test takers still tend to skip more questions than their male counterparts in the Social Science 1 section. Note that our sample for these regressions contains only the quantitative track students who are taking the exam for the first time. Given that they are maximizing their quantitative scores, the math and science sections are the highest stakes sections. This provides further confirmation that the gender differences in guessing behavior is driven by additional mechanisms other than risk aversion. Gender gap in unwillingness to guess loses significance only when we control for the share of skipped questions in Math 1 section.

Table 10 reports the results for Literature section which are quite different than the results for other sections. At first glance, it appears that there is no significant gender gap in the share of skipped questions in Literature 1 section. On the other hand, we find that male applicants seem to skip more questions in the Literature 1 section once we condition on the share of skipped questions in the Social Science 1, Science 1, and Math 1 sections. As before, this result suggests that risk aversion alone does not explain the gender differences in guessing behavior.

Our findings so far suggest that the gender gap in willingness to guess is not only due to gender differences in risk aversion. We find that there is still some gender gap in willingness to guess even while controlling for the willingness to guess in other sections of different importance to the score calculations. Yet, we can not distinguish whether this finding is caused by the heterogeneity across subjects. We find that female test takers are more willing to guess than male test takers in literature while they remain hesitant when it comes to the math, science, and social science sections.

Our next step is to analyze the sections that are equally important to the most important sections discussed previously but which cover more sophisticated material. This part of the analysis allows us to study gender differences in willingness to guess in more difficult sections while controlling for less difficult but equally important sections covering the same subject. As the more sophisticated social science and literature sections are not factored into quantitative scores, this analysis will look solely at the Math 2 and Science 2 sections.

We will first focus on the Science 2 section. This section is equally important to the Science 1 section for quantitative track score calculations but its questions cover what the test makers believe to be more difficult material. In Table 11, the first column shows that the gender gap is equal to 6.2 percentage points which is slightly higher than the gender gap in willingness to guess in the Science 1 section. The second column shows that controlling for self-assessment reduces the gender gap to 3.9 percentage points. For this regression, the self-assessments in math and science have significant coefficients. The gender gap remains unaffected while controlling for the share of skipped questions in the Literature 1 and Social Science 1 sections, but by additionally controlling for the share of skipped questions in the Science 1 section it is reduced to 3.7 percentage points. That is, there remains some gender differences in the share of skipped questions in the more difficult Science 2 section even while controlling for the guessing behavior in other sections that are either less important or equally important and less difficult. In the last two columns, this remaining gap disappears once we add in self-assessment variables.

Lastly, we analyze the gender gap in the share of skipped questions in the Math 2 section in Table 12. The first column shows that female test takers skip 9.1 percentage points more than their male counterparts. This gap goes down to 7.3 percentage points after controlling for the self assessment variables. Of these variables, it appears that only the self-assessment in math is significantly related to the guessing behavior on this section. Introducing controls for the share of skipped questions in the Literature and Social Science 1 sections reduces the gender gap only slightly. However, introducing the share of skipped questions in the Science 1 section reduces it to 6.1 percentage points. In the fifth column, we also control for the willingness to guess in the Math 1 section. After controlling for this, we find a remaining gender gap of 0.039. That is, almost half of the gender gap remains unexplained in the more difficult math section after we condition on all other sections.

In column 6 of Table 12, we restrict the sample to the applicants who answered at least one question from the social science section to rule out the possibility that we might be capturing gender differences in time management rather than willingness to guess. If a student has the time to answer a question in the Social Science 1 section, we assume that they must have completed the math sections as they are the most important sections for the test takers who were on the quantitative track. We find very similar estimates with this sample which provides suggestive evidence that our results are not driven by gender differences in time management.

Math is the only subject where there is a remaining gender gap in willingness to guess even after controlling for the share of skipped questions in all other sections. This raises the question of if this is also true of applicants who are in the equally-weighted high school track (which emphasizes math, literature, and social science). We provide results for the share of skipped questions in the Math 1 and Math 2 sections in Table 13 and Table 14 respectively for these students. We find no evidence of a gender gap in willingness to guess in the math sections among equally-weighted track students.

5.2 Gender Differences in Self-Assessment

Our results show that there are other factors than risk aversion that lead to a gender gap in guessing on multiple choice tests. We provide suggestive evidence that gender differences in self-assessment partly explain the gender gap in willingness to guess. We report the gender differences in reported self-assessments in Table 15. We find that female applicants are less likely to report that they find themselves good at math, science, and social science than their male peers. This phenomenon is especially pronounced among quantitative track applicants as these gender differences in reported self-assessment are not present among applicants from other tracks. In all tracks, females seem to be more likely to report that they are good at literature.

One interesting finding from the previous section is that self-assessment in math is significantly related to the share of skipped questions in all non-literature sections. Within the literature sections, the gender gap in the share of skipped questions reverses.

In order to understand the gender differences in self-assessment further, we estimated a linear probability model of reporting that they find themselves good at certain subjects conditional on the correct answers in the test in the corresponding subjects. Table 16 shows that females are 8.6%, 8%, and 5.5% less likely to report that they find themselves successful in math, science, and social science respectively after conditioning on the number of correct answers in the corresponding sections. On the other hand, females are 12.4% more likely to report that they think they are good at Literature after controlling for the correct answers in Literature sections[12]. These results suggest that males are more likely to be overly confident in math, science, and social Sciences while females are more likely to be overly confident in literature. These results are consistent with the pattern of the gender differences in willingness to guess across subjects. Moreover, females might be less likely to report themselves as good at a subject, even though they think they are good at it. If self-assessment could be measured without this potential bias, then controlling for self-assessment might have explained a larger part of gender differences in guessing.

[12] We also add other controls as described in the previous estimations.

6 Conclusion

Multiple past studies have found that male students are more likely to guess on standardized exams than female students. Our paper adds to this literature by examining how gender differences in willingness to guess differ across subjects, difficulty levels, and stakes using administrative data from Turkey. Through variation in section weights imposed by test taker high school track, we were able to effectively vary the stakes of the exam between sections.

This allowed us to find that not only is the gender difference in willingness to guess largest on sections that cover mathematics, but that this difference is only significant for the test takers on the track that places the most emphasis on these sections. This is notable as we would expect the female students on this track to be positively selected toward an affinity for mathematics. Across all non-literature sections for this track, female students were less willing to guess than their male peers.

Furthermore, by using a supplemental survey answered by the test takers within our data set, we were able to proxy self confidence for the subjects tested by the exam. We found that a positive self assessment on a given subject was related to guessing behavior on that subject. However, for quantitative track test takers, the mathematics self assessment question was found to be related to guessing behavior on other non-math sections of the exam. In addition to this, we provide evidence for gender differences in self assessment conditional on the exam results. Male test takers were more confident than their exam results would suggest in math, science, and social science while female test takers were more confident in their literature abilities.

References

S. Pelin Akyol, James Key, and Kala Krishna. Hit or miss? test taking behavior in multiple choice exams. *National Bureau of Economic Research Working Papers*, (22401), July 2016. doi: 10.3386/w22401. URL `http://www.nber.org/papers/w22401`.

Ghazala Azmat, Caterina Calsamiglia, and Nagore Iriberri. Gender differences in response to big stakes. *Journal of the European Economic Association*, 14 (6):1372–1400, 2016. doi: 10.1111/jeea.12180. URL `http://dx.doi.org/10.1111/jeea.12180`.

Katherine Baldiga. Gender differences in willingness to guess. *Management Science*, 60(2):434–448, 2014. doi: 10.1287/mnsc.2013.1776. URL `https://doi.org/10.1287/mnsc.2013.1776`.

Brad M. Barber and Terrance Odean. Boys will be Boys: Gender, Overconfidence, and Common Stock Investment*. *The Quarterly Journal of Economics*, 116(1):261–292, 02 2001. ISSN 0033-5533. doi: 10.1162/003355301556400. URL `https://doi.org/10.1162/003355301556400`.

Sylvia Beyer. Gender differences in the accuracy of grade expectancies and evaluations. *Sex Roles*, 41, 08 1999. doi: 10.1023/A:1018810430105.

Thomas Breda and Son Thierry Ly. Professors in core science fields are not always biased against women: Evidence from france. *American Economic Journal: Applied Economics*, 7(4):53–75, October 2015. doi: 10.1257/app.20140022. URL `http://www.aeaweb.org/articles?id=10.1257/app.20140022`.

Katherine Baldiga Coffman and David Klinowski. The Impact of Penalties for

Wrong Answers on the Gender Gap in Test Scores. *Harvard Business School Working Papers*, 09 2020.

Rachel Croson and Uri Gneezy. Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74, June 2009. doi: 10.1257/jel.47.2.448. URL `http://www.aeaweb.org/articles?id=10.1257/jel.47.2.448`.

Angela L. Duckworth and Martin E. P. Seligman. Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological Science (0956-7976)*, 16(12):939 – 944, 2005. ISSN 09567976. URL `http://lp.hscl.ufl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=buh&AN=18943661&site=eds-live`.

Avraham Ebenstein, Victor Lavy, and Sefi Roth. The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4):36–65, October 2016. doi: 10.1257/app.20150213. URL `http://www.aeaweb.org/articles?id=10.1257/app.20150213`.

Catherine C. Eckel and Philip J. Grossman. Men, women and risk aversion: Experimental evidence. 1:1061 – 1073, 2008. ISSN 1574-0722. doi: https://doi.org/10.1016/S1574-0722(07)00113-8. URL `http://www.sciencedirect.com/science/article/pii/S1574072207001138`.

María Paz Espinosa and Javier Gardeazabal. Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, 54(5): 415 – 425, 2010. ISSN 0022-2496. doi: https://doi.org/10.1016/j.jmp.2010.06.001. URL `http://www.sciencedirect.com/science/article/pii/S0022249610000623`.

María Paz Espinosa and Javier Gardeazabal. Do Students Behave Rationally in Multiple Choice Tests? Evidence from a Field Experiment. *Journal of*

Economics and Management, 9(2):107–135, July 2013. URL https://ideas.repec.org/a/jec/journl/v9y2013i2p107-135.html.

Antonio Filippin and Paolo Crosetto. A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11):3138–3160, 2016. doi: 10.1287/mnsc.2015.2294. URL https://doi.org/10.1287/mnsc.2015.2294.

Patricia Funk and Helena Perrone. Gender differences in academic performance: The role of negative marking in multiple-choice exams. *Center for Economic and Policy Research Working Paper*, 12 2016.

Ben-Shakhar Gershon and Sinai Yakov. Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement*, 28(1):23–35. doi: 10.1111/j.1745-3984.1991.tb00341.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3984.1991.tb00341.x.

Mary Hirschfeld, Robert L. Moore, and Eleanor Brown. Exploring the gender gap on the gre subject test in economics. *The Journal of Economic Education*, 26(1):3–15, 1995. doi: 10.1080/00220485.1995.10844852. URL https://www.tandfonline.com/doi/abs/10.1080/00220485.1995.10844852.

Nagore Iriberri and Pedro Rey-Biel. Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. *Center for Economic and Policy Research Discussion Paper*, 2019. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3341348.

Stepan Jurajda and Daniel Munich. Gender gap in performance under competitive pressure: Admissions to czech universities. *The American Economic Review*, 101(3):514–518, 2011. ISSN 00028282. URL http://www.jstor.org/stable/29783799.

Daniel Montolio and Pere A. Taberner. Gender differences under test pressure and their impact on academic performance: A quasi-experimental design. *Institut d'Economia de Barcelona Working Paper*, 2018.

P.Karen Murphy and Patricia A. Alexander. A motivated exploration of motivation terminology. *Contemporary Educational Psychology*, 25(1):3 – 53, 2000. ISSN 0361-476X. doi: https://doi.org/10.1006/ceps.1999.1019. URL http://www.sciencedirect.com/science/article/pii/S0361476X99910196.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much. *The Quarterly Journal of Economics*, 122:1067–1101, 08 2007. doi: 10.1162/qjec.122.3.1067.

Muriel Niederle and Lise Vesterlund. Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2): 129–44, June 2010. doi: 10.1257/jep.24.2.129. URL http://www.aeaweb.org/articles?id=10.1257/jep.24.2.129.

Evren Ors, Frédéric Palomino, and Eloïc Peyrache. Performance gender gap: Does competition matter? *Journal of Labor Economics*, 31(3):443–499, 2013. doi: 10.1086/669331. URL https://doi.org/10.1086/669331.

Tuomas Pekkarinen. Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior and Organization*, 115:94 – 110, 2015. ISSN 0167-2681. doi: https://doi.org/10.1016/j.jebo.2014.08.007. URL http://www.sciencedirect.com/science/article/pii/S0167268114002261.

Lara Perez-Felkner, Samantha Nix, and Kirby Thomas. Gendered pathways: How mathematics ability beliefs shape secondary and postsecondary course and degree field choices. *Frontiers in Psychology*, 8:386, 2017. ISSN 1664-

1078. doi: 10.3389/fpsyg.2017.00386. URL https://www.frontiersin.org/article/10.3389/fpsyg.2017.00386.

Gerhard Riener and Valentin Wagner. Gender differences in willingness to compete and answering multiple-choice questions? The role of age. *Economics Letters*, 164(C):86–89, 2018. doi: 10.1016/j.econlet.2018.01. URL https://ideas.repec.org/a/eee/ecolet/v164y2018icp86-89.html.

Fernando Saltiel. What's math got to do with it? Multidimensional ability and the gender gap in STEM. (1201), 2019. URL https://ideas.repec.org/p/red/sed019/1201.html.

Perihan Saygin. Gender bias in standardized tests: evidence from a centralized college admissions system. *Empirical Economics*, 03 2019. doi: 10.1007/s00181-019-01662-z.

William B. Walstad and Denise Robson. Differential Item Functioning and Male-Female Differences on Multiple-Choice Tests in Economics. *The Journal of Economic Education*, 28(2):155–171, June 1997. doi: 10.1080/00220489709595917. URL https://ideas.repec.org/a/taf/jeduce/v28y1997i2px155-171.html.

7 Figures and Tables

Table 1: Differences in Rankings for 1 Point Increase in EW1 Test Score

Percentiles	Score	Change in Ranking
10	177	13,390
25	194	10,971
50	215	6,654
90	255	2,338
95	265	1,689

Source: OSYM 2008 Test Score Distributions

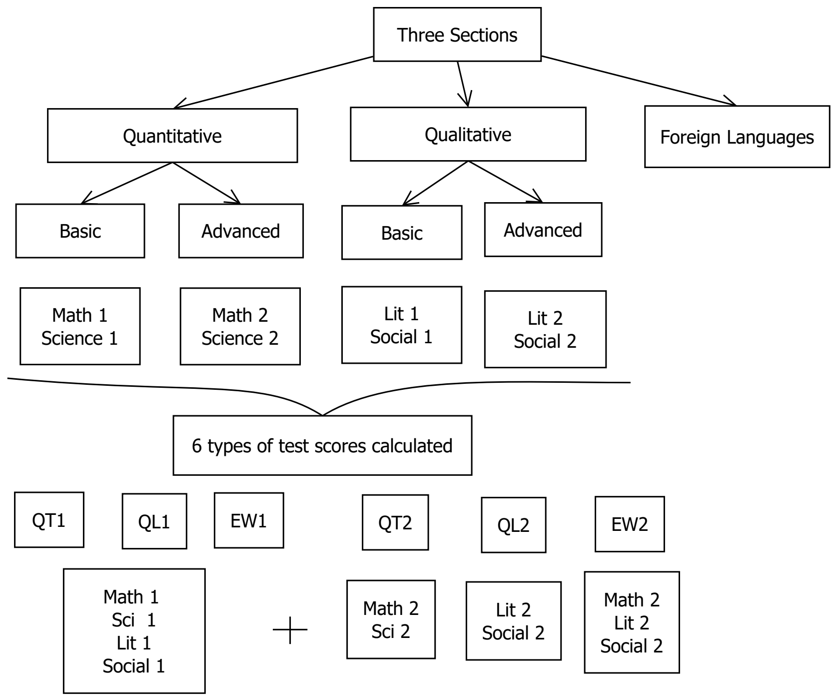
Table 2: Coefficients by Sections for All Test Scores

Scores	Lit 1	Social 1	Math 1	Science 1	Lit 2	Social 2	Math2	Science 2
Qualitative 1	1.0	0.7	0.3	0.2	-	-	-	-
Quantitative 1	0.3	0.2	1.0	0.7	-	-	-	-
Equally Weighted 1	0.8	0.3	0.9	0.2	-	-	-	-
Qualitative 2	0.5	0.35	0.3	0.2	0.5	0.35	-	-
Quantitative 2	0.3	0.2	0.5	0.35	-	-	0.5	0.35
Equally Weighted 2	0.4	0.3	0.45	0.2	0.4	-	0.45	-

Source: OSYM 2008 Application Manual

Note: The university entrance test has two main sections, Quantitative and Qualitative, in addition to a foreign language section. The main sections each have two subsections consisting of the following fields: Social Sciences (history, geography and philosophy), Science (Biology, Chemistry and Physics), Mathematics, and Literature. Each section of the test counts toward the final scores with different coefficients depending on the type of score as listed above.

Figure 1: University Entrance Test: Sections and Categories



Note: The university entrance test has two main sections, Quantitative and Qualitative, in addition to a foreign language section. These sections each have two subsections: basic and advanced. Quantitative section includes Science (Biology, Chemistry and Physics) and Mathematics whereas Qualitative section includes Social Sciences (history, geography and philosophy) and Literature. The Foreign Language section is an additional test. Regardless a student's choice of subject in high school, each student answers essentially 4 sections which are Literature 1, Social Sciences 1, Mathematics 1, and Science 1. Based on these sections Quantitative 1, Qualitative 1, and Equally Weighted 1 test scores are calculated for each student. Advanced sections with its subsections such as Literature 2, Social Sciences 2, Mathematics 2, and Science 2 are more sophisticated requiring more detailed knowledge in these subjects. These advanced sections are only relevant for subject specific test scores such as Quantitative 2, Qualitative 2, and Equally Weighted 2. Therefore, students choose to answer the questions of the subsections that pertain to their high school subject. For example, a student who wants to maximize her Quantitative 2 test score, she would answer questions from Math 2 and Science 2 sections in addition to 4 basic sections.

Table 3: Descriptive Statistics by Gender

	Female	Male	Gap
Panel A: All Sample			
High School GPA	76.53	72.03	4.50***
	(11.21)	(11.58)	
Standardized High School GPAs			
Equally Weighted High School GPA	84.69	80.89	3.79***
	(8.69)	(9.44)	
Quantitative High School GPA	82.44	78.53	3.91***
	(9.63)	(10.26)	
Qualitative High School GPA	85.89	82.23	3.66***
	(7.89)	(8.68)	
OSS Scores			
Test Score Equally Weighted 1	212.55	206.03	6.53***
	(35.90)	(42.80)	
Test Score Quantitative 1	188.20	188.75	-0.55
	(38.71)	(45.26)	
Test Score Qualitative 1	219.11	209.58	9.53***
	(34.24)	(42.05)	
Test Score Equally Weighted 2	153.68	145.22	8.46***
	(83.63)	(86.58)	
Test Score Quantitative 2	111.46	106.15	5.31*
	(98.32)	(100.30)	
Test Score Qualitative 2	111.57	96.25	15.32***
	(101.90)	(101.46)	
High School Tracks			
Equally-Weighted Track	0.48	0.43	0.05***
	(0.50)	(0.49)	
Quantitative Track	0.29	0.34	-0.05***
	(0.46)	(0.47)	
Qualitative Track	0.12	0.14	-0.02***
	(0.32)	(0.35)	
OSS exam retake	0.78	0.84	-0.06***
	(0.41)	(0.37)	
Observations	3544	6439	9983
Panel B: Only First-Takers Sample			
High School GPA	78.85	72.44	6.40***
	(12.66)	(13.61)	
Standardized High School GPAs			
Equally Weighted High School GPA	89.24	84.96	4.28***
	(8.82)	(10.83)	
Quantitative High School GPA	87.12	82.58	4.55***
	(10.06)	(11.84)	
Qualitative High School GPA	90.00	85.90	4.10***
	(7.96)	(9.90)	
OSS Scores			
Test Score Equally Weighted 1	226.25	215.98	10.26***
	(37.92)	(50.79)	
Test Score Quantitative 1	204.08	201.86	2.22
	(41.82)	(53.99)	
Test Score Qualitative 1	230.16	217.15	13.01***
	(35.38)	(48.30)	
Test Score Equally Weighted 2	160.89	150.27	10.62*
	(92.75)	(95.40)	
Test Score Quantitative 2	133.93	127.82	6.12
	(104.54)	(109.53)	
Test Score Qualitative 2	93.62	71.46	22.16***
	(107.29)	(99.37)	
High School Tracks			
Equally-Weighted Track	0.41	0.34	0.07**
	(0.49)	(0.47)	
Quantitative Track	0.36	0.41	-0.05*
	(0.48)	(0.49)	
Qualitative Track	0.10	0.11	-0.01
	(0.30)	(0.31)	
Observations	768	1024	1792

Notes: Panel A shows the summary statistics for the whole sample whereas Panel B reports the summary statistics for the first time takers. Standard deviations for females and males are in parentheses.

Table 4: Share of Skipped Questions: Only First takers - All Tracks

	Female Mean/sd	Male Mean/sd	Gap b	Female Mean/sd	Male Mean/sd	Gap b	Female Mean/sd	Male Mean/sd	Gap b
Sh. of Skipped in Math 1	0.50	0.44	0.06**	0.82	0.76	0.06	0.64	0.63	0.01
	(0.26)	(0.29)		(0.22)	(0.30)		(0.25)	(0.29)	
Sh. of Skipped in Science 1	0.26	0.21	0.04*	0.92	0.84	0.07	0.94	0.89	0.05**
	(0.23)	(0.25)		(0.22)	(0.31)		(0.16)	(0.25)	
Sh. of Skipped in Math 2	0.60	0.53	0.07***	0.99	0.93	0.05*	0.77	0.78	-0.01
	(0.26)	(0.29)		(0.10)	(0.22)		(0.24)	(0.30)	
Sh. of Skipped in Science 2	0.37	0.33	0.03	1.00	0.95	0.05**	1.00	0.97	0.02**
	(0.31)	(0.32)		(0.00)	(0.21)		(0.05)	(0.14)	
Sh. of Skipped in Literature 1	0.20	0.22	-0.02	0.14	0.16	-0.02	0.14	0.15	-0.01
	(0.25)	(0.25)		(0.17)	(0.21)		(0.21)	(0.22)	
Sh. of Skipped in Social 1	0.50	0.44	0.06*	0.14	0.12	0.02	0.13	0.12	0.00
	(0.33)	(0.37)		(0.15)	(0.20)		(0.15)	(0.19)	
Sh. of Skipped in Literature 2	0.99	0.98	0.01	0.15	0.26	-0.11**	0.18	0.25	-0.07***
	(0.09)	(0.12)		(0.17)	(0.30)		(0.21)	(0.29)	
Sh. of Skipped in Social 2	1.00	1.00	0.00	0.27	0.29	-0.01	0.75	0.64	0.11***
	(0.01)	(0.06)		(0.20)	(0.31)		(0.35)	(0.41)	
Total Share of Skipped - QT	0.46	0.40	0.05**						
	(0.22)	(0.24)							
Most Relevant Share of Skipped - QT	0.44	0.39	0.05**						
	(0.22)	(0.24)							
Total Share of Skipped - QL				0.58	0.53	0.05			
				(0.17)	(0.21)				
Most Relevant Share of Skipped - QL				0.19	0.23	-0.03			
				(0.15)	(0.24)				
Total Share of Skipped - EW							0.66	0.63	0.03*
							(0.14)	(0.19)	
Most Relevant Share of Skipped - EW							0.58	0.57	0.01
							(0.20)	(0.24)	
Observations	280	421	701	74	114	188	313	348	661

Notes: Standard deviations are in parentheses.

Table 5: Share of Skipped in All Sections Excl. Less Relevant Sections

	(Equally-Weighted)	(Qualitative)	(Quantitative)
Female	0.025	-0.042	0.065***
	(0.021)	(0.029)	(0.016)
Equally Weighted High School GPA	-0.000		
	(0.002)		
Qualitative High School GPA		-0.000	
		(0.002)	
Quantitative High School GPA			-0.009***
			(0.001)
Observations	661	188	701

Standard errors in parentheses
* p¡0.10, ** p¡0.05, *** p¡0.010

Table 6: Share of Skipped: QT Students

	(Math)	(Sci.)	(Lit.)	(Soc.)	(Math2)	(Sci2)
Female	0.065***	0.052***	-0.028	0.059**	0.091***	0.062***
	(0.023)	(0.017)	(0.023)	(0.023)	(0.019)	(0.021)
Observations	676	692	692	701	693	700

Standard errors in parentheses
* p¡0.10, ** p¡0.05, *** p¡0.010

Table 7: Share of Skipped: Math 1 - QT Students

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.065***	0.051**	0.063***	0.047**	0.040*	0.022
	(0.023)	(0.024)	(0.022)	(0.022)	(0.024)	(0.018)
Successful Math		-0.099***			-0.064***	
		(0.021)			(0.021)	
Successful in Science		0.010			0.011	
		(0.033)			(0.028)	
Successful in Literature		0.017			0.016	
		(0.025)			(0.024)	
Successful Social Sciences		0.018			0.038	
		(0.039)			(0.039)	
Sh. of Skipped in Social 1			0.157***	0.124***	0.123***	0.073**
			(0.042)	(0.039)	(0.041)	(0.030)
Sh. of Skipped in Literature 1			0.333***	0.247***	0.244***	0.207***
			(0.041)	(0.044)	(0.045)	(0.042)
Sh. of Skipped in Science 1				0.345***	0.332***	0.172***
				(0.039)	(0.038)	(0.036)
Sh. of Skipped in Math 2						0.415***
						(0.048)
Observations	676	676	669	662	662	659

Standard errors in parentheses
* p¡0.10, ** p¡0.05, *** p¡0.010

Table 8: Share of Skipped: Science 1 - QT Students

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.052***	0.036**	0.048***	0.029*	0.020	0.016
	(0.017)	(0.018)	(0.018)	(0.017)	(0.018)	(0.016)
Successful Math		-0.057***			-0.019	
		(0.018)			(0.020)	
Successful in Science		-0.048**			-0.049**	
		(0.022)			(0.023)	
Successful in Literature		0.014			0.015	
		(0.019)			(0.018)	
Successful Social Sciences		-0.007			-0.015	
		(0.032)			(0.035)	
Sh. of Skipped in Literature 1			0.241***	0.160***	0.170***	0.109***
			(0.034)	(0.034)	(0.035)	(0.033)
Sh. of Skipped in Social 1			0.103***	0.067***	0.060**	0.032
			(0.027)	(0.025)	(0.025)	(0.021)
Sh. of Skipped in Math 1				0.248***	0.239***	0.159***
				(0.033)	(0.033)	(0.029)
Sh. of Skipped in Science 2						0.314***
						(0.035)
Observations	692	692	683	662	662	662

Standard errors in parentheses
* p¡0.10, ** p¡0.05, *** p¡0.010

Table 9: Share of Skipped: Social 1 - QT Students

	(1)	(2)	(3)	(4)	(5)
Female	0.059**	0.045*	0.054**	0.040	0.031
	(0.023)	(0.024)	(0.023)	(0.026)	(0.026)
Successful Math		-0.069**			-0.028
		(0.028)			(0.025)
Successful in Science		-0.006			-0.005
		(0.032)			(0.025)
Successful in Literature		-0.010			0.010
		(0.025)			(0.024)
Successful Social Sciences		-0.081**			-0.107***
		(0.039)			(0.034)
Sh. of Skipped in Literature 1			0.413***	0.344***	0.354***
			(0.055)	(0.058)	(0.056)
Sh. of Skipped in Science 1			0.191***	0.135**	0.122**
			(0.053)	(0.054)	(0.055)
Sh. of Skipped in Math 1				0.179***	0.178***
				(0.054)	(0.059)
Observations	701	701	683	662	662

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010

Table 10: Share of Skipped: Literature 1 - QT Students

	(1)	(2)	(3)	(4)	(5)
Female	-0.028	-0.022	-0.052**	-0.058***	-0.045**
	(0.023)	(0.020)	(0.020)	(0.020)	(0.018)
Successful Math		-0.036			0.009
		(0.029)			(0.026)
Successful in Science		0.024			0.034
		(0.026)			(0.022)
Successful in Literature		-0.043**			-0.048**
		(0.019)			(0.020)
Successful Social Sciences		0.040			0.058*
		(0.032)			(0.033)
Sh. of Skipped in Social 1			0.228***	0.186***	0.191***
			(0.036)	(0.035)	(0.033)
Sh. of Skipped in Science 1			0.246***	0.174***	0.185***
			(0.032)	(0.035)	(0.035)
Sh. of Skipped in Math 1				0.193***	0.191***
				(0.039)	(0.039)
Observations	692	692	683	662	662

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010

Table 11: Share of Skipped: Science 2 - QT Students

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.062***	0.039*	0.060**	0.037*	0.029	0.009
	(0.021)	(0.023)	(0.024)	(0.021)	(0.020)	(0.019)
Successful Math		-0.071***				
		(0.022)				
Successful in Science		-0.076***				
		(0.021)				
Successful in Literature		0.022				
		(0.026)				
Successful Social Sciences		-0.013				
		(0.044)				
Sh. of Skipped in Literature 1			0.260***	0.138***	0.089**	0.097***
			(0.046)	(0.041)	(0.036)	(0.035)
Sh. of Skipped in Social 1			0.162***	0.110***	0.081*	0.033
			(0.046)	(0.040)	(0.041)	(0.037)
Sh. of Skipped in Science 1				0.499***	0.462***	0.321***
				(0.055)	(0.057)	(0.056)
Sh. of Skipped in Math 2						0.501***
						(0.064)
Observations	700	700	691	682	662	659

Standard errors in parentheses
* p¡0.10, ** p¡0.05, *** p¡0.010

Table 12: Share of Skipped: Math 2 - QT Students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.091***	0.073***	0.082***	0.061***	0.039**	0.032*	0.027*
	(0.019)	(0.019)	(0.019)	(0.020)	(0.017)	(0.019)	(0.016)
Successful Math		-0.098***					
		(0.020)					
Successful in Science		-0.022					
		(0.023)					
Successful in Literature		0.009					
		(0.023)					
Successful Social Sciences		-0.006					
		(0.027)					
Sh. of Skipped in Literature 1			0.180***	0.081***	-0.017	-0.028	-0.054*
			(0.035)	(0.029)	(0.028)	(0.035)	(0.028)
Sh. of Skipped in Social 1			0.197***	0.157***	0.090***	0.123***	0.057**
			(0.037)	(0.033)	(0.028)	(0.033)	(0.022)
Sh. of Skipped in Science 1				0.403***	0.279***	0.282***	0.083**
				(0.049)	(0.052)	(0.053)	(0.041)
Sh. of Skipped in Math 1					0.387***	0.390***	0.314***
					(0.049)	(0.054)	(0.043)
Sh. of Skipped in Science 2							0.425***
							(0.042)
Observations	693	693	684	677	659	600	659

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010

Table 13: Share of Skipped: Math 1 - EW Students

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.030	0.033	0.003	0.001	-0.005	-0.005
	(0.031)	(0.031)	(0.031)	(0.031)	(0.021)	(0.022)
Successful Math		-0.213***				-0.106***
		(0.039)				(0.030)
Successful in Literature		-0.020				-0.001
		(0.034)				(0.030)
Successful Social Sciences		0.042				0.018
		(0.028)				(0.024)
Sh. of Skipped in Social 1			0.217**	0.101	0.054	0.052
			(0.082)	(0.100)	(0.061)	(0.063)
Sh. of Skipped in Science 1			0.534***	0.532***	0.382***	0.379***
			(0.063)	(0.066)	(0.082)	(0.082)
Sh. of Skipped in Literature 1				0.189***	0.128***	0.134***
				(0.063)	(0.046)	(0.048)
Sh. of Skipped in Math 2					0.456***	0.426***
					(0.043)	(0.042)
Observations	659	659	658	641	641	641

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010

Table 14: Share of Skipped in Math 2 - EW Students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.029	0.034	0.010	0.013	0.019	0.019	0.021
	(0.026)	(0.025)	(0.033)	(0.032)	(0.030)	(0.021)	(0.021)
Successful Math		-0.184***					-0.078**
		(0.041)					(0.036)
Successful in Science		0.029					-0.062
		(0.073)					(0.063)
Successful in Literature		-0.031					-0.009
		(0.025)					(0.018)
Successful Social Sciences		0.039**					0.020
		(0.019)					(0.016)
Sh. of Skipped in Social 1			0.198**	0.104	0.010	-0.020	-0.024
			(0.090)	(0.111)	(0.116)	(0.073)	(0.074)
Sh. of Skipped in Science 1			0.343***	0.338***	0.321***	0.052	0.057
			(0.069)	(0.070)	(0.068)	(0.108)	(0.106)
Sh. of Skipped in Literature 1				0.134*	0.097	0.013	0.019
				(0.068)	(0.061)	(0.048)	(0.047)
Sh. of Skipped in Literature 2					0.185***	0.142***	0.143***
					(0.043)	(0.037)	(0.036)
Sh. of Skipped in Math 1						0.496***	0.470***
						(0.047)	(0.050)
Observations	660	660	659	642	637	636	636

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010

Table 15: Self-Assessment by Gender: Only First takers

	Female	Male	Gap
Quantitative Track Students			
Successful Math	0.41	0.51	-0.10**
	(0.49)	(0.50)	
Successful in Science	0.21	0.32	-0.10**
	(0.41)	(0.47)	
Successful in Literature	0.36	0.21	0.15***
	(0.48)	(0.41)	
Successful Social Sciences	0.10	0.16	-0.06*
	(0.30)	(0.37)	
Observations	280	421	701
Equally-Weighted Track Students			
Successful Math	0.12	0.09	0.03
	(0.33)	(0.29)	
Successful in Science	0.01	0.01	0.00
	(0.10)	(0.09)	
Successful in Literature	0.56	0.32	0.24***
	(0.50)	(0.47)	
Successful Social Sciences	0.32	0.35	-0.03
	(0.47)	(0.48)	
Observations	313	348	661
Qualitative Track Students			
Successful Math	0.01	0.02	-0.00
	(0.12)	(0.13)	
Successful in Science	0.00	0.00	0.00
	(0.00)	(0.00)	
Successful in Literature	0.47	0.24	0.24**
	(0.50)	(0.43)	
Successful Social Sciences	0.32	0.31	0.02
	(0.47)	(0.46)	
Observations	74	114	188

Notes: Standard deviations for females and males are in parentheses.

Table 16: I am successful: Science-Math Track

	(Math)	(Sci.)	(Lit.)	(Social)
Female	-0.086**	-0.080**	0.124***	-0.055*
	(0.040)	(0.037)	(0.038)	(0.029)
Correct Answers in Math 1	0.006			
	(0.006)			
Correct Answers in Math 2	0.026***			
	(0.004)			
Correct Answers in Science 1		0.006		
		(0.005)		
Correct Answers in Science 2		0.018***		
		(0.004)		
Correct Answers in Lit 1			0.019***	
			(0.005)	
Correct Answers in Lit 2			0.010	
			(0.007)	
Correct Answers in Social 1				0.008***
				(0.002)
Correct Answers in Social 2				-0.025
				(0.029)
Observations	701	701	701	701

Standard errors in parentheses

* p¡0.10, ** p¡0.05, *** p¡0.010