Gendered Effects of Labels on Advanced Course Enrollment

Paola Ugalde A.
Louisiana State University

April 28, 2025

Abstract

This paper investigates gender differences in how high school students react to standardized test performance labels regarding their advanced math and English enrollment decisions. Using a regression discontinuity design, I find that women labeled as not proficient in math are less likely to enroll in advanced math courses than their proficient-labeled peers. In English, the effect of labels on women's enrollment decisions is smaller and nosier. While, on average, men enroll in advanced classes at a lower rate than women, men's likelihood of enrollment is not impacted by the labels they receive, regardless of subject. These findings highlight unintended consequences of testing practices that affect human capital investment decisions differentially by gender, potentially contributing to the persistent underrepresentation of women in male-dominated fields.

1 Introduction

Since the introduction of school accountability systems, considerable attention has been devoted to the gender gaps in achievement. However, less attention has been paid to how males and females perceive, interpret, and respond to the feedback provided in the context of standardized testing. Students' reactions to this feedback influence their academic decisions, with potential long-term implications for their human capital development. In this paper, I investigate the gender differences in the decision-making processes of high school students regarding enrollment in advanced classes.[1] I study their decisions after receiving a score and a label summarizing their performance on subject-specific standardized tests.

Like most states, North Carolina has implemented a test-based accountability program to assess students' knowledge about specific subjects. Each subject tested has a threshold score, and students who score above it are labeled *proficient*. This label means they have sufficient command of skills for the respective course and are prepared for further studies in the subject (North Carolina Department of Public Instruction, 2016a). Students who score below are labeled *not proficient*. There is evidence that these types of discrete performance summaries, common in state testing programs (May et al., 2009), can influence students' educational decisions. For example, Papay et al. (2016) finds that earning a favorable label positively affects the post-secondary education decisions of urban low-income students. I use a regression discontinuity approach to identify the causal effect of the proficiency labels on students' educational investment decisions and the gender difference therein. I focus on the labels students received in their Math and English End-of-Course (EOC) tests and the subsequent decision to enroll in each subject's regular or advanced version of the class for the following year.[2] These advanced classes are more challenging and cover more topics than regular classes.

I find that around the proficiency cutoff the labels influence women's decisions to enroll

[1]For the purpose of this paper, gender is defined as sex assigned at birth.

[2]During my study period, there is one Math EOC test at the end of Math 1 and one English EOC test at the end of English 2 taken during the first years of high school.

in advanced math courses. Specifically, women labeled as not proficient on their math EOC test are less likely to enroll in advanced math classes than their proficient-labeled female peers. In the case of English, the proficiency label has a smaller, noisier effect suggesting that women labeled as not proficient are slightly less likely to enroll in advanced English compared to their proficient-labeled peers. While men are generally less likely to enroll in advanced classes than women, performance labels do not impact their course decisions, regardless of subject.

Additionally, women labeled not proficient in either subject are less likely to state they will attend a 4-year college than their proficient female peers.[3] Males' likelihood is the same regardless of their label. These findings suggest that proficiency labels given early in high school generate gender-specific reactions, persisting until the end of high school and influencing college attendance.

Considering that the labels merely reiterate information already present in the continuous score and that students just above and just below the cutoff are similar other than in their label, one would not expect a discontinuous change in the likelihood of enrolling in advanced courses. Thus, the finding that women's advanced enrollment decisions are influenced by their proficiency labels highlights an unintended consequence of the state's testing policy that could have significant long-term effects by leading to different educational outcomes. I find that advanced enrollment correlates with better high school academic outcomes and higher rates of planning to attend a 4-year college. From this perspective, non-proficient women may be under-enrolling in advanced classes and missing out on these potential benefits. Research also shows that high school students' course-taking behavior is linked not only to their academic outcomes in high school but also in college (Long et al., 2012; Aughinbaugh, 2012). Moreover, the decision to enroll in advanced math courses contributes positively to future earnings (Rose and Betts, 2004; Joensen and Nielsen, 2009; Goodman, 2019), and increases women's likelihood of majoring in male-dominated fields (Levine and Zimmerman, 1995;

[3] I only observe what students plan to do after high school, but I do not know if they executed those plans.

Card and Payne, 2021). This is particularly relevant given the persistent underrepresentation of women in high-paying fields like STEM and Economics (Porter and Serra, 2020; Jiang, 2021).[4]

This paper contributes to the growing literature investigating the effects of early grades and performance labels on academic choices (Papay et al., 2016; Avery and Goodman, 2022; Brownstein, 2023). My findings on the gendered effects of labels highlight the potential unexpected biases of this widely used evaluation metric. I also add to the evidence that early feedback can change students' education investment decisions, which has important implications for the design of effective educational interventions.

Most of the research on grade sensitivity focus on college student's behavior.[5] I contribute to this body of work by investigating students' responsiveness to performance earlier in life, which can provide valuable insights into critical human capital decisions that arise later on. Moreover, the findings support the existing evidence that women are more responsive to grades than men, particularly in male-dominated fields (Rask and Tiefenthaler, 2008; Ost, 2010; Owen, 2010; Goldin, 2015; Kugler et al., 2021; Kaganovich et al., 2021; Ahn et al., 2022; Ugalde A., 2024), and reveal similar gender differences among high school students, with stronger patterns in math, a stereotypically male subject.

This paper also contributes to the broader literature on gender stereotypes and their role in shaping academic behavior. A large body of research documents the effects of stereotype threat—the concern of confirming negative group stereotypes—on performance outcomes (Spencer et al., 1999; Steele, 1997; Spencer et al., 1999). More recent work in economics has shown that gender differences in how individuals interpret feedback can influence beliefs and choices, especially in domains shaped by gender stereotypes (Favara, 2012; Bordalo et al., 2019; Coffman et al., 2023). The gender-differentiated responses to performance

[4]See Altonji et al. (2012, 2014, 2016) for evidence of higher wages in STEM and Economics fields.

[5]See Chizmar (2000); Rask and Tiefenthaler (2008); Ost (2010); Owen (2010); Main and Ost (2014); Goldin (2015); Astorne-Figari and Speer (2019); Bestenbostel (2021); Owen (2021); Kugler et al. (2021); Kaganovich et al. (2021); Ahn et al. (2022). These references are intended as examples of this literature instead of an exhaustive list.

labels observed in this paper, particularly the contrast between math, a male-stereotyped field (Breda et al., 2020), and English, are consistent with the idea that stereotypes may shape not only performance but also how students internalize and respond to feedback.

My results raise the question of why the performance label affects women's academic decisions but not men's.[6] Like most of the work on grade sensitivity, I cannot precisely identify the mechanisms driving this gender difference. However, evidence shows that men and women often react differently to positive and negative feedback. For instance, women tend to attribute negative feedback to lack of ability (Roberts and Nolem-Hoeksema, 1989; Shastry et al., 2020), update their beliefs more pessimistically (Berlin and Dargnies, 2016; Coffman et al., 2024), and are less confident in their math ability (Ellis et al., 2016). Parents and teachers could also contribute to this differential reaction by influencing the student's enrollment decision differently by gender and subject (Tungodden and Willén, 2022; Gentrup and Rjosk, 2018; Lavy and Sand, 2018).

Regardless of the mechanism, the unexpected gendered effects of labels could exacerbate gender gaps observed in male-dominated fields like STEM and Economics later in life. To address this critical issue, education authorities should prioritize the development of testing systems with a neutral impact on students' decisions across genders.

The rest of the paper is organized as follows. Section 2 discusses the North Carolina standardized testing system and the administrative data used in the analysis. Section 3 explains the empirical strategy. Sections 4 and 5 present the results. Section 6 discusses the results. Finally, Section 7 concludes.

[6]This difference does not inherently indicate that either gender is "making a mistake." Such differences can result from rational decision-making based on each gender's priors and uncertainty about their ability.

2 Institutional Background and Data

2.1 Background

The North Carolina Department of Public Instruction (NCDPI) uses standardized end-of-course (EOC) tests to "sample a student's knowledge of subject-related concepts as specified in the North Carolina Standard Course of Study and to provide a global estimate of the student's mastery of the material in a particular content area" (North Carolina Department of Public Instruction, 2013) for high school accountability purposes. During high school, every student must take an EOC test at the end of the first math course (Math 1), the second English course (English 2) and the Biology course. I focus on Math 1 and English 2 tests, usually taken during 9^{th} and 10^{th} grade, respectively. Each of these tests represent at least 20% of the final course grade (North Carolina Department of Public Instruction, 2016b, 2016c).

Based on the test scores, students are classified into five achievement levels. Students that receive at least an achievement level of 3 are considered proficient, which means that they have sufficient command of skills for the respective course and are prepared for further studies in that subject (North Carolina Department of Public Instruction, 2016a). For each test, students receive an Individual Student Report (ISR) that includes the test score (scale score), achievement level and a description of it, proficiency (yes/no), percentile, an on track for college and career ready indicator (yes/no), and the Quantile Framework for math and Lexile Framework for English.[7]

For each subject, there is an established threshold above which students are considered proficient. These clear cutoffs allow for the application of the empirical strategy described in Section 3. The NCDPI establishes these thresholds through a process called standard setting. This process involves panels of educators who assess test items in multiple rounds

[7]See Appendix Figure A1, for sample ISR. The scale score allows comparisons of scores across different test forms. The Quantile Framework for math"describes what mathematics the student likely already understands and what the student is ready to learn in the future". The Lexile Framework"describe how strong a student's reading is." (North Carolina Department of Public Instruction, 2019b)

and recommend cut scores that distinguish different performance levels. The thresholds are determined based on the panel's judgment of the level of knowledge and skills that students at each performance level are expected to demonstrate. The cut scores are recommended prior to observing test scores and are not a function of the distribution of scores.[8] For the period of interest (2013-2014), the thresholds were established in July 2013 and remained unchanged until 2019-2020 when the NCDPI went through a new standard setting process for the newly adopted curriculum.

Each scholar year, during the spring semester (around March or April), students make the choices for the classes they will be taking the next year. When choosing the classes, students often must decide whether to enroll in a regular or honors version of a given course. This is the case for Math 2, for instance. The honors version is a class with a higher level of difficulty, which studies the topics in a deeper way than the regular version and sometimes covers more topics. Given the higher level of difficulty, students taking honors classes get quality points that make grades from honors classes have a higher weight in GPA calculations. For example, a C in a honors class is equivalent to a B in a regular class (North Carolina Department of Public Instruction, 2022).

Once students advance to higher grades, the available options can be more than just honors and regular. For example, some high schools offer Advanced Placement (AP) classes in English Language and Literature and International Baccalaureate (IB) English classes. AP and IB courses are college-level classes in which students can earn college credit depending on performance on a test at the end of the course. High school students get even more quality points for taking AP and IB courses than for honor classes.

However, quality points and more knowledge are not the only benefits of taking advanced classes. Taking more rigorous math, English, or science courses during high school is associated with a higher probability of high school graduation and 4-year college attendance (Long et al., 2012; Aughinbaugh, 2012).[9] College students who took advanced classes during

[8]For more details see North Carolina Department of Public Instruction (2016b,c)

[9]The National Center for Education Statistics (2019) report about math, science and reading instruction

high school tend to complete more college credits, have higher GPAs, and are more likely to graduate (Long et al., 2012). Moreover, taking advanced math classes during high school has been linked to higher future earnings (Rose and Betts, 2004; Joensen and Nielsen, 2009; Goodman, 2019).

2.2 Data

In order to determine if there is a gender difference in the reaction to the proficiency label received on the EOC tests, I use administrative records from the North Carolina Education Research Data Center (NCERDC). The analysis focuses on the cohorts that began public high school in the fall of 2013 and 2014. Since the objective is to study the effect of the labels on class choices for the next scholar year, I restrict the samples to those students taking the relevant EOC tests during the fall semesters. This guarantees that students have received the exam results before making course decisions. Additionally, I restrict the samples to students with observable transcript information for the following school year.

The Math 1 sample includes 27,997 students that began high school in the 2013-2014 or 2014-2015 school years and took the Math 1 EOC test at the end of the fall semester during their freshman year.[10] The English 2 sample includes 72,395 students that began 10th grade in the fall of 2013 or 2014 and that took the English 2 EOC test at the end of that semester.[11] Table 1 shows summary statistics for both samples. They are balanced in terms of gender with 51% of women. More than half the students are white (54% math, 57% English), around 25% Hispanic and 12% or less are black. The proportion of economically disadvantaged students (EDS) is 52% and 43% for the math and English samples, respectively. In terms of

finds a positive correlation between taking advanced math classes and 4-year college acceptances.

[10] About 16% of the students took the Math 1 class during the fall of their freshman year, 51% took it after the fall semester and 33% took it during 8th grade (some middle schools offer 9^{th} grade-level math classes, which allows students to gain high school credits and take 10^{th} classes during their freshman year). Fall students performed better in academic terms than the after fall, but worse than the students that took the class during middle school. See Appendix Table A1.

[11] Around 38% of the students took the English 2 class during the fall of their sophomore year, 62% took it after the fall semester and 7% during their freshman year. There are not economically significant differences between the fall and after fall students. See Appendix Table A2

academic outcomes, 63% (35%) of the students in the math (English) sample were deemed proficient on their middle school math test. On the middle school reading test these numbers are 53% and 38% for the math and English samples, respectively.[12] Nearly half (49%) of the students in the math sample and 35% of the students in the English sample were labeled as *not proficient* on the Math 1 and English 2 EOC tests, respectively.

Additionally, I am interested in the effect of the performance labels on outcomes closer to the end of high school, like the number of higher-level advanced classes in math and English that students take and the plans to attend a 4-year higher education institution. Therefore, in section 5 I further restrict the samples to students who graduated at the end of their fourth year.[13]

3 Empirical Strategy

The existence of a cutoff that determines the proficiency status suggests a regression discontinuity (RD) approach. As an identification assumption, RD requires that students just above and just below the cutoff be similar except for their labels. The RD design estimates the size of a discontinuity. In other words, it identifies the effect of the label on the likelihood of enrolling in the advanced version of the class for the next school year within a bandwidth around the cutoff.

Given that the main objective is to establish if men and women react differently to their performance label on the EOC test, the gender difference in discontinuity is estimated by the interaction of a female indicator and a treatment variable defined as receiving the *not proficient* label in the EOC test.[14] This approach is sometimes called difference-in-discontinuity because it combines regression discontinuity (RD) and differences-in-differences

[12]The reading and math tests during 8^{th} are part of the end-of-grade (EOG) exams that North Carolina students take at the end of the year from grades 3 to 8 in order to measure their performance on "the goals, objectives, and grade-level competencies specified in the North Carolina Standard Course of Study" (North Carolina Department of Public Instruction, 2019a).

[13]The main results are robust to using the restricted samples.

[14]Alternatively, treatment could be defined as being labeled *proficient*, and the conclusions would be the same.

(DD) (Buser and Yuan, 2019, Grembi et al., 2016; Eggers et al., 2018; Galindo-Silva et al., 2020; Canaan et al., 2022). I estimate the following model:

$$\begin{aligned} Y_{ijkt} =& \beta_0 + \beta_1 \mathrm{F}_i + \beta_2 \text{Non-Prof}_{ijkt} + \beta_3 (\mathrm{F}_i \cdot \text{Non-Prof}_{ijkt}) \\ &+ f(S_{ijkt}) + \text{Non-Prof}_{ijkt} \cdot f(S_{ijkt}) + \gamma \mathbf{X}_{ij} + \eta_j + \nu_t + \epsilon_{ijkt} \end{aligned} \tag{1}$$

where $k \in \{\text{math, English}\}$, and Y_{ijkt} is the outcome variable for student i at high school j that took the test for subject k during school year t. F_i is an indicator variable equal to one for females. Non-Prof_{ijkt} is an indicator that takes value one when the student is labeled not proficient on the EOC test. η_j and ν_t are high school and year fixed effects, respectively. $\mathbf{X}$ includes controls like middle school test scores, race, EDS. $f(S_{ijkt})$ represents a function of the EOC test score, which is the running variable in this case. The interaction between $f(S_{ijkt})$ and Non-Prof_{ijkt} allows for different slopes above and below the cutoff. The main results in the next sections are estimated using $f(S_{ijkt})$ as a first-degree polynomial, however all the results are robust to using a second-degree polynomial instead. The parameter of interest is β_3, the coefficient on the interaction between being a female and non-proficient that estimates the gender difference in the discontinuity. Standard errors are clustered at the school level. To assess the sensitivity of the findings to the clustering approach, I also examine the robustness of the results to clustering at alternative levels, including district and test score, and to using robust standard errors.

Following Grembi et al. (2016), I compute the optimal bandwidths separately for men and women using the algorithm developed by Calonico et al. (2014, 2019). I then use the average of these bandwidths to estimate Equation 1. The resulting bandwidths for the main results are 6.43 for math and 5.55 for English. Due to the discrete nature of the running variable, I present results for bandwidths of ±6 for math and ±5 for English in section 4. The findings remain consistent when using bandwidths of ±7 and ±6, respectively. Figure 3 further displays point estimates across a bandwidth range from 2 to 15. As a robustness check, I also apply the optimal bandwidth selection method from Xu (2017), given that the outcomes considered are not continuous.[15]

3.1 Validity of Design

In this subsection, I provide evidence that supports the use of the empirical strategy discussed above. The validity of this design requires that students do not perfectly manipulate their test scores and that all other factors that play a role when deciding whether to take the advanced or regular version of a class are continuously related to the test scores. Evidence of those supports the assumption that students above and below the proficiency threshold are similar to each other in terms of relevant characteristics, except their proficiency label.

3.1.1 Test Score Manipulation

An exogenous running variable is required because manipulation of the test scores can lead to identification problems. I apply the tests described in McCrary (2008) and Frandsen (2017) where the idea is to test the continuity of the running variable density at the cutoff. For both subjects and genders, the hypothesis of manipulation cannot be rejected at usual significance levels, regardless of the test. This is not surprising given the distribution of the test scores in Appendix Figure A2. In both cases, there are spikes in the distribution around the cutoff. However, the spikes seem to be similar across genders in both subjects, which suggests that if any manipulation exists, it is similar for men and women. Nevertheless, the validity of the design is only compromised when the agents can "precisely" manipulate the running variable (Roberts and Whited, 2013), and this kind of manipulation is dubious in the case of the EOC tests in North Carolina.

For instance, one can think that students have some control over their scores but they cannot predict them with certainty or marginally change them around the cutoff. There are several test versions, and the exact number of correct answers required to achieve proficiency

[15] As with Calonico et al. (2014, 2019), I compute the optimal bandwidths separately for men and women and use their average as the bandwidth for estimation.

is not public information and varies by test version. Given that the test scores are used to assess teacher and high school performance, one might be concerned that the teachers are manipulating the results. However, this seems very implausible because tests are not graded by the professor. Instead the NCDPI has in place a centralized grading system with a rigorous protocol to ensure the security of the materials before and during the tests and to avoid any manipulation when transporting and scanning the answer sheets at the NCDPI offices (North Carolina Department of Public Instruction, 2016b, 2016c).[16]

Another explanation for the shape of the distributions is the performance measurement system that was in place in North Carolina at the time. Under that system, high school performance depended heavily on the percentage of students considered proficient on the EOC tests (North Carolina Department of Public Instruction, 2015). Such proficiency-count systems create strong incentives for the teachers to direct resources and attention to students on the margin of being above the proficiency threshold, and to pay less attention to the distribution's tails (Macartney et al., 2018).

On the other hand, given that the interest is in the gender difference in the discontinuity, it is important to guarantee that if any manipulation exists, it is similar across genders. Following the visual test proposed by Grembi et al. (2016), Figure 1 plots the difference between male and female test score densities for Math 1 and English 2, along with a third-degree polynomial fitted to the data separately above and below the threshold and 95% confidence intervals. These figures support the assumption that the difference in the densities across genders is continuous at the cutoff.

3.1.2 Continuity of Predetermined Covariates

Given that the main interest is the gender difference in the discontinuity, it is important to rule out the possibility of discontinuities in the predetermined covariates varying by gender. In order to do so, I estimate model (2) for each subject $k \in \{\text{math, English}\}$ with each of the

[16] Dee et al. (2019) documents the elimination of teacher test score manipulation once a centralized grading system was adopted in New York.

predetermined controls included in $\mathbf{X}$ as dependent variables within the optimal bandwidth used in the main results (±6 for math and ±4 for English).

$$\begin{aligned}X_{ijk} =& \beta_0 + \beta_1 \mathrm{F}_i + \beta_2 \text{Non-Prof}_{ijk} + \beta_3 (\mathrm{F}_i \cdot \text{Non-Prof}_{ijk}) \\ &+ f(S_{ijk}) + \text{Non-Prof}_{ijk} \cdot f(S_{ijk}) + \epsilon_{ijk}\end{aligned} \tag{2}$$

Results are presented in Tables 2 and 3 for math and English, respectively. There are no gender differences for any of the math covariates. There is a small (2%) gender difference in the proportion of black students in the English 2 sample. However, the main results control for all these covariates in order to avoid any biases due to the discontinuities.

The continuity of the covariates, ignoring the possibility of gender differences, is studied by estimating instead model (3) for each subject $k \in \{\text{math, English}\}$ with each of the predetermined controls included in $\mathbf{X}$ as dependent variables.

$$X_{ijk} = \beta_0 + \beta_1 \text{Non-Prof}_{ijk} + f(S_{ijk}) + \text{Non-Prof}_{ijk} \cdot f(S_{ijk}) + \epsilon_{ijk} \tag{3}$$

The results are presented in Appendix Figures A3 and A4 for math and English, respectively. They suggest that there are no statistically significant discontinuities at the cutoff except for the proportion of black students in the English sample (in line with the results of the previous analysis).

Overall, I find no evidence of economically significant discontinuities in covariates at the threshold for both math and English tests, which supports the chosen empirical strategy and the results presented in the paper.

4 Early High School Outcomes

This section presents the main results of the study, which are first illustrated in Figure 2. The proficiency cutoff is represented by zero. To the left of zero students are labeled non-proficient. The markers represent the proportion of students of each gender that take honors Math 2 or advanced English 3 within each 1-point bin. Advanced English 3 includes the honors class and AP or IB classes when offered. The lines are fitted values from the estimation of model (1) without any controls.[17]

The level differences across genders indicate that women are more likely to enroll in honors Math 2 and advanced English 3 than men. This result is consistent with previous literature that finds women are more likely to take advanced courses during high school (Shettle et al., 2007; Nord et al., 2011; Long et al., 2012), and with the literature that investigates women's advantages in educational attainment.[18]

Despite the level differences across genders, the main objective is to establish if men and women react differently to their label on the EOC test. This differential reaction is illustrated by the discontinuities at the threshold. In panel 2a, there is no discontinuity at the cutoff for men (blue solid line). This indicates that the likelihood of men enrolling in honors Math 2 does not change with their proficiency label on the EOC test. On the other hand, there is a discontinuity at the cutoff for women (orange dashed line). Non-proficient women are less likely to enroll in honors Math 2 than proficient ones. In other words, there is a differential reaction between women and men to their proficiency label on the Math 1 EOC test. In the case of English, panel 2b shows that there is no discontinuity at the cutoff for either gender. Neither men's nor women's likelihood of enrolling in Advanced English 3 changes with their proficiency status.

I further analyze this in Table 4, which presents the estimation of model (1) for Math in columns (1) and (2), and English in columns (3) and (4).[19] The outcome variable is indicated at the top of the columns. For math, it equals one when the student enrolls in honors Math 2, zero otherwise. For English, the outcome variable is equal to one when the student enrolls in the advanced version of English 3, and zero otherwise. The specifications in columns (2)

[17]Columns (1) and (3) in Table 4.

[18]See for example Goldin et al. (2006); Bailey and Dynarski (2011); Autor and Wasserman (2013); Bertocchi and Bozzano (2019); Delaney and Devereux (2021); Cappelen et al. (2023).

[19]See Table A3 for bandwidths ± 7 and ± 6 for math and English, respectively.

and (4) in addition to the middle school test scores control for school and year fixed effects, EDS, and race. The sample size differences occur because middle school test scores could not be obtained for all students in the sample. Nonetheless, the proportion of students enrolling in honors Math 2 and advanced English 3 is around 23% and 42%, respectively, regardless of the specification.

The coefficient on the female indicator is positive and statistically significant (p-value $<$ 0.01). This suggests a gender difference in the likelihood of enrolling in advanced classes that favors women. This gap corresponds to the level difference illustrated in Figure 2 and discussed earlier.

The coefficient on non-proficient (Non-Prof) measures the discontinuity on men's likelihood of enrolling in advanced classes caused by the change in performance label at the threshold. Regardless of the specification and subject, the coefficient is small and not statistically different from zero. This indicates that men's probability of enrolling in honors Math 2 or advanced English 3 does not change due to their proficiency label on the respective EOC test.

The coefficient of interest is the interaction between being a woman and getting a non-proficient label (F*Non-Prof) on the respective EOC test. This coefficient estimates the gender difference in the discontinuity. A statistically significant coefficient suggests that the impact of the non-proficient label on women's likelihood of enrolling in advanced classes differs from that on men. In other words, it indicates a differential reaction between women and men to their performance labels on the EOC tests.

In Table 4 column (1), the coefficient for math is negative and statistically significant (p-value $<$ 0.01). This result shows that women respond differently than men to their proficiency label on the math EOC test. Specifically, women labeled as not proficient are less likely to enroll in honors Math 2 compared to those labeled proficient. Column (2) includes controls for school and year fixed effects, race, economically disadvantaged status, and middle school test scores in math and reading. These middle school scores act as proxies for students' ability in these subjects prior to high school, as well as indicators of potential comparative advantages. The literature suggests that women tend to have a comparative advantage in verbal skills, while men have one in math (Aucejo and James, 2021; Breda and Napp, 2019; Delaney and Devereux, 2019). Despite the inclusion of these controls, the coefficient of interest remains largely unchanged, and its statistical significance is robust to alternative clustering choices.[20,21]

In Table 4 column (3), the coefficient on the interaction between female and being labeled non-proficient on the English 2 EOC test is marginally not statistically significant (p-value = 0.115). This suggests that within 5 points of the cutoff, women's reactions to their performance label on the English EOC test are not statistically different from men's. The absence of a gender difference is likely due to insufficient statistical power from small sample sizes within narrow bandwidths. The inclusion of controls in column (4) improves the precision of the estimate.[22] It shows that women labeled as not proficient are slightly less likely to enroll in advanced English 3 compared to those labeled proficient (p-value < 0.05). Therefore, women also react differently from men to their proficiency label on the English EOC test, although these results appear somewhat noisier than those for math.[23]

A potential concern is that students who fail the corresponding math or English courses, partly due to their test performance, may be mechanically less likely to enroll in advanced courses. However, failure rates are low—3% for Math 1 and 2% for English 2. Moreover, the results remain robust even after excluding students who failed the courses, as shown in Appendix Table A8.

The results in Table 4 are qualitatively robust across different bandwidths. Figure 3 shows the estimates of the gender difference in the discontinuity for various bandwidths around

[20]See Table A4 in the Appendix for results that add each control separately.

[21]See Appendix Table A5 for results using different clustering levels, including district and test score. Column (4) also presents results when test scores are standardized by year; results are similar in both magnitude and significance.

[22]See Table A6 in the Appendix for the inclusion of each control separately.

[23]See Appendix Table A7 for results using different clustering levels, including district and test score. Significance disappears only when clustering at the test score level. Column (4) also presents results when test scores are standardized by year; estimates are similar in magnitude and significance.

the cutoff for both subjects. For math, using the optimal bandwidth selection methods from Calonico et al. (2014, 2019) and Xu (2017) yields similar results, with bandwidths between ±6 and ±7.[24] Across all bandwidths, the gender difference estimates consistently range between 4 and 7 percentage points and are statistically significant. Comparing these estimates to the proportion of women who score within 6 points of the cutoff and enroll in Math 2 (27%), roughly translates to a 15 to 26% gap in the likelihood of proficient and non-proficient women enrolling in honors Math 2.

For English, using the optimal bandwidth selection methods from Calonico et al. (2014, 2019) produces bandwidths between ±5 and ±6. In contrast, Xu (2017)'s method results in wider bandwidths, between ±10 and ±11. However, across all bandwidths, the estimates of the gender difference consistently range between 2 and 4 percentage points and are statistically significant for bandwidths of 5 or wider. Comparing these estimates to the proportion of women who score within 5 points of the cutoff and enroll in advanced English 3 (47%), roughly translates to a 4.2 to 8.5% gap in the likelihood of proficient and non-proficient women enrolling in advanced English 3.

5 End of High School Outcomes

In this section, I estimate model (1) using as dependent variable end-of-high school outcomes such as the total number of advanced classes taken in each subject, or plans to attend a 4-year college. These results provide insights into the longer-term effects of the gender differences in reaction to performance earlier in high school. Results are reported for the preferred specification that includes controls for middle school test scores, school and year fixed effects, EDS, and race at their respective optimal bandwidths.

Table 5 presents the results for both math and English for different outcomes: the probability of taking at least one higher-level advanced class, the total number of higher-level advanced courses taken during high school, the plans to attend a 4-year college, and the likeli-

[24] Xu (2017)'s method accounts for the discontinuous nature of the outcome variable.

hood of dropping out. The set of higher-level advanced math classes includes any senior-level honors class, as well as any AP, IB, or community college math course. Higher-level advanced English classes correspond to any junior or senior honors English class, as well as any AP, IB or community college English course. This set of courses includes more than the courses considered as possible options for English 3 in the main results.

Table 5 columns (1)-(2) and (5)-(6) show that, as with the beginning of high school outcomes, women are more likely to take higher-level advanced classes in both subjects. Despite this level difference across genders, there is evidence of a gender gap in the student's reactions to their proficiency label on the EOC test. On the extensive margin, men labeled proficient and not proficient are equally likely to enroll in higher-level advanced math or English classes. On the other hand, the likelihood of ever taking higher-level math advanced classes is four percentage points lower among women labeled not proficient (p-value<0.01) compared to proficient ones. On the intensive margin, women who receive the proficient label enroll in a significantly greater number of higher-level advanced classes in both subjects than women non-proficient-labeled women.

Regardless of the subject, women are more likely to state plans to attend a 4-year college (columns (3) and (7)). This is consistent with the literature that finds female advantages in educational attainment (Goldin et al., 2006; Bailey and Dynarski, 2011). However, there is a gendered effect of the performance label. The likelihood of planning to attend a 4-year college is the same between proficient and non-proficient labeled male students. On the other hand, in both math and English, non-proficient-labeled women are less likely to state they will attend a 4-year college than their proficient-labeled female peers. This implies that the proficiency label on a test early in high school generates differential reactions across genders, persisting until the end of high school and influencing crucial outcomes such as college attendance.

Finally, columns (4) and (8) in Table 5 show that dropout rates are not significantly affected by students' EOC test performance. Unlike advanced course enrollment and college plans, there is no gendered effect on the likelihood of dropping out. This suggests that while proficiency labels influence academic choices and aspirations, they do not impact students' decisions to remain in school.

6 Discussion

Using a regression discontinuity approach, I find evidence of a gender difference in reaction to the proficiency label received on standardized tests early in high school. Men's likelihood of enrolling in advanced classes earlier or later in high school remains the same regardless of the label they receive. Their plans to attend college are also unaffected. The same is not true for women. Women labeled as proficient are more likely to enroll in advanced math classes through high school and more likely to plan to attend college than non-proficient-labeled women. The effect of being labeled proficient in English although positive, is smaller and noisier than math's.

Since the labels summarize the information already present in the continuous score and students just above and just below the cutoff are similar other than in their labels, we would not expect a discontinuous change in advanced course enrollment. Thus, the findings highlight the unintended gendered effects of a widely used evaluation metric that could have long-term consequences. The course-taking behavior of high school students is associated with their academic outcomes in high school and college (Long et al., 2012; Aughinbaugh, 2012). Enrollment in advanced math courses contributes positively to future earnings (Rose and Betts, 2004; Joensen and Nielsen, 2009; Goodman, 2019) and increases women's likelihood of majoring in male-dominated fields (Levine and Zimmerman, 1995; Card and Payne, 2021). This is particularly relevant for women's labor market outcomes, considering their persistent underrepresentation in high-paying fields like STEM and Economics (Porter and Serra, 2020; Jiang, 2021).[25]

As mentioned in section 2.1, students are classified into five achievement levels based

[25]See Altonji et al. (2012, 2014, 2016) for evidence of higher wages in STEM and Economics fields.

on their test scores. The main analysis focuses on the threshold that defines proficiency, i.e., level three and above. Tables A9 and A10 in the Appendix explore potential gender effects across other achievement levels on advanced course enrollment for math and English, respectively. The proficiency threshold (2 vs. 3, column (2)) confirms the results discussed in Section 4. There is a gender effect at the lowest performance levels, with women in the lowest level enrolling in advanced courses less often than those in the second-lowest, while men show no significant response. This effect is smaller in math compared to the more salient 2 vs. 3 threshold. In contrast, for math, differences between the higher levels—3 vs. 4 and 4 vs. 5 (columns (3) and (4))—do not reveal significant differences for women. This suggests that once female students are labeled proficient, further distinctions in performance within the proficient range have less influence on their decision-making compared to crossing the proficiency threshold. However, in English, results in Table A10 reveal a significant positive effect for women at the 4 vs. 5 threshold suggesting that even within the proficient range, performance distinctions can shape enrollment decisions.[26] This contrast indicates that in certain subjects, the relevance of labels may extend beyond the proficiency cutoff, and highlights the importance of future research examining how students respond to finer performance distinctions, particularly in fields that are not male-dominated.

Given the different policy implications, it is important to consider whether non-proficient women are under-enrolling in advanced classes or proficient women are over-enrolling. While these decisions may reflect a rational optimization process that accounts for uncertainty and prior beliefs about ability, they can lead to different educational outcomes. Figure 4 compares average Math 2 grades, high school GPA, and the proportion planning to attend a 4-year college for students enrolled in honors versus regular Math 2. Honors enrollment generally correlates with better outcomes, regardless of EOC proficiency. Proficient students in honors classes tend to perform better in terms of course grades and GPA than their proficient peers in regular classes. Although proficient students in regular classes achieve similar or slightly

[26] There is also a small negative effect for men at this threshold.

better grades than non-proficient students in honors classes, the gap shifts in favor of non-proficient honors students when quality points are factored in (see Figure A5).[27] Additionally, honors students are more likely to report plans to attend a 4-year college. The patterns are qualitatively similar when comparing regular versus advanced English courses (see Figures 5 and A6). This suggests non-proficient women may be under-enrolling in advanced classes and missing out on potential benefits. However, these findings should be interpreted with caution, as they reflect purely correlational relationships rather than causal effects.[28]

The differential effect of the performance label on women's and men's academic decisions does not necessarily imply that either gender is "making a mistake." These differences can result from rational decision-making based on each gender's priors and uncertainty about their ability. Although I cannot precisely identify the mechanisms driving this gender difference, the literature points towards differences in how men and women react to positive and negative feedback. Experimental evidence suggests that women are more likely to attribute negative feedback to ability, while men tend to attribute it to bad luck (Roberts and Nolem-Hoeksema, 1989; Shastry et al., 2020). Women tend to update their beliefs more pessimistically than men after receiving negative feedback, even when controlling for performance (Berlin and Dargnies, 2016; Coffman et al., 2024). Moreover, there is experimental evidence that women are not only less willing to compete but also less likely to do it after losing (Buser and Yuan, 2019), which is consistent with women opting out of "harder" classes after performing poorly on standardized tests. Also, women are less confident in their math ability (Ellis et al., 2016) and more sensitive to feedback on tests that measure math ability than verbal ability (Kiefer and Shih, 2006).[29]

Parents and teachers could also contribute to the gendered effect of performance labels by influencing students' enrollment decisions differently by gender and subject. There is

[27] Quality points increase the GPA weight of honors grades; for example, a C in an honors class is equivalent to a B in a regular class.

[28] See Tables 5 and A11 for the effects of the proficiency labels on plans to attend college and GPA, respectively.

[29] Kiefer and Shih (2006) find the opposite pattern holds for men.

evidence that boys' parents choose a competitive task for their sons more often than parents of girls (Tungodden and Willén, 2022). Teachers have higher math achievement expectations for boys than for girls and higher reading achievement expectations for girls (Gentrup and Rjosk, 2018). Lavy and Sand (2018) concludes that primary school teachers' biased expectations in favor of boys encourage them to enroll in advanced math and science courses during high school but discourage girls.

7 Conclusion

This study highlights the gendered effects of standardized test performance labels on students' subsequent academic decisions. The proficiency labels women receive on their tests significantly impacts their advanced enrollment decisions throughout high school, especially in math. The labels have no effect on men's course decisions, regardless of subject.

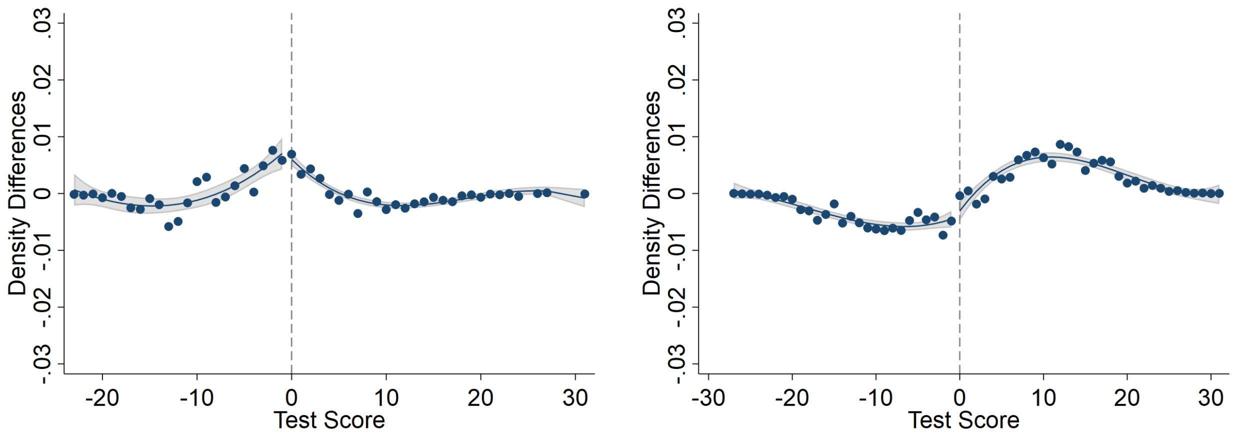
This gendered effect of labels underscores the potential unintended consequences that the widespread practice of summarizing test performance into discrete proficiency categories can have on academic decisions. Given the relationship between advanced math coursework and future earnings potential, especially in male-dominated, high-paying fields, this labeling effect could contribute to broader labor market gender inequalities. Therefore, policymakers should consider these gendered effects when designing evaluation policies that promote optimal investment decisions for all students, regardless of gender.

The methodology used in this paper does not allow me to determine precisely the mechanisms through which these gendered effects operate. The literature points towards gender differences in reaction to positive and negative feedback and the role that teachers and parents can play in enrollment decisions. Despite these potential explanations, the differential impact of the performance labels on women's and men's academic choices warrants further research that investigates the underlying mechanisms driving this gendered effect.

8 Figures

Figure 1: Test Score Densities Differences

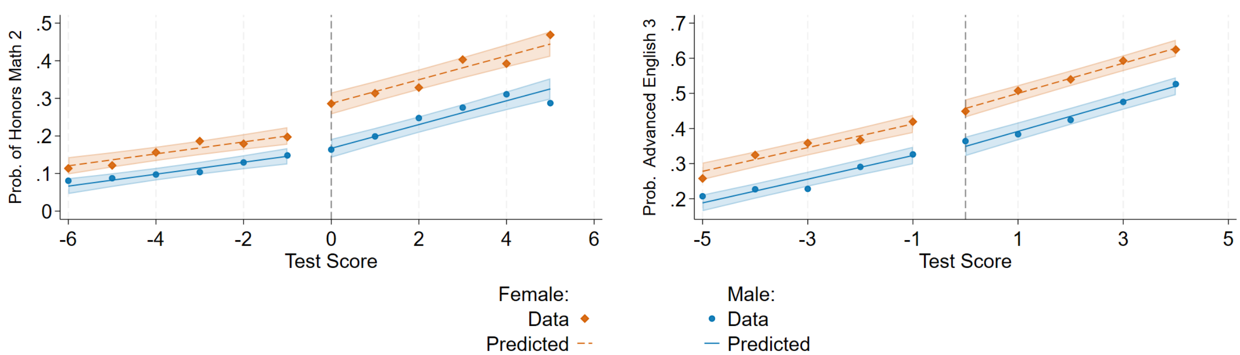
(a) Math 1. (b) English 2.



Note: This is a visual test of the continuity of the difference between female and male densities at the cutoff. The central lines are third-degree local polynomials fitted to the data separately above and below the threshold, the shaded area represent 95% confidence intervals. Scores are normalized such that a score of 0 or more means proficiency.

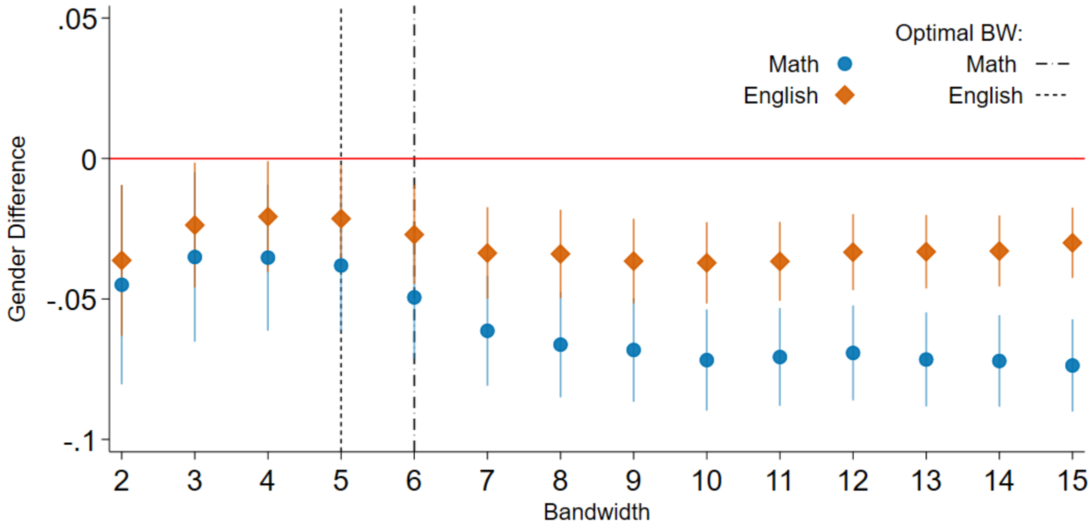
Figure 2: Discontinuity in the Likelihood of Taking Advanced Classes

(a) Math 1. (b) English 2.



Note: Test scores are normalized such that a score of 0 or more means proficiency. Lines are fitted values from regressing a dependent variable that equals one when taking honors Math 2 or advanced English 3 on indicators for being female and non-proficient on the Math 1 EOC test or English 2 EOC test, respectively, the interaction of those two variables, and a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold, within ±6 of the cutoff in panel (a), and ±5 in panel (b). Standard errors are clustered at school score level. The markers represent the proportion of students by gender that take honors Math 2 or advanced English 3 within each 1 point bin.

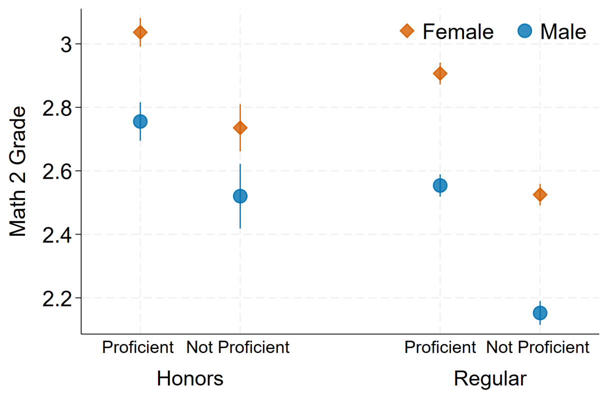
Figure 3: Estimates of the Difference in the Discontinuity for Different Bandwidths



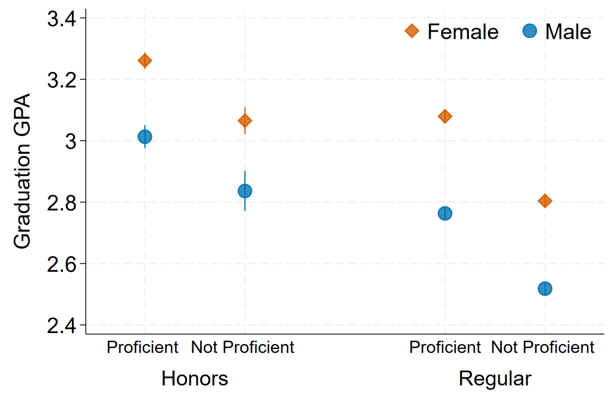
Note: Spikes represent 90% confidence intervals. Markers are the coefficients for being a non-proficient female in the preferred specification: the indicator for taking advanced English 3 or honors Math 2 class regressed on indicators for female, being non-proficient on the English 2 or Math 1 EOC test, the interaction of those two variables, a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold, high school and year FE, indicators for EDS, race, and middle school test scores. BW: bandwidth. Optimal bandwidths are calculated following the procedure described in section 3 using the algorithm developed by Calonico et al. (2014, 2019).

Figure 4: Outcomes by Proficiency Level Math 1 EOC test and Course Enrollment Type

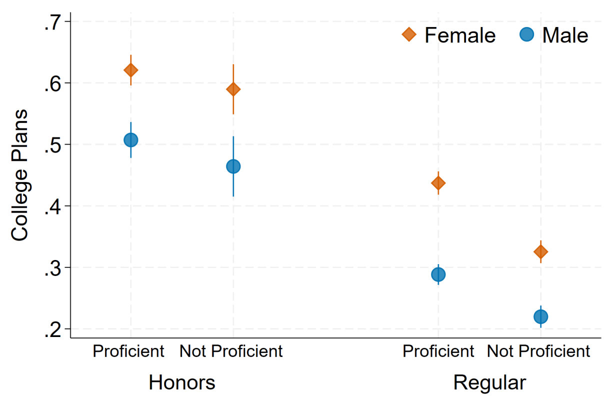
(a) Math 2 Final Grade



(b) High School GPA



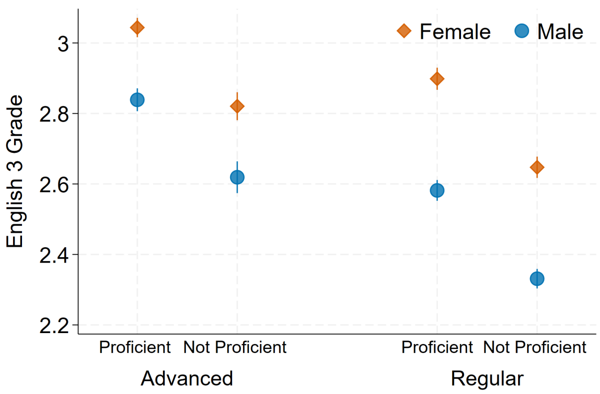
(c) Plans to Attend College



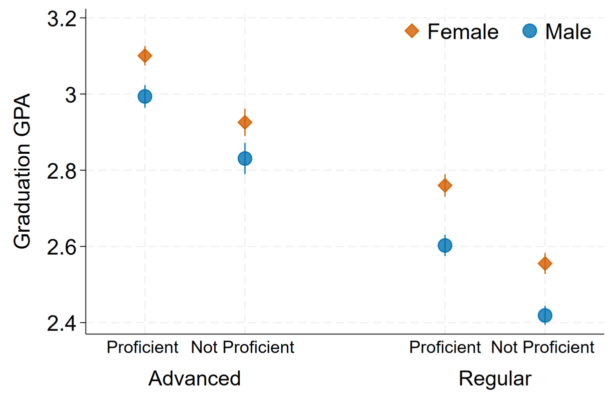
Note: Markers represent the average of the variable indicated at te top of each panel for the students in the math sample that score within ±6 from the proficiency threshold in their Math 1 EOC test. Spikes represent 95% confidence intervals.

Figure 5: Outcomes by Proficiency Level in English 2 EOC test and Course Enrollment Type

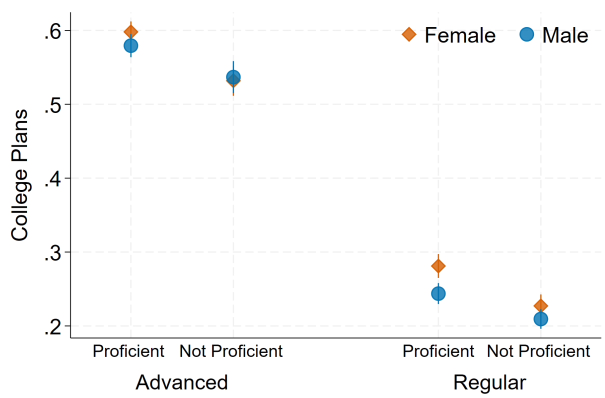
(a) English 3 Final Grade



(b) High School GPA



(c) Plans to Attend College



Note: Markers represent the average of the variable indicated at te top of each panel for the students in the English sample that score within ±5 from the proficiency threshold in their English 2 EOC test. Spikes represent 95% confidence intervals.

9 Tables

Table 1: Sample Characteristics

	Math Sample	English Sample
Female	0.51	0.51
Black	0.26	0.24
Hispanic	0.13	0.11
White	0.54	0.58
EDS	0.51	0.42
Non-Prof. Middle School Math test	0.63	0.35
Non-Prof. Middle School Reading test	0.53	0.38
Non-Prof. Math 1 test	0.49	-
Non-Prof. English 2 test	-	0.35
N	27,997	72,395

Note: Table presents sample proportions of variables of interest. Math sample includes students that took the Math 1 class during the fall semester of their freshman year. English sample includes students that took the English 2 class during the fall semester of their sophomore year. Non-Prof.: non-proficiency, EDS: Economically Disadvantaged Student. *Significant at 10%, **5%, ***1%.

Table 2: Continuity of the Covariates for Math 1 Test by Gender

	Middle School Math test	Middle School Reading test	Black	Hispanic	EDS
	(1)	(2)	(3)	(4)	(5)
Female (F)	-0.030***	0.150***	0.015*	0.009	-0.001
	(0.011)	(0.015)	(0.008)	(0.007)	(0.010)
Non-Prof.	-0.008	-0.010	-0.006	0.018	-0.005
	(0.021)	(0.032)	(0.015)	(0.013)	(0.017)
F*Non-Prof.	-0.017	0.019	0.012	-0.020*	0.012
	(0.018)	(0.025)	(0.012)	(0.011)	(0.015)
Mean	-0.10	-0.07	0.24	0.13	0.51
N	14,184	14,152	14,918	14,918	14,918

Note: Dependent variable indicated at the top of each column. Each column follows the same specification: dependent variable regressed on a variable equal to one when the student is a woman, a variable equal to one when the student is deemed non-proficient on the Math 1 EOC test, the interaction of those two and a first degree polynomial of the test score with flexible slopes above and below the proficiency cutoff and high school and year FE. Test scores are standardized such that mean is zero and standard deviation is one. EDS: Economically Disadvantaged Student. Bandwidth of ± 6 around the cutoff. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table 3: Continuity of the Covariates for English 2 Test by Gender

	Middle School Math test	Middle School Reading test	Black	Hispanic	EDS
	(1)	(2)	(3)	(4)	(5)
Female (F)	-0.178***	-0.100***	0.042***	0.005	0.044***
	(0.011)	(0.010)	(0.006)	(0.005)	(0.008)
Non-Prof.	-0.013	-0.003	0.005	-0.000	-0.014
	(0.020)	(0.017)	(0.012)	(0.009)	(0.014)
F*Non-Prof.	-0.019	0.000	-0.020*	0.008	0.016
	(0.018)	(0.015)	(0.011)	(0.008)	(0.012)
Mean	-0.13	-0.16	0.28	0.13	0.49
N	25,178	25,175	27,072	27,072	27,072

Note: Dependent variable indicated at the top of each column. Each column follows the same specification: dependent variable regressed on a variable equal one when the student is a woman, a variable equal to one when the student is deemed non-proficient on the English 2 EOC test, the interaction of those two and a first degree polynomial of the test score with flexible slopes above and below the proficiency cutoff, and high school and year FE. Test scores are standardized such that mean is zero and standard deviation is one. EDS: Economically Disadvantaged Student. Bandwidth of ±5 around the cutoff. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table 4: Probability of Enrolling in Advanced Math and English Classes

	Honors Math 2		Advanced English 3	
	(1)	(2)	(3)	(4)
Female (F)	0.119***	0.108***	0.108***	0.134***
	(0.011)	(0.011)	(0.009)	(0.008)
Non-Prof.	-0.005	-0.007	0.008	0.011
	(0.014)	(0.014)	(0.012)	(0.012)
F*Non-Prof.	-0.065***	-0.049***	-0.018	-0.021*
	(0.014)	(0.013)	(0.012)	(0.012)
Math Test MS		0.100***		0.128***
		(0.007)		(0.005)
Reading Test MS		0.003		0.070***
		(0.005)		(0.005)
Controls		✓		✓
Bandwidth	±6		±5	
Mean	0.23	0.23	0.42	0.42
R^2	0.06	0.25	0.06	0.28
N	14,918	14,146	27,072	25,137

Note: In columns (1)-(2) the dependent variable is one when taking Honors Math 2, zero otherwise. In columns (3)-(4), it is one when taking advanced English 3 class, zero otherwise. Each specification includes indicator variables for being a female (F), for being non-proficient (Non-Prof) on the Math 1 EOC test (columns (1)-(2)) or English 2 EOC test (Columns (3)-(4)); and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold. Reading Test MS: standardized middle school reading test score. Math test MS: standardized middle school math test score. Controls include school and year FE, race, and economically disadvantage status. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table 5: Effect of Non-Proficiency on End of High School Outcomes.

	Math 1 EOC Test				English 2 EOC Test			
	Prob. higher-level advanced math class	Number of higher-level advanced math classes	4-year College Plan	Dropout Prob.	Prob. of higher-level advanced English class	Number higher-level advanced English classes	4-year College Plan	Dropout Prob.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female (F)	0.121***	0.157***	0.141***	-0.006**	0.141***	0.320***	0.085***	-0.001
	(0.011)	(0.016)	(0.013)	(0.003)	(0.008)	(0.018)	(0.007)	(0.001)
Non-Prof.	0.027	0.053**	0.012	0.001	0.008	0.022	0.020*	-0.002
	(0.018)	(0.023)	(0.021)	(0.005)	(0.013)	(0.024)	(0.011)	(0.002)
F*Non-Prof.	-0.057***	-0.073***	-0.040**	-0.000	-0.013	-0.056**	-0.023**	-0.000
	(0.015)	(0.021)	(0.017)	(0.004)	(0.011)	(0.023)	(0.011)	(0.002)
Bandwidth	±6	±6	±5	±5	±5	±5	±6	±7
Mean	0.29	0.36	0.37	0.01	0.50	0.92	0.38	0.01
R^2	0.23	0.23	0.19	0.07	0.27	0.32	0.19	0.04
N	11,710	11,710	11,157	12,287	23,907	23,907	27,760	33,216

Note: Dependent variable indicated at the top of each column. Prob. higher-level advanced math class: one if the student ever took a higher-level advanced math class, zero otherwise. Prob. of higher-level advanced English class: one if the student ever took a higher-level advanced English class, zero otherwise. 4-year College Plan: one if student plans to attend a 4-year college, zero otherwise. Each column follows the same specification: dependent variable regressed on an indicator equal one when the student is female, an indicator equal to one when the student is deemed non-proficient in the respective EOC test indicated at the top, the interaction of those two and a first degree polymonial of the test score with flexible slopes above and below the proficiency cutoff. All columns include year and high school fixed effects, controls for EDS, race, and middle school test scores. EDS: Economically Disadvantaged Student. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Acknowledgments

I am grateful to Esteban Aucejo and Basit Zafar for their guidance and support. I thank Zachary Tobin, Naci Mocan, Jacob French, Daniel Silverman, and the seminar participants at Arizona State University for helpful comments and suggestions.

Data Availability

The data used in this study were provided by the North Carolina Education Research Data Center (NCERDC) and cannot be shared due to confidentiality agreements. Researchers may request access to the data directly from the NCERDC.

Declaration of generative AI and AI-assisted technologies in the writing process.

During the preparation of this work the author used ChatGPT in order to improve the readability and language of the manuscript. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the article.

References

Ahn, T., P. Arcidiacono, A. Hopson, and J. Thomas (2022): "Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors," *Working Paper*.

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

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

Altonji, J. G., L. B. Kahn, and J. D. Speer (2014): "Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs," *American Economic Review*, 104, 387–393.

Astorne-Figari, C. and J. D. Speer (2019): "Are changes of major major changes? The roles of grades, gender, and preferences in college major switching," *Economics of Education Review*, 70, 75–93.

Aucejo, E. and J. James (2021): "The Path to College Education: The Role of Math and Verbal Skills," *Journal of Political Economy*, 129.

Aughinbaugh, A. (2012): "The effects of high school math curriculum on college attendance: Evidence from the NLSY97," *Economics of Education Review*, 31, 861–870.

Autor, D. and M. Wasserman (2013): "Wayward sons: The emerging gender gap in labor markets and education." .

Avery, C. and J. Goodman (2022): "Ability signals and rigorous coursework: Evidence from AP Calculus participation," *Economics of Education Review*, 88, 102237.

Bailey, M. J. and S. M. Dynarski (2011): "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion," Working Paper 17633, National Bureau of Economic Research.

Berlin, N. and M.-P. Dargnies (2016): "Gender differences in reactions to feedback and willingness to compete," *Journal of Economic Behavior & Organization*, 130, 320–336.

Bertocchi, G. and M. Bozzano (2019): "Gender Gaps in Education," *CEPR Discussion Papers 14082*.

Bestenbostel, A. (2021): "Do Grade Signals Drive the Gender Gap in STEM? Evidence From a Regression Discontinuity," *Working Paper*.

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

Breda, T., E. Jouini, C. Napp, and G. Thebault (2020): "Gender stereotypes can explain the gender-equality paradox," *Proceedings of the National Academy of Sciences*, 117, 31063–31069.

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

Brownstein, J. (2023): "The Effect of Test Score Performance Labels on Post-Secondary Educational Outcomes: Evidence from Michigan 1," *Working Paper*.

Buser, T. and H. Yuan (2019): "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad," *American Economic Journal: Applied Economics*, 11, 225–52.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019): "Regression

Discontinuity Designs Using Covariates," *The Review of Economics and Statistics*, 101, 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Canaan, S., S. Fischer, P. Mouganie, and G. C. Schnorr (2022): "Keep Me In, Coach: The Short- and Long-Term Effects of Targeted Academic Coaching," *Upjohn Institute Working Paper 22-370*.

Cappelen, A. W., R. Falch, and B. Tungodden (2023): "Experimental Evidence on the Acceptance of Males Falling Behind," *NHH Dept. of Economics Dicsussion Paper 13*.

Card, D. and A. A. Payne (2021): "HIGH SCHOOL CHOICES AND THE GENDER GAP IN STEM," *Economic Inquiry*, 59, 9–28.

Chizmar, J. F. (2000): "A Discrete-Time Hazard Analysis of the Role of Gender in Persistence in the Economics Major," *The Journal of Economic Education*, 31, 107–118.

Coffman, K., M. R. Collis, and L. Kulkarni (2023): "Stereotypes and Belief Updating," *Journal of the European Economic Association*, 22, 1011–1054.

Coffman, K., M. P. Ugalde Araya, and B. Zafar (2024): "A (dynamic) investigation of stereotypes, belief-updating, and behavior," *Economic Inquiry*, 62, 957–983.

Dee, T. S., W. Dobbie, B. A. Jacob, and J. Rockoff (2019): "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations," *American Economic Journal: Applied Economics*, 11, 382–423.

Delaney, J. M. and P. J. Devereux (2019): "Understanding gender differences in STEM: Evidence from college applications." *Economics of Education Review*, 72, 219–238.

——— (2021): "Gender and Educational Achievement: Stylized Facts and Causal Evidence," *Working Papers 202103, School of Economics, University College Dublin*.

Eggers, A. C., R. Freier, V. Grembi, and T. Nannicini (2018): "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions," *American Journal of Political Science*, 62, 210–229.

Ellis, J., B. K. Fosdick, and C. Rasmussen (2016): "Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit," *PLOS ONE*, 11, 1–14.

Favara, M. (2012): "The cost of acting "girly": gender stereotypes and educational choices," *IZA Discussion Paper Series*, 1–63.

Frandsen, B. R. (2017): "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete," in *Regression Discontinuity Designs*, Emerald Publishing Ltd, vol. 38 of *Advances in Econometrics*, 281–315.

Galindo-Silva, H., N. H. Somé, and G. Tchuente (2020): "Does Obamacare Care? A Fuzzy Difference-in-Discontinuities Approach," GLO Discussion Paper Series 666, Global Labor Organization (GLO).

Gentrup, S. and C. Rjosk (2018): "Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling?" *Educational Research and Evaluation*, 24, 295–323.

Goldin, C. (2015): "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College," .

Goldin, C., L. F. Katz, and I. Kuziemko (2006): "The Homecoming of American College Women: The Reversal of the College Gender Gap," *Journal of Economic Perspectives*, 20, 133–156.

Goodman, J. (2019): "The Labor of Division: Returns to Compulsory High School Math Coursework," *Journal of Labor Economics*, 37, 1141–1182.

Grembi, V., T. Nannicini, and U. Troiano (2016): "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8, 1–30.

Jiang, X. (2021): "Women in STEM: Ability, preference, and value," *Labour Economics*, 70, 101991.

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

Kaganovich, M., M. Taylor, and R. Xiao (2021): "Gender Differences in Persistence in a Field of Study," CESifo Working Paper Series 9087, CESifo.

Kiefer, A. and M. Shih (2006): "Gender Differences in Persistence and Attributions in Stereotype Relevant Contexts." *Sex Roles*, 54, 859–868.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021): "Choice of majors: are women really different from men?" *Economics of Education Review*, 81, 102079.

Lavy, V. and E. Sand (2018): "On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases," *Journal of Public Economics*, 167, 263–279.

Levine, P. B. and D. J. Zimmerman (1995): "The Benefit of Additional High-School Math and Science Classes for Young Men and Women," *Journal of Business & Economic Statistics*, 13, 137–149.

Long, M. C., D. Conger, and P. Iatarola (2012): "Effects of High School Course-Taking on Secondary and Postsecondary Success," *American Educational Research Journal*, 49, 285–322.

Macartney, H., R. McMillan, and U. Petronijevic (2018): "Teacher Performance and Accountability Incentives," NBER Working Paper 24747, National Bureau of Economic Research.

Main, J. B. and B. Ost (2014): "The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis," *Journal of Economic Education*, 45, 1–10.

May, H., I. Perez-Johnson, J. Haimson, S. Sattar, and P. Gleason (2009): "Using State Tests in Education Experiments: A Discussion of the Issues," (NCEE 2009-013), National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

National Center for Education Statistics (2019): "2015 Survey Questionnaires Results: Classroom Instructions for Mathematics, Reading, and Science (NCES 2018-147)," https://www.nationsreportcard.gov/sq_classroom/#mathematics.

Nord, C., S. Roey, R. Perkins, M. Lyons, N. Lemanski, J. Brown, and J. Schuknecht (2011): "The Nation's Report Card: America's High School Graduates (NCES 2011-462)," .

North Carolina Department of Public Instruction (2013): "North Carolina End-of-Course Tests," http://www.ncpublicschools.org/accountability/testing/eoc/.

——— (2015): "2013–14 School Performance Grades (A–F) for North Carolina Public Schools," https://www.dpi.nc.gov/media/5814/open.

——— (2016a): "North Carolina End-of-Course Assessment of NC Math 1," http://www.ncpublicschools.org/docs/accountability/testing/achievelevels/math1achvlvl16.pdf.

——— (2016b): "The North Carolina Testing Program Technical Report 2012–2015 English Language Arts/Reading Assessments (ELA) End-of-Grade 3–8 and End-of-Course English II," https://www.dpi.nc.gov/media/9619/open.

——— (2016c): "The North Carolina Testing Program Technical Report 2012–2015 Mathematics Assessments End-of-Grade 3–8 and End-of-Course Math I," https://files.nc.gov/dpi/documents/accountability/testing/technotes/mathtechreport1215.pdf.

——— (2019a): "End-of-Grade (EOG)," https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/state-tests/end-grade-eog/.

——— (2019b): "Lexile® and Quantile® Measures in North Carolina," https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/lexiler-and-quantiler-measures-north-carolina.

——— (2022): "Course Code Manual Revised March 2022," https://www.dpi.nc.gov/media/14282/open.

Ost, B. (2010): "The role of peers and grades in determining major persistence in the sciences," *Economics of Education Review*, 29, 923–934.

Owen, A. L. (2010): "Grades, gender, and encouragement: A regression discontinuity analysis," *Journal of Economic Education*, 41, 217–234.

Owen, S. (2021): "Ahead of the Curve Grade Signals, Gender, and College Major Choice," *Working Paper*.

Papay, J. P., R. J. Murnane, and J. B. Willett (2016): "The Impact of Test Score Labels on Human-Capital Investment Decisions," *Journal of Human Resources*, 51, 357–388.

Porter, C. and D. Serra (2020): "Gender Differences in the Choice of Major: The Importance of Female Role Models," *American Economic Journal: Applied Economics*, 12, 226–54.

Rask, K. and J. Tiefenthaler (2008): "The role of grade sensitivity in explaining the gender imbalance in undergraduate economics," *Economics of Education Review*, 27, 676–687.

Roberts, M. R. and T. M. Whited (2013): "Chapter 7 - Endogeneity in Empirical Corporate Finance1," Elsevier, vol. 2 of *Handbook of the Economics of Finance*, 493–572.

Roberts, T.-A. and S. Nolem-Hoeksema (1989): "Sex Differences in Reactions to Evaluative Feedback," *Sex Roles*, 21, 725–747.

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

Shastry, G. K., O. Shurchkov, and L. L. Xia (2020): "Luck or skill: How women and men react to noisy feedback," *Journal of Behavioral and Experimental Economics*, 88, 101592.

Shettle, C., S. Roey, J. Mordica, R. Perkins, C. Nord, J. Teodorovic, J. Brown, M. Lyons, C. Averett, and D. Kastberg (2007): "The Nation's Report Card: America's High School Graduates (NCES 2007-467)," .

Spencer, S. J., C. M. Steele, and D. M. Quinn (1999): "Stereotype Threat and Women's Math Performance," *Journal of Experimental Social Psychology*, 35, 4–28.

Steele, C. M. (1997): "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, 52, 613–629.

Tungodden, J. and A. Willén (2022): "When Parents Decide: Gender Differences in Competitiveness," CESifo Working Paper Series 9516, CESifo.

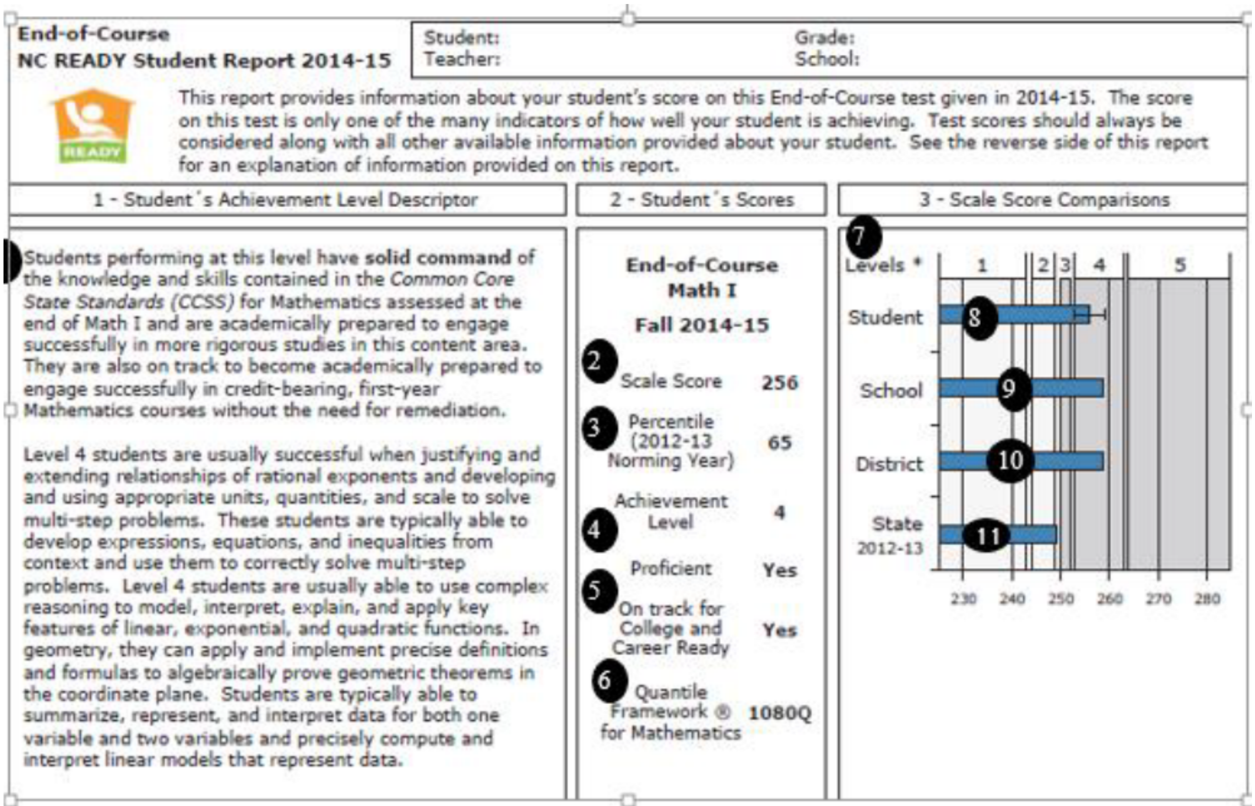
Ugalde A., P. (2024): "Gender, Grade Sensitivity, and Major Choice," Working paper.

Xu, K.-L. (2017): "Regression discontinuity with categorical outcomes," *Journal of Econometrics*, 201, 1–18.

A Appendix

Figure A1: Individual Student Report

(a) Math



End-of-Course
NC READY Student Report 2014-15

Student: Grade:
Teacher: School:

READY

This report provides information about your student's score on this End-of-Course test given in 2014-15. The score on this test is only one of the many indicators of how well your student is achieving. Test scores should always be considered along with all other available information provided about your student. See the reverse side of this report for an explanation of information provided on this report.

1 - Student's Achievement Level Descriptor

Students performing at this level have **solid command** of the knowledge and skills contained in the *Common Core State Standards (CCSS)* for Mathematics assessed at the end of Math I and are academically prepared to engage successfully in more rigorous studies in this content area. They are also on track to become academically prepared to engage successfully in credit-bearing, first-year Mathematics courses without the need for remediation.

Level 4 students are usually successful when justifying and extending relationships of rational exponents and developing and using appropriate units, quantities, and scale to solve multi-step problems. These students are typically able to develop expressions, equations, and inequalities from context and use them to correctly solve multi-step problems. Level 4 students are usually able to use complex reasoning to model, interpret, explain, and apply key features of linear, exponential, and quadratic functions. In geometry, they can apply and implement precise definitions and formulas to algebraically prove geometric theorems in the coordinate plane. Students are typically able to summarize, represent, and interpret data for both one variable and two variables and precisely compute and interpret linear models that represent data.

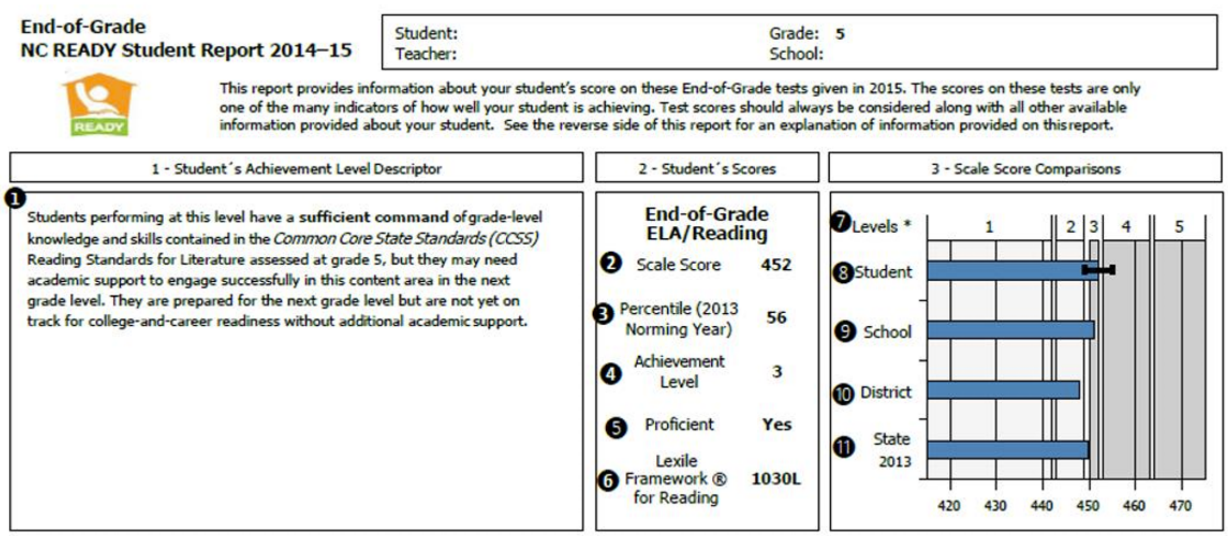
2 - Student's Scores

End-of-Course Math I
Fall 2014-15

2 Scale Score	256
3 Percentile (2012-13 Norming Year)	65
4 Achievement Level	4
Proficient	Yes
5 On track for College and Career Ready	Yes
6 Quantile Framework ® for Mathematics	1080Q

3 - Scale Score Comparisons

(b) English



End-of-Grade
NC READY Student Report 2014–15

Student: Grade: 5
Teacher: School:

READY

This report provides information about your student's score on these End-of-Grade tests given in 2015. The scores on these tests are only one of the many indicators of how well your student is achieving. Test scores should always be considered along with all other available information provided about your student. See the reverse side of this report for an explanation of information provided on this report.

1 - Student's Achievement Level Descriptor

1 Students performing at this level have a **sufficient command** of grade-level knowledge and skills contained in the *Common Core State Standards (CCSS)* Reading Standards for Literature assessed at grade 5, but they may need academic support to engage successfully in this content area in the next grade level. They are prepared for the next grade level but are not yet on track for college-and-career readiness without additional academic support.

2 - Student's Scores

End-of-Grade ELA/Reading

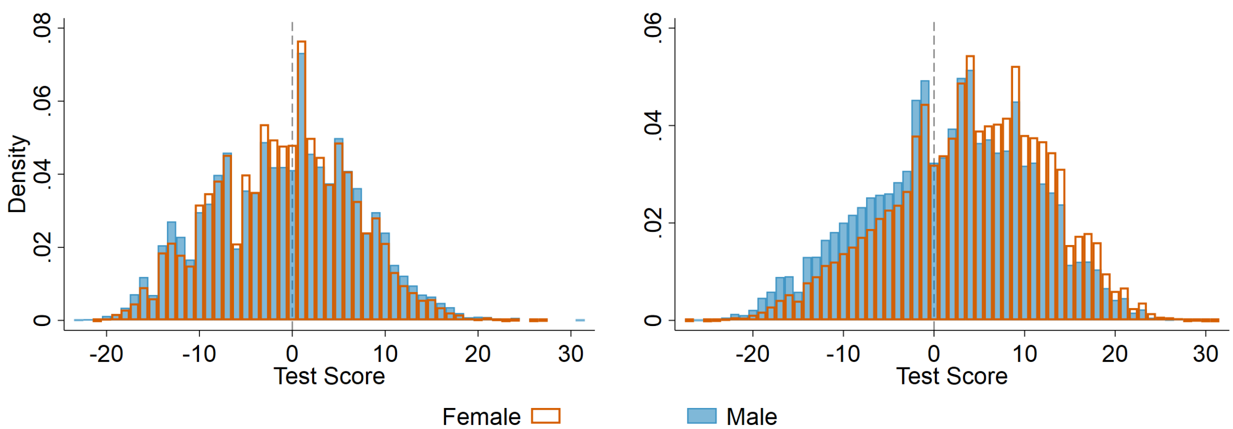
2 Scale Score	452
3 Percentile (2013 Norming Year)	56
4 Achievement Level	3
5 Proficient	Yes
6 Lexile Framework ® for Reading	1030L

3 - Scale Score Comparisons

Source: North Carolina Department of Public Instruction (2016b,c)

Figure A2: Test Scores Distributions

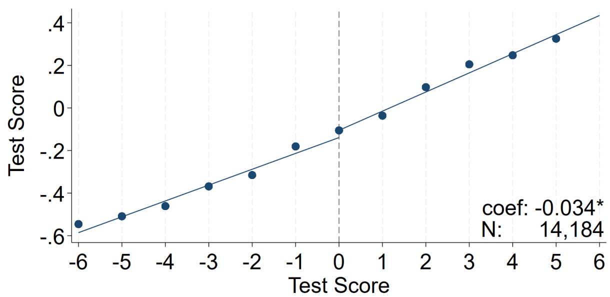
(a) Math 1. (b) English 2.



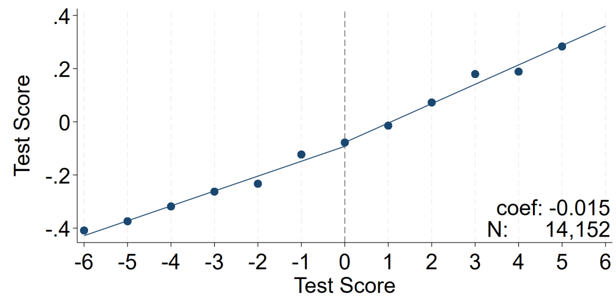
Note: Scores are normalized such that a score of 0 or more means proficiency.

Figure A3: Continuity of Covariates for Math 1 Test

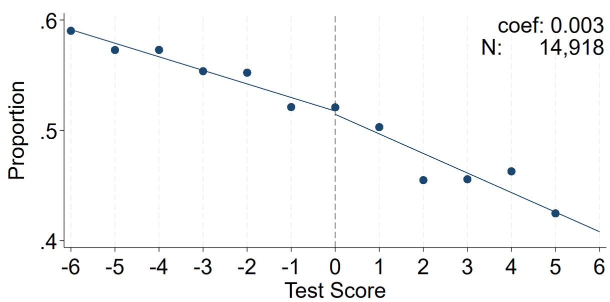
(a) Middle School Math Test Score



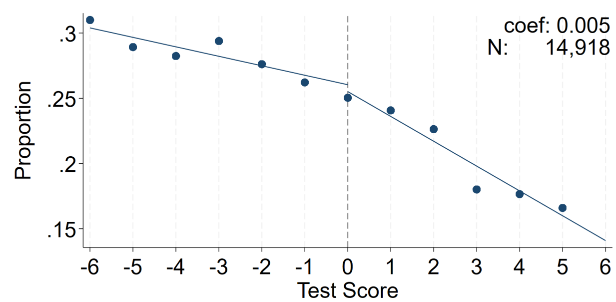
(b) Middle School Reading Test Score



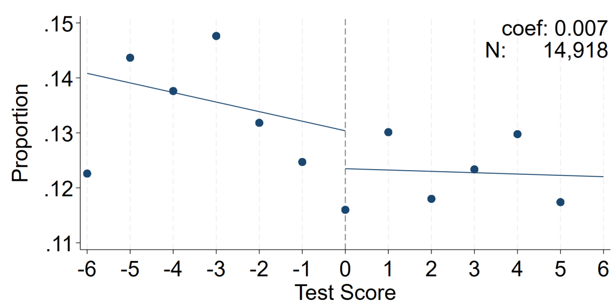
(c) EDS



(d) Black

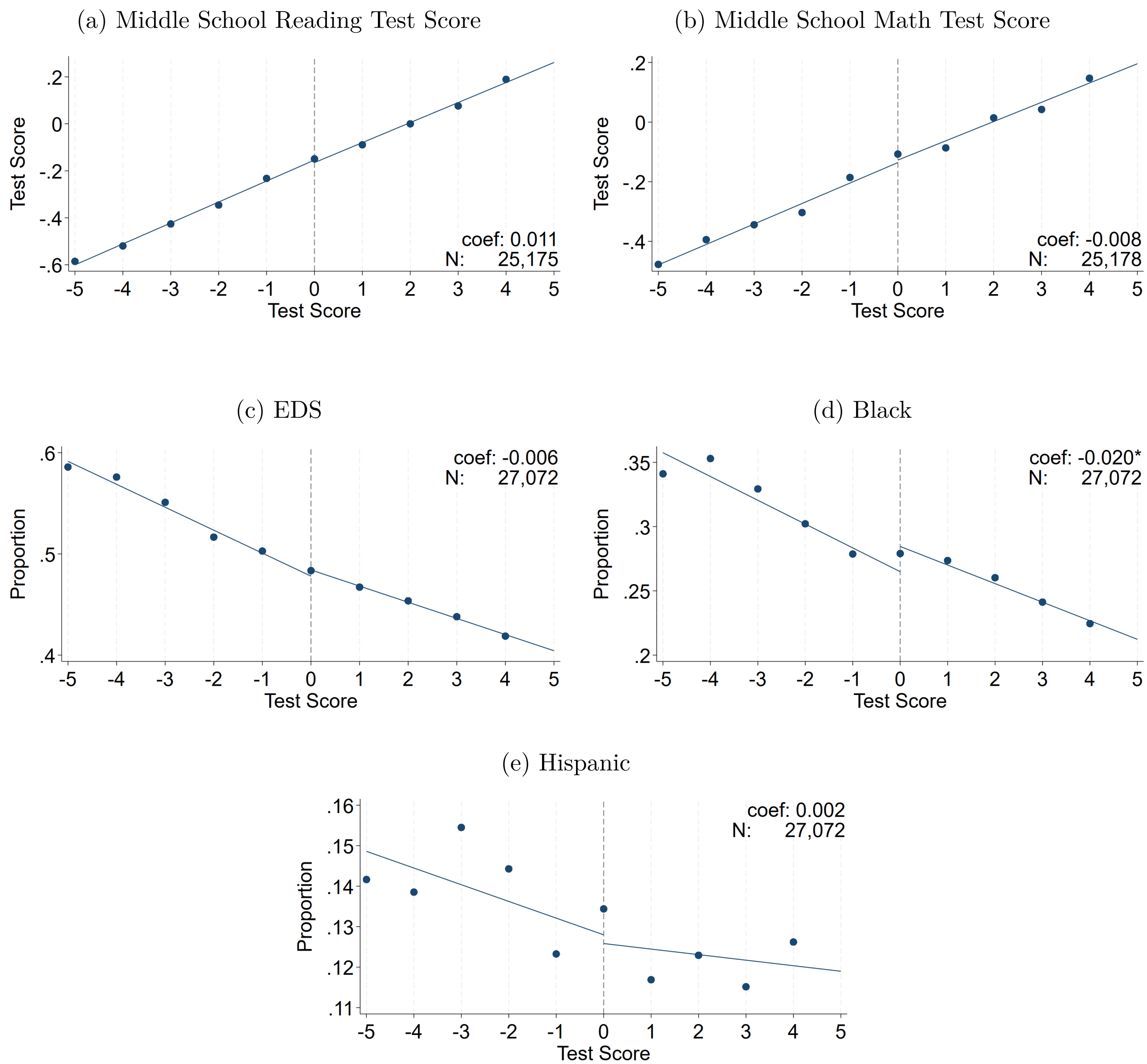


(e) Hispanic



Note: Math test scores are normalized such that a score of 0 or more means proficiency. Lines are fitted values from regressing the corresponding covariate on an indicator for being non-proficient on the Math 1 test. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold, within ±6 of the cutoff. Standard errors clustered at school level. Middle school test scores are standardized by year. The dots are averages within each 1 point bin. Minority: black or hispanic students. EDS: Economically Disadvantaged Student. Coef: estimated discontinuity, i.e. the coefficient for non-proficiency. *Significant at 10%, **5%, ***1%.

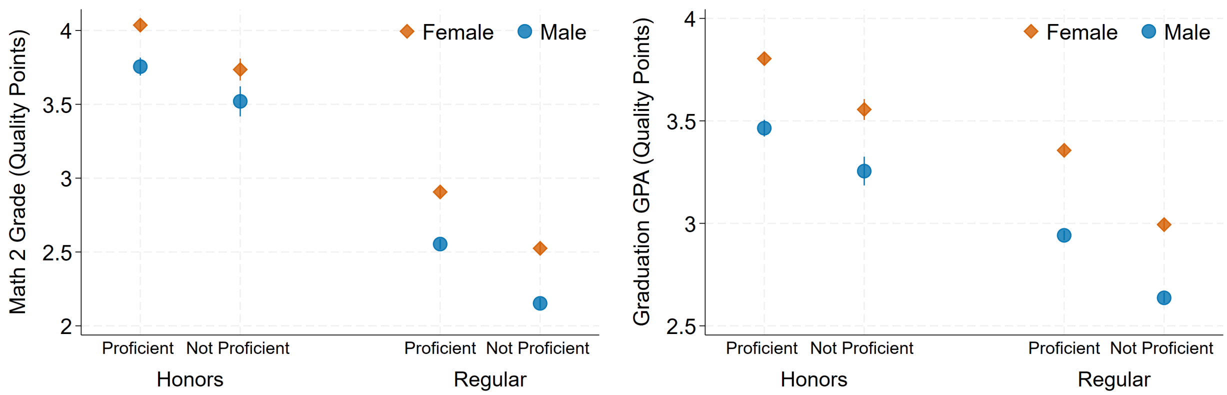
Figure A4: Continuity of Covariates for English 2 Test



Note: English test scores are normalized such that a score of 0 or more means proficiency. Lines are fitted values from regressing the corresponding covariate on an indicator for being non-proficient on the English 2 test. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold, within ±4 of the cutoff. Standard errors clustered at school level. Middle school test scores are standardized by year. The dots are averages within each 1 point bin. Minority: black or hispanic students. EDS: Economically Disadvantaged Student. Coef: estimated discontinuity, i.e. the coefficient for non-proficiency. *Significant at 10%, **5%, ***1%.

Figure A5: Outcomes by Proficiency Level Math 1 EOC test and Course Enrollment Type

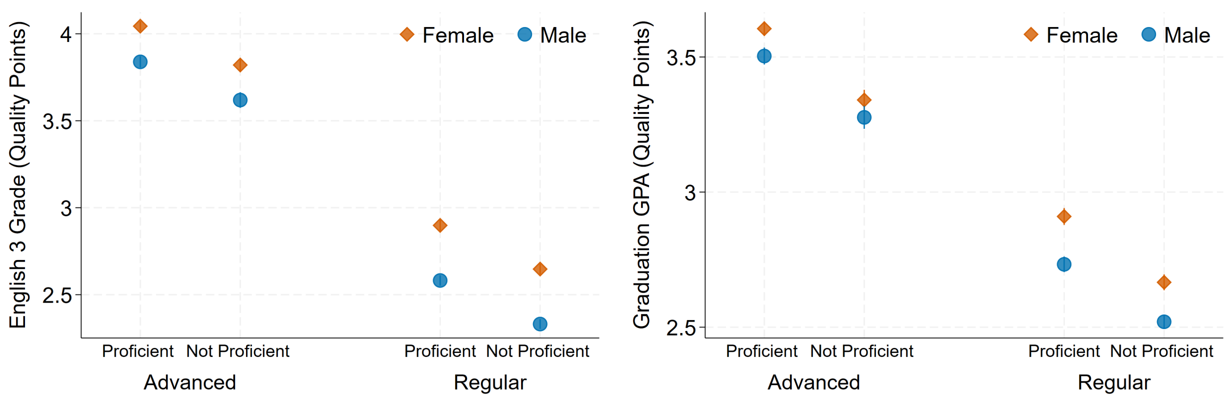
(a) Math 2 Final Grade (b) High School GPA



Note: Markers represent the average final grade and graduation GPA considering quality points for the students in the math sample that score within ±6 from the proficiency threshold in their Math 1 EOC test. Spikes represent 95% confidence intervals.

Figure A6: Outcomes by Proficiency Level in English 2 EOC test and Course Enrollment Type

(a) English 3 Final Grade (b) High School GPA



Note: Markers represent the average of the variable final grade and graduation GPA considering quality points for the students in the English sample that score within ±5 from the proficiency threshold in their English 2 EOC test. Spikes represent 95% confidence intervals.

Table A1: Comparison Between Fall, After Fall and Middle School Samples

	Fall	After Fall (AF)	Difference Fall - AF	Middle School (MS)	Difference Fall - MS
	(1)	(2)	(3)	(4)	(5)
Female	0.51	0.47	0.03***	0.51	-0.00
Black	0.26	0.33	-0.07***	0.18	0.08***
Hispanic	0.13	0.15	-0.03***	0.10	0.03***
White	0.54	0.45	0.10***	0.64	-0.10***
EDS	0.51	0.59	-0.08***	0.31	0.20***
Non-Prof. Math 1 test	0.49	0.62	-0.13***	0.24	0.25***
Non-Prof. MS Math test	0.63	0.68	-0.05***	0.23	0.39***
Non-Prof. MS Reading test	0.53	0.57	-0.04***	0.18	0.34***
N	27,997	129,541		76,505	

Note: Table presents sample proportions of variables of interest. Fall sample refers to the students that took the Math 1 class during the fall semester of their freshman year. After Fall sample refers to the students that took Math 1 at any other time after the fall semester of their freshman year. Middle School sample refers to the students that took Math 1 during middle school. The difference columns show the mean difference between the Fall sample and the After Fall sample or Middle School Sample, respectively, for each variable and its significance. Non-Prof.: non-proficiency, MS: middle school, EDS: Economically Disadvantaged Student. *Significant at 10%, **5%, ***1%.

Table A2: Comparison Between Fall and After Fall Samples

	Fall	After Fall	Difference
Female	0.51	0.49	0.02***
Black	0.24	0.27	-0.03***
Hispanic	0.11	0.12	-0.01***
White	0.58	0.52	0.05***
EDS	0.42	0.46	-0.04***
Non-Prof. English 2 test	0.35	0.34	0.01**
Non-Prof. Middle School Math test	0.35	0.32	0.03***
Non-Prof. Middle School Reading test	0.38	0.35	0.03***
N	72,395	143,057	

Note: Table presents sample proportions of variables of interest. Fall sample refers to the students that took the English 2 class during the fall semester of their sophomore year. After Fall sample refers to the students that took English 2 at any other time after the fall semester of their sophomore year. The difference column shows the mean difference between the two samples for each variable and its significance. Non-Prof.: non-proficiency, EDS: Economically Disadvantaged Student. *Significant at 10%, **5%, ***1%.

Table A3: Probability of Enrolling in Advanced Math and English Classes

	Honors Math 2		Advanced English 3	
	(1)	(2)	(3)	(4)
Female (F)	0.125***	0.114***	0.111***	0.135***
	(0.010)	(0.010)	(0.008)	(0.008)
Non-Prof.	0.001	-0.000	0.011	0.013
	(0.013)	(0.012)	(0.011)	(0.010)
F*Non-Prof.	-0.075***	-0.061***	-0.026**	-0.027**
	(0.012)	(0.012)	(0.011)	(0.011)
Math Test MS		0.095***		0.129***
		(0.006)		(0.004)
Reading Test MS		0.006		0.067***
		(0.005)		(0.005)
Controls		✓		✓
Bandwidth	±7		±6	
Mean	0.23	0.24	0.42	0.42
R^2	0.08	0.25	0.08	0.29
N	17,339	16,417	31,495	29,199

Note: In columns (1)-(2) the dependent is one when taking Honors Math 2, zero otherwise. In columns (3)-(4), it is one when taking advanced English 3 class, zero otherwise. Each specification includes indicator variables for being a female (F), for being non-proficient (Non-Prof) on the Math 1 EOC test (columns (1)-(2)) or English 2 EOC test (Columns (3)-(4)); and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold. Reading Test MS: standardized middle school reading test score. Math test MS: standardized middle school math test score. Controls include school and year FE, race, and economically disadvantage status. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A4: Probability of Taking Honors Math 2. ±6 bandwidth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female (F)	0.119***	0.108***	0.108***	0.108***	0.107***	0.108***	0.108***	0.110***
	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
Non-Prof.	-0.005	-0.006	-0.006	-0.006	-0.006	-0.007	-0.007	-0.008
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
F*Non-Prof.	-0.065***	-0.054***	-0.053***	-0.052***	-0.052***	-0.049***	-0.049***	-0.050***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Math Test MS						0.101***	0.100***	
						(0.007)	(0.007)	
Reading Test MS							0.003	
							(0.005)	
(Math-Reading) Test MS								0.028***
								(0.004)
School FE		✓	✓	✓	✓	✓	✓	✓
Year FE			✓	✓	✓	✓	✓	✓
EDS				✓	✓	✓	✓	✓
Race					✓	✓	✓	✓
Mean	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
R^2	0.06	0.22	0.22	0.23	0.23	0.25	0.25	0.23
N	14,918	14,918	14,918	14,918	14,918	14,184	14,146	14,146

Note: The dependent variable is the same across all specifications: one when taking honors Math 2, zero otherwise. Each specification includes indicator variables for being a female (F), for being non-proficient (Non-Prof) on the Math 1 EOC test, and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold. Reading Test MS: standardized middle school reading test score. Math test MS: standardized middle school math test score. (Math-Reading) Test MS: difference between the standardized math and reading middle school test scores. School FE: school fixed effects. EDS: Economically Disadvantaged Student. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A5: Probability of Taking Honors Math 2, Standard Errors Robustness

	(1)	(2)	(3)	(4)
Female (F)	0.108***	0.108***	0.108***	0.108***
	(0.009)	(0.011)	(0.010)	(0.011)
Non-Prof.	-0.007	-0.007	-0.007	-0.008
	(0.013)	(0.015)	(0.013)	(0.014)
F*Non-Prof.	-0.049***	-0.049***	-0.049***	-0.047***
	(0.012)	(0.015)	(0.013)	(0.014)
SE	Robust	Clustered District	Clustered Test Score	Clustered School
Mean	0.23	0.23	0.23	0.24
R^2	0.25	0.25	0.25	0.24
N	14,146	14,146	14,146	13,611

Note: The dependent variable is the same across all specifications: one when taking honors Math 2, zero otherwise. Each column follows the same specification: dependent variable regressed on an indicator equal one when the student is female, an indicator equal to one when the student is deemed non-proficient on the Math 1 EOC test, the interaction of those two and a first degree polymonial of the test score with flexible slopes above and below the proficiency cutoff. All columns include year and high school fixed effects, controls for EDS, race, and middle school test scores. EDS: Economically Disadvantaged Student. Column (4) uses the same specification, but test scores are standardized by year and recentered such that the cutoff is at zero. Bandwidth of ±6 for columns (1)-(3), and a bandwidth of ±0.1261 in column (4). Standard errors are in parentheses. *Significant at 10%, **5%, ***1%.

Table A6: Probability of Taking Advanced English 3. ±5 bandwidth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female (F)	0.108***	0.097***	0.097***	0.103***	0.104***	0.118***	0.134***	0.111***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Non-Prof.	0.008	0.011	0.011	0.009	0.009	0.009	0.011	0.010
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
F*Non-Prof.	-0.018	-0.022**	-0.022**	-0.020*	-0.020*	-0.024**	-0.021*	-0.023*
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Reading Test MS						0.119***	0.070***	
						(0.005)	(0.005)	
Math Test MS							0.128***	
							(0.005)	
(Math-Reading) Test MS								0.054***
								(0.004)
School FE		✓	✓	✓	✓	✓	✓	✓
Year FE			✓	✓	✓	✓	✓	✓
EDS				✓	✓	✓	✓	✓
Race					✓	✓	✓	✓
Mean	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
R^2	0.06	0.21	0.21	0.23	0.23	0.25	0.28	0.24
N	27,072	27,072	27,072	27,072		25,175	25,137	25,137

Note: The dependent variable is the same across all specifications: one when taking an advanced English 3 class, zero otherwise. Each specification includes indicator variables for being a female (F), for being non-proficient (Non-Prof) on the English 2 EOC test; and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold. Reading Test MS: standardized middle school reading test score. Math test MS: standardized middle school math test score. (Math-Reading) Test MS: difference between the standardized math and English middle school test scores. School FE: school fixed effects. EDS: Economically Disadvantaged Student. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A7: Probability of Taking Advanced English 3, Standard Errors Robustness

	(1)	(2)	(3)	(4)
Female (F)	0.134***	0.134***	0.134***	0.136***
	(0.007)	(0.009)	(0.005)	(0.008)
Non-Prof.	0.011	0.011	0.011	0.013
	(0.012)	(0.011)	(0.007)	(0.011)
F*Non-Prof.	-0.021**	-0.021*	-0.021	-0.023**
	(0.011)	(0.012)	(0.012)	(0.011)
SE	Robust	Clustered District	Clustered Test Score	Clustered School
Mean	0.42	0.42	0.42	0.44
R^2	0.28	0.28	0.28	0.28
N	25,137	25,137	25,137	27,650

Note: The dependent variable is the same across all specifications: one when taking an advanced English 3 class, zero otherwise. Each column follows the same specification: dependent variable regressed on an indicator equal one when the student is female, an indicator equal to one when the student is deemed non-proficient on the English 2 EOC test, the interaction of those two and a first degree polymonial of the test score with flexible slopes above and below the proficiency cutoff. All columns include year and high school fixed effects, controls for EDS, race, and middle school test scores. EDS: Economically Disadvantaged Student. Column (4) uses the same specification, but test scores are standardized by year and recentered such that the cutoff is at zero. Bandwidth of ± 5 for columns (1)-(3), and a bandwidth of ± 0.2743 in column (4). Standard errors are in parentheses. *Significant at 10%, **5%, ***1%.

Table A8: Probability of Enrolling in Advanced Math and English Classes

	Honors Math 2		Advanced English 3	
	(1)	(2)	(3)	(4)
Female (F)	0.104***	0.105***	0.131***	0.131***
	(0.011)	(0.011)	(0.008)	(0.009)
Non-Prof.	-0.009	-0.009	0.006	0.007
	(0.015)	(0.015)	(0.012)	(0.012)
F*Non-Prof.	-0.045***	-0.046***	-0.020*	-0.021*
	(0.014)	(0.014)	(0.012)	(0.012)
Only Passing Students		✓		✓
Bandwidth	±6		±5	
Mean	0.23	0.23	0.42	0.42
R^2	0.25	0.25	0.28	0.28
N	12,812	12,667	22,928	22,618

Note: In columns (1)-(2) the dependent is one when taking Honors Math 2, zero otherwise. In columns (3)-(4), it is one when taking advanced English 3 class, zero otherwise. Each specification includes indicator variables for being a female (F), for being non-proficient (Non-Prof) on the Math 1 EOC test (columns (1)-(2)) or English 2 EOC test (Columns (3)-(4)); and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the proficiency threshold, high school and year FE, indicators for EDS, race, and middle school test scores. Sample restricted to students without missing course grade. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A9: Probability of Enrolling Honors Math 2 for Different Thresholds

Levels	1 vs. 2	2 vs.3	3 vs. 4	4 vs. 5
	(1)	(2)	(3)	(4)
Female (F)	0.052***	0.101***	0.129***	0.126**
	(0.010)	(0.014)	(0.012)	(0.051)
Below Cutoff	0.021*	-0.009	0.003	0.047
	(0.012)	(0.015)	(0.022)	(0.063)
F*Below Cutoff	-0.031**	-0.046***	-0.026	-0.031
	(0.013)	(0.015)	(0.018)	(0.057)
Bandwidth	±5	(-6,3)	(-3,5)	±3
Mean	0.11	0.19	0.34	0.66
R^2	0.23	0.24	0.25	0.33
N	9,214	10,672	9,954	1,334

Note: In all columns the dependent variable is one when taking Honors Math 2, zero otherwise. Each specification includes indicator variables for being a female (F), for being below the test score cutoff between the two performance levels indicated at the top on the Math 1 EOC test; and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the threshold, high school and year FE, indicators for EDS, race, and middle school test scores. In columns (2) and (3) the right and left bandwidths, respectively, are such that there is not overlap of observations between the different achivement levels. The bandwidths for columns (1) and (4) and left and right bandwidths for columns (2) and (3) are estimated following the same proceduce described in section 3. Column (3) presents the results for the threshold that determines the proficiency label. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A10: Probability of Enrolling Advanced English 3 for Different Thresholds

Levels	1 vs. 2	2 vs.3	3 vs. 4	4 vs. 5
	(1)	(2)	(3)	(4)
Female (F)	0.091***	0.134***	0.139***	0.031***
	(0.008)	(0.010)	(0.008)	(0.008)
Below Cutoff	0.009	0.012	-0.017	-0.027*
	(0.012)	(0.013)	(0.016)	(0.014)
F*Below Cutoff	-0.032***	-0.025*	0.002	0.024**
	(0.011)	(0.013)	(0.012)	(0.011)
Bandwidth	±6	(-5,3)	(-3,6)	±4
Mean	0.21	0.36	0.58	0.93
R^2	0.24	0.26	0.27	0.19
N	17,149	18,246	24,065	8,429

Note: In all columns the dependent variable is one when taking Advanced English 3, zero otherwise. Each specification includes indicator variables for being a female (F), for being below the test score cutoff between the two performance levels indicated at the top on the English 2 EOC test; and the interaction of those two variables. Additionally all specifications include a first-degree polynomial of the running variable with different slopes above and below the threshold, high school and year FE, indicators for EDS, race, and middle school test scores. In columns (2) and (3) the right and left bandwidths, respectively, are such that there is not overlap of observations between the different achivement levels. The bandwidths for columns (1) and (4) and left and right bandwidths for columns (2) and (3) are estimated following the same proceduce described in section 3. Column (3) presents the results for the threshold that determines the proficiency label. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.

Table A11: Effect of Non-Proficiency on Graduation GPA.

	Math 1 EOC Test		English 2 EOC Test	
	Graduation GPA	Graduation GPA (Quality Points)	Graduation GPA	Graduation GPA (Quality Points)
	(1)	(2)	(3)	(4)
Female (F)	0.307***	0.416***	0.248***	0.320***
	(0.014)	(0.016)	(0.014)	(0.015)
Non-Prof.	-0.022	-0.027	-0.019	-0.020
	(0.023)	(0.027)	(0.022)	(0.023)
F*Non-Prof.	-0.025	-0.056**	-0.006	-0.027
	(0.020)	(0.022)	(0.021)	(0.022)
Bandwidth	±5	±5	±5	±5
Mean	2.86	3.12	2.75	3.03
R^2	0.33	0.36	0.38	0.42
N	9,898	11,031	9,964	12,143

Note: Dependent variable indicated at the top of each column. Graduation GPA (Quality Points) takes into account the quality points that students in advanced and honors courses get due to the higher difficulty of those classes. Each column follows the same specification: dependent variable regressed on an indicator equal one when the student is female, an indicator equal to one when the student is deemed non-proficient in the respective EOC test indicated at the top, the interaction of those two and a first degree polymonial of the test score with flexible slopes above and below the proficiency cutoff. All columns include year and high school fixed effects, controls for EDS, race, and middle school test scores. EDS: Economically Disadvantaged Student. Standard errors clustered at school level are reported in parentheses. *Significant at 10%, **5%, ***1%.