Targeted Acceleration in Middle School Math: Impacts on College Entry, Degree Completion, and STEM

David Card and Laura Giuliano*

February 2025

Abstract

The traditional mathematics sequence in U.S. schools leaves many students ill-prepared to enter selective engineering and science programs. Attempts in several states to begin algebra in middle school have faltered; the latest Common Core curriculum delays algebra until high school, leaving no room for calculus by twelfth grade. We study a *targeted* multi-year accelerated curriculum, offered in all middle schools of a large urban school district, that enables students to finish Algebra I and Geometry before high school. Entry to the program is based on 5th-grade scores in the statewide math test, with a cutoff at the 80th percentile, enabling a highly credible regression discontinuity analysis. Following students for 13 years after entry, we show that the accelerated curriculum has significant positive effects on the probabilities that marginally eligible girls enter selective colleges, complete a college degree within 6 years, and major in STEM or business/economics. In contrast, we find negligible effects on the degree completion or major choices of boys, implying that GEM brings girls up to the level of similar-scoring boys in attainment of degrees in the most highly-paid fields. To address concerns that the introduction of an accelerated track hurts lower-scoring students, we compare students with similar math scores in our district and those in a nearby district without an accelerated program. We find no evidence of negative (or positive) spillovers on students with scores below the cutoff.

* Card: University of California, Berkeley and NBER, card@berkeley.edu; Giuliano: University of California, Santa Cruz and NBER, lgiulian@ucsc.edu. We are extremely grateful to Cynthia Park for assistance in accessing and interpreting the data used in this study, and to Tatiana Reyes for extraordinary research assistance. For helpful comments, we also thank Sarah Cohodes, Serena Canaan, Glenn Ellison, seminar participants at UC Berkeley, and session participants at the 2024 NBER Economics of Talent Meeting and 2025 AEA Annual Meeting. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A190175 to the NBER. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

The question of when students should begin the study of algebra is highly controversial. It also matters: UC Berkeley's College of Engineering, for example, tells prospective students that "*Four years of mathematics, especially including calculus*" demonstrates the necessary aptitude in science and math to be considered for admission.[1] Unfortunately, the traditional U.S. high school math sequence of Algebra I, Geometry, Algebra II, and Pre-calculus leaves no room for calculus by the end of high school, let alone in time to be included in a college application. The conflict between the "hidden curriculum" required by many selective college programs and the traditional math curriculum has led to wide debate and policy swings. In the 2000's several states including California began promoting "Algebra for All" in 8th grade.[2] Just a few years later, this effort was abandoned. More recently some districts have tried to ***delay*** algebra for even high-ability students until 9th grade, citing concerns that tracking harms less advanced students (see e.g., Huffaker et al. 2024 and Huffaker, 2023). Nevertheless, the demand for advanced math in middle school is strong (Yoder, 2024), and enrollment in Algebra and Geometry in 7th and 8th grades remains relatively high in many states.[3]

In this paper we study the impacts of a targeted accelerated math curriculum available in the middle schools of a large urban school district in Florida ("the District"). The curriculum, known as "Great Explorations in Mathematics" (GEM) includes pre-algebra in 6th grade, high-school algebra in 7th grade, and high-school geometry in 8th grade, effectively accelerating students two years beyond the traditional curriculum, and one year beyond the "advanced" tracks offered in many schools, including those in the District and other districts in Florida. Eligibility for the GEM program is based on scores in the state's 5th-grade math test, with a threshold at roughly the 80th percentile. Thus, marginally eligible participants – for whom the causal effects of GEM participation are identified in a regression discontinuity (RD) analysis – are relatively

[1] This information is from the front page of the website https://engineering.berkeley.edu/admissions/undergraduate-admissions/ retrieved August 20, 2024.

[2] The National Mathematics Advisory Panel (2008) recommended that "School districts should ensure that all prepared students have access to an authentic algebra course and should prepare more students than at present to enroll in such a course by Grade 8. See Domina et al. (2015) for an analysis of the California initiative, which is widely perceived as having hurt lower-ability students. Similar findings are reported by Clotfelter et al. (2015) from an analysis of an 8th grade algebra initiative in North Carolina in the early 2000s.

[3] For example, Peters and Carter (2023, Table 2) report that average enrollment rates in Geometry in 8th grade in 2017-28 were around 18% in Utah, 13% in Virginia, and 8-9% in Florida, Maryland, Minnesota, and Washington. At the opposite extreme the rate was below 0.5% in Arkansas, Alabama, Louisiana, Mississippi, and West Virginia.

well qualified.[4] Indeed, nearly all the students who barely miss the GEM threshold will enter an advanced math track in middle school.

We use information on middle and high school course-taking and grades, combined with information from the National Student Clearinghouse on college enrollment and degree completion, to study the effects of entry to the GEM accelerated mathematics sequence in 6th grade. Focusing on students whose 5th-grade scores are within a narrow bandwidth ($\pm 0.6\sigma's$) of the entry threshold, we follow 23,000 students for 13 years after 5th grade (i.e., allowing up to 6 years after they would normally graduate from high school). This relatively large sample enables us to conduct an informative RD analysis and make meaningful comparisons of the effects of GEM along two key dimensions that have been identified in previous work: gender, and participation in the free or reduced price lunch (FRL) program.[5] We also compare students with higher and lower predicted probabilities of completing all three years of the program.

For students as a whole, we find that entry to the GEM accelerated program leads to large gains in the probability of taking algebra and geometry in middle school (+20 percentage points (ppt) and +43 ppt respectively); a smaller but still highly significant +9 ppt rise in the probability of taking and passing calculus in high school; and a marginally significant +3.7 ppt effect on the probability of obtaining a bachelor's degree in STEM or business/economics fields within 6 years after high school. Importantly, ***all*** of the effects on college-related outcomes are attributable to girls. Entry to GEM leads to large increase in the probability that girls complete a degree in a STEM field (from 8% to 13%), pushing their rate up to that of boys. There is a similar gain in majoring in business or economics (from 7% to 11%), again catching up to the level of boys. In contrast, the estimated impacts on college-related outcomes of boys are small and sometimes even negative, though never statistically significant.

GEM is demanding: only about one-half of marginally qualified girls, and one-third of boys, complete all three years of the program. Classifying students by their predicted probability of completing the program, we find similar impacts on degree completion and STEM/business majoring for lower- versus higher-probability girls, but no systematic effects on either group of

[4] The closest antecedent to our work is a recent study by Dougherty et al. (2017), which evaluates test-based admission to Algebra in 8th grade for students in Wake County NC. The threshold for this program is set at about the 20th percentile of the state test.

[5] Dougherty et al. (2017) find that admission to algebra in 8th grade has a relatively large and significant effect on the probability of taking pre-calculus in 11th grade for girls, but no effect for boys. They also find that the gains are concentrated among higher income students and are negligible for lower-income students.

boys. Thus the essential heterogeneity in the program's impact appears to be gender, rather than better or worse matching to the demands of the curriculum.

A concern for our RD estimates is that students who barely *miss* the GEM threshold may experience a causal *reduction* in their academic performance, arising from the removal of better-performing peers from their math classes (Payolo, 2020), or from discouragement effects associated with missing the cutoff (Ellison and Swanson, 2022).[6] To address this, we use data from another large urban school district in the same state that has no GEM program and compare boys and girls with similar 5th-grade math scores in the two districts. At scores below the GEM threshold, post-secondary outcomes, including rates of STEM and business degree completion, are very similar in the two districts, with no evidence of negative (or positive) effects for District students of either gender who just miss the threshold. But at scores above the threshold, girls in the District see sizeable gains relative to those in the comparison district, consistent with our RD estimates.

Our empirical findings contribute to two main literatures. The first is a relatively small set of studies that have tried to identify the causal effects of math acceleration on student outcomes. Research on the Algebra-for-All initiative in California (Domina et al, 2015) and on a related initiative in North Carolina (Clotfelter et al., 2015) showed that pushing all students into algebra in 8th grade was harmful for weaker students – an effect that led to the suspension of the initiatives. Subsequent research by Dougherty et al. (2017) showed that a more targeted approach – leaving out the lowest-scoring quintile of students – leads to a positive effect on math course-taking of girls, including enrollment in pre-calculus in 11th grade, with little effect (positive or negative) on boys. (They do not follow students to college). Cohodes (2020) examines an accelerated curriculum program for high-achievers in 4th-6th grades and finds positive effects on high school graduation and college entry. She does not follow students long enough to measure degree completion, nor does she present results separately by gender.[7]

[6] Alternatively, students who barely miss the threshold may experience an increase in performance arising from their new status as the top students in their math class – see Denning et al. (2023) for example.

[7] An older study by Altonji (1996) studied the effects of high school course taking on college completion and wages. He used the mean shares of students taking different courses in a students' high school as instrumental variables, potentially identifying the effects of access to course options. He found relatively small effects of math or science courses on college outcomes or wages.

We also contribute to a much bigger literature on gender differences in college major choice (see McNally 2020 for a recent review).[8] Women and men choose different fields of study in college, with a lower share of women graduating in science, technology, engineering and math (STEM) fields, and in business-related fields. As noted by Altonji et al. (2016), graduates of these fields are among the most highly paid college workers, and observational comparisons suggest that differences in field of study can explain about one fifth of the gender earnings gap among college graduates in the US.[9] While there is a positive correlation between math preparation in high school and entry to STEM fields in college (see e.g., Speer, 2017; Card and Payne, 2021), there is almost no causal evidence that *increases* in math preparation lead to a *rise* in STEM majoring by young women.[10] Our results suggest that exposure to an accelerated math curriculum in middle school can help close the gender gap in both STEM and business-related degree attainment.

I. Background: District Setting and the GEM Program

The District is one of the 10 largest school districts in the country, serving around 20,000 students per grade. Its student body is racially diverse, with roughly equal numbers of Hispanic, Black and White students, just under one-half of whom are eligible for the FRL program. During the past two decades the District operated approximately 140 elementary schools with grades K-5, about 80 mainstream middle schools offering grades 6-8, and 70 conventional high schools. It also has a wide variety of charter and special service schools.

The District's mainstream middle schools offer three primary tracks of mathematics instruction (taught in separate classrooms): regular, advanced and GEM. There is also a course stream with additional instructional time for weaker students. As shown in the flowchart in Figure 1, both the regular and advanced streams follow the standard math curriculum prescribed

[8] Recent studies include Ceci et al. (2009), Ceci et al. (2014), Perez-Felkner et al. (2014), Dasgupta and Stout (2014), Wiswall et al. (2014), Sax et al. (2016), Ellis et al. (2017), Sassler et al. (2017), Brenoe and Zolitz (2018), Charlesworth and Banaji (20190), Cimpian et al. (2020), Cimpian and King (2024) and Goulas et al. (2024). Some of these studies focus on a narrower subset of STEM subjects (physics, engineering and computer science), which are the fields with the lowest shares of female students.

[9] See Appendix Table A1, where we present a simple analysis of data on 31-35 year old college graduates in the 2010-2018 American Community Survey. We show that adding two dummies that indicate whether a college graduate has a degree in STEM or business-related fields reduces the gender gap in earnings by 14%.

[10] Owen (2023) conducts an RCT among college students informing them of their relative performance on STEM-antecedent courses but finds no effect on major choice.

by the State of Florida in grades 6-8.[11] Students in the advanced track who perform well in 7th-grade math, however, can take Algebra I in 8th grade. In addition, since the 2000-2001 school year, the District's middle schools have offered the Great Explorations in Mathematics (GEM) track, which compresses the standard 3-year middle school math curriculum into 6th grade, and offers Algebra I in 7th grade and Geometry in 8th grade. The availability of GEM in all its mainstream middle schools (including those located in lower-income and minority neighborhoods) reflects the District's long-term effort to expand access to advanced academic programs for all children.

To qualify for GEM, students must have a 5th-grade score of 380 or above in Math and 300 or above in Reading on the Florida Comprehensive Assessment Tests (FCAT's).[12] Entry to GEM also requires a teacher's recommendation and parental consent. Students must earn at least a B grade to stay in the program in 7th and 8th grades; otherwise they return to the advanced track (as shown in Figure 1). This "off ramp" feature helps address concerns that marginal students may be unable to keep up with the faster pace of GEM. Even those who only spend a year in GEM, however, have an extra pathway to completing Algebra 1 by the end of middle school relative to those who miss the cutoff and enter the advanced track in 6th grade (see Figure 1).

II. Empirical Framework

a. Regression Discontinuity Model

We use a "fuzzy RD" approach to analyze the effect of entering GEM in 6th grade on subsequent student outcomes (Lee and Lemieux, 2010). Specifically, we estimate a locally linear first-stage model for entering GEM, with the 5th-grade math score as a running variable and a potential jump at the GEM cutoff. We estimate a parallel set of reduced-form models for outcomes like college enrollment and degree attainment. Under standard assumptions (e.g., Hahn, Todd and van der Klaauw, 2001), the ratio of the estimated discontinuity in the outcome variable at the GEM cutoff to the estimated discontinuity in the fraction of students who enter

[11] The content of the regular and advanced courses is based on a curricular planning system operated by the State of Florida, known as CPALMS. The "advanced" track courses in the Flowchart (which was assembled by the District) correspond to "accelerated" courses in current CPALMS terminology.

[12] The statewide tests have changed over time but the GEM cutoff in math has remained at about the 80th percentile of the math test, while the minimum standard for the reading test has remained at around the median.

GEM provides a causal estimate of the effect of GEM entry on students whose GEM entry status switches as their math score crosses the 380 point threshold.

Specifically, our first-stage model takes the form:

$$GEM6_i = \pi \cdot 1[M_i \geq 380] + g(M_i) + u_i, \quad (1)$$

where $GEM6_i$ is an indicator that student i entered GEM in 6th grade and $g(M_i)$ is a piece-wise linear function of the student's math score (M_i) that allows a slope change at $M_i = 380$:

$$g(M_i) = g_0 + g_1(M_i - 380) + g_2(M_i - 380) \times 1[M_i \geq 380].$$

We estimate (1) using data for students with math scores in a symmetric bandwidth BW around the 380 point threshold, i.e., using students with scores in the interval $[380 - BW, 380 + BW]$.

Likewise, we estimate reduced-form models for the jump in an outcome of interest, Y_i, at the 380-point threshold of the form:

$$Y_i = \delta \cdot 1[M_i \geq 380] + h(M_i) + v_i, \quad (2)$$

where $h(M_i)$ is a piece-wise linear function analogous to $g(M_i)$. Our IV estimator of the treatment effect of gifted status on the mean outcome of the compliers is $\beta = \delta/\pi$. We construct this using a two-stage least squares (2SLS) estimator for the second-stage model with GEM entry in 6th grade on the right-hand side:

$$Y_i = \beta \cdot GEM6_i + f(M_i) + e_i, \quad (3)$$

treating the indicator $1[M_i \geq 380]$ as an instrumental variable, and parameterizing $f(M_i)$ in the same way as $g(.)$ and $h(.)$. Our baseline models have no controls other than cohort fixed effects (4 dummies) to allow for trends over time in GEM participation and college outcomes, but we show that combinations of baseline characteristics that predict our outcomes trend smoothly at the 380-point threshold, and we also present models with controls.

For our main analysis, we focus on a bandwidth of 30 points, which is equivalent to $\pm 0.6\sigma's$ (standard deviation units) of the statewide math test. The first-stage and reduced-form relationships in our data are all relatively close to linear, so (as we show below) our estimates of δ and π are relatively stable across bandwidths. We also present estimates using alternative bandwidths, including the bandwidth selected by the algorithm of Calonico, Cattaneo and Titiunik (2014) (hereafter, CCT): in most cases the CCT selection is close to 30.

The entry rules for GEM specify that the student has a score of 380 or higher on the 5th-grade statewide math test and 300 or more on the 5th-grade reading test. The latter introduces a source of "fuzziness" to the first stage model (1), since not all students who pass the 380 point

math threshold have an eligible reading score, but it does not invalidate our basic approach. In fact, over 90% of students who barely satisfy the math test criterion have reading scores of 300 or above, so the requirement excludes relatively few students. It is slightly more constraining for boys than girls, and for FRL students than non-FRL students, leading to slightly smaller first stage effects of passing the math threshold for boys and FRL recipients.

b. Compliers

Our fuzzy RD design identifies the average treatment effect of entry to GEM for the subset of *compliers* whose GEM status changes at the math score cutoff. We estimate the characteristics of the compliers using the method of Abadie (2002), applied to an RD setting. Specifically, we estimate the mean of characteristic X of the compliers by the coefficient μ_x from a 2SLS regression of the form:

$$X_i \cdot GEM6_i = \mu_x \cdot GEM6_i + k(M_i) + \zeta_i, \quad (4)$$

using $1[M_i \geq 380]$ as an instrumental variable for $GEM6_i$, and including a piece-wise linear control function $k(M_i)$ similar to the ones included in equations (1)-(3). We use the same approach to estimate the mean potential outcomes $Y_i(1)$ and $Y_i(0)$ for compliers. In the former case we fit a 2SLS model with $Y_i \cdot GEM6_i$ as the dependent variable, and interpret the estimated coefficient of the endogenous regressor $GEM6_i$ as an estimate of the mean of $Y_i(1)$. In the latter case we use $Y_i \cdot (1 - GEM6_i)$ as the dependent variable and switch the sign of the 2SLS estimate of the coefficient on $GEM6_i$ to estimate the mean of $Y_i(0)$.

III. Data and Analysis Sample

a. Data Sources

We obtained anonymized student-level data from the District's centralized record system for all children who were in 5th grade between the 2003 and 2007 academic years. These records contain information on gender, age, race/ethnicity, and FRL/ELL status. They also include state-wide achievement test scores in reading and math for grades 3-8; course enrollments for grades 6-12, and course grades for middle and high school. We obtain information on college enrollment and degree completion from data reported to the District by the National Student Clearinghouse (NSC), which are available for all students who completed high school in the District. As discussed below, there is no evidence of a sample selection problem arising from discontinuities at the GEM cutoff in the fraction of students who stay through 12th grade.

b. Analysis Sample

Appendix Table A2 summarizes the derivation of our analysis sample. We start with 98,709 children who were in 5th grade in the District in the academic years 2003-2007. As shown in column 1 of the table, half these children are female, 43% participated in the FRL program, and their racial composition is 35% White non-Hispanic ("White"), 34% Black non-Hispanic ("Black"), and 24% Hispanic. We then eliminate the 20% of grade 5 students who were not also in grades 4 and 6 in the District (column 2); this has little effect on the characteristics of the sample. Next, in column 3, we eliminate the roughly 30% of students who left the District before 12th grade. This leads to some rise in the 4th grade test scores of the remaining group, and a slight increase in the share of girls (51% vs. 49%). About 18% participated in GEM and 29% participated in another advanced math track math class in 6th grade, leaving about 53% in the regular math track.

A potential concern for our sample is that students who enter GEM may be more or less likely to stay in the District and have college data, leading to differential selection biases on the two sides of the 380 point threshold for GEM entry. We investigate this issue in Appendix Figure A1, where we plot the probability of being included in the analysis sample against 5th-grade math scores. For the population as a whole (Panel A), there is a very small and statistically insignificant negative discontinuity (t=0.5) in the probability of sample inclusion around the GEM threshold. As shown in the other 4 panels of the figure, the estimated discontinuities are very similar for boys and girls, and for FRL versus non-FRL students.

For our "*RD analysis sample*," shown in column 4 of Table 1, we focus on students who were in the analysis sample described in column 3 and had a 5th-grade math score within 30 points of the GEM threshold of 380 points. This subset of 23,255 students has nearly even gender composition but its 31% FRL rate is below the District average of 43%. Relative to the overall District population, the analysis sample over-represents White students and under-represents Black students, reflecting racial differences in test scores. Not surprisingly, students with scores near the GEM cutoff had higher math (+0.51σ's) and reading scores (+0.41σ's) in 4th grade than the average student who persisted to 12th grade. About 29% ultimately entered GEM in 6th grade. Most of the others (49% of the RD analysis sample) were in the advanced track math class in 6th grade; only 22% were in the basic math stream.

The last four columns of the table show the mean characteristics of students in our RD analysis sample separately by gender and FRL status. Most characteristics are similar across gender and FRL lines, though 4th-grade math scores are slightly higher for boys and 4th-grade reading scores are higher for girls, while test scores in both domains are lower for FRL students.

IV. Empirical Results

a. Validity of RD Design and First Stage

We begin our main empirical analysis with Figure 2. The left panel of this figure probes the potential validity of our RD design by plotting the *predicted* probability of on-time high school graduation and completion of a bachelor's degree in STEM or business/economics within 6 years against 5th-grade math scores of students in our RD analysis sample. We use 1-point bins in this and all subsequent plots, reflecting the granularity of the underlying FCAT scores. In the RD analysis sample we have around 380 observations per bin, so the confidence interval for a bin mean extends ± 5 percentage points (ppt) when the mean is close to 0.5. To construct the dependent variable, we fit a simple linear prediction model using students in the RD sample whose 5th-grade math scores are below the GEM threshold. The explanatory variables are student demographic characteristics (gender, age, race/ethnicity, FRL status, English Language Learner status), gifted status, and 4th-grade math and reading scores on the statewide tests.

As discussed by Lee (2008), in a valid RD design there should be no discontinuity in the distribution of predetermined characteristics of sample members at the RD threshold: thus, we are looking for any evidence of a jump at 380 points in our predicted probability of majoring in STEM or business. The predicted major share rises steadily with 5th-grade scores but there is no indication of a jump at the GEM threshold. We also constructed similar figures by gender and FRL status. These are reported in Appendix Figure A2, and show no indication of discontinuous changes in student characteristics at the 380-point threshold.

We also looked for evidence of any discontinuity in the density of test scores around the threshold, as suggested by McCrary (2006). Plots of the density, shown in Appendix Figure A3, show no sign of bunching or other manipulation of the scores, which is expected given that the test is administered statewide and graded electronically.

The right panel of Figure 2 shows the first-stage relationship between 5th-grade math scores and GEM entry. For reference, we also plot the relationship between test scores and the probability that a student either enters GEM ***or*** takes advanced math in 6th grade. Focusing on the

first stage for GEM, we note three things. First, the relationships to the right and left of the 380 point cutoff are nearly linear, apart from a small upward trend in GEM participation for students with scores up to 6 points below the threshold.[13] Second, among the students with scores near the cutoff, there are both "GEM always takers" (as evidenced by the ~10% GEM entry rate for students with scores just below the 380 cutoff) and "GEM never takers" (as evidenced by the roughly 40% of students with scores just above the 380 cutoff who do not enter GEM). Third, there is a large jump in the probability of GEM entry – around 50 ppt in magnitude – implying that our RD design has a powerful first stage ($t \approx 18.3$ for the pooled sample).

Comparing the first stage for GEM entry to the RD model for entering *either* advanced math or GEM shows that nearly all the students who move into GEM when their score crosses the 380 point threshold come from the advanced math track. Only about 3 ppt of the total jump in the GEM enrollment share comes from students who would have been in the regular track.

Appendix Figure A4 presents a parallel set of first-stage plots by gender and FRL status. All four of these figures look very similar to Figure 2B, with relatively low rates of GEM participation to the left of the 380 point threshold and large jumps in the probability of entering GEM at the cutoff, of roughly 50-55 ppt in magnitude. We also see that for all four subgroups, there is a much smaller rise (<5 ppt) in the probability of GEM or advanced track math in 6th grade, implying that most of the students who enter GEM would otherwise have enrolled in the advanced track.

b. Characteristics of Compliers

Table 1 shows the characteristics of the compliers for the GEM entry cutoff, estimated using the approach of equation (4). Panel A shows baseline (pre-GEM) characteristics, while Panels B and C present estimated potential outcomes without the GEM treatment (i.e., the means of $Y_i(0)$), in middle and high school and in college, respectively. These mean potential outcomes – which we refer to the "untreated complier means" – are analogous to the means for the *control group* in a simple 2-arm RCT: they represent the counterfactual outcomes for the compliers who are admitted to GEM and have realized outcomes $Y_i(1)$ – which we refer to as "treated complier means." For reference, the first column of the table shows the means of the same variables for

[13] This appears to be driven by a practice of filling the 6th-grade GEM classroom with nearly-eligible students at schools where relatively few students meet the 380 threshold.

all the students in our RD analysis sample (i.e., all those who graduated on time from the District and had a 5th-grade math score between 350 and 410 points).

Looking first at the baseline characteristics in Panel A we see that on average, the compliers are broadly similar to students in the overall analysis sample. Looking within the complier population, however, there are some interesting differences by gender: female compliers appear to be somewhat less advantaged, with a higher shares of Black and FRL students (but a lower share of Hispanics). The girls also have nearly 0.1 $\sigma's$ higher 4th-grade reading scores than the boys, while the boys are 0.05 $\sigma's$ ahead in their 4th-grade math scores. Likewise, comparisons between FRL and non-FRL compliers show relatively big differences in their racial composition, reading scores, and school peers.

Turning to Panel B, we first show the 6th-grade math course enrollments of the untreated compliers. As expected given our discussion of Figure 2B, 94% are in the advanced track math and only 6% in the basic math track, with rates that do not vary too much by subgroup (though interestingly boys are more likely to be in the basic math track than girls despite their higher average 4th grade math scores). About 60% of untreated compliers complete algebra by the end of middle school, but only 2% also complete geometry, reflecting the lack of a pathway to geometry in the absence of GEM. Untreated compliers have fairly strong math GPA's (close to 3.4 in 6th grade, falling to 3.0 by 8th grade). As is often the case, girls have higher course grades than boys, even though by construction all the compliers have similar 5th-grade math scores.

About three quarters of the untreated compliers took an advanced math class in high school, and 23% took calculus class in high school, with slightly higher rates for girls and non-FRL students. Among the calculus takers, a higher fraction of girls achieved a grade of B- or better (0.18/0.24=75%) than males (0.14/0.21=66%). A large share of the untreated compliers who took calculus (roughly 60%) took it through an AP or International Baccalaureate (IB) class, with fairly similar shares across subgroups.

Nearly all high school students in the District take the PSAT, and about 90% of those in our RD analysis sample also took the SAT or ACT. The untreated compliers have an average PSAT math score (in national percentile units) of 69.5, with a slightly higher mean for the boys

than the girls (+1.5 percentiles).[14] Average percentile scores in the PSAT reading test are a little lower and are 3.7 percentiles higher for girls.

Finally, turning to the post-secondary outcomes in Panel C, we see that 97% of the untreated compliers graduated high school on time, with similar rates by gender but a lower rate for FRL students. (Of course, our analysis sample requires that students were still *enrolled* in high school 6 years after entering 6th grade, so these high rates are partly mechanical.) 76% graduate on time and enter college the next year, with a 5 ppt higher rate for girls than boys and an even wider gap (9 ppt) between FRL and non-FRL students. Just over 60% graduate on time and enter *selective* colleges, with an 8 ppt higher rate for girls than boys, and a 15 ppt gap by FRL status.

Our main outcomes of interest are degree completion and field of study. Among the untreated compliers, rates of degree completion and majoring in different fields closely parallel the overall means for our RD analysis sample: 55% complete a college degree within 6 years, 43% compete a bachelor's degree, and 19% complete a bachelor's degree in a STEM or business-related field. The degree completion rates are substantially higher for untreated complier females than males, and for non-FRL than FRL participants.[15] Despite their +14 ppt advantage in completion of a bachelor's degree, however, girls are less likely to complete a degree in STEM or business-related fields. Indeed, only 30% (=0.15/0.50) of the untreated female compliers who complete a bachelor's degree within 6 years graduate in a STEM or business field, versus 67% (=0.24/0.36) of the untreated male compliers.

c. Impacts of GEM Entry on Math Courses in Middle and High School

With this background, we turn to an analysis of the impacts of GEM entry on subsequent student outcomes, using the framework of equation (3). Figure 3 shows the reduced-form relationships between two main pre-college outcomes: whether the student completed algebra before high school, and whether she or he took calculus in high school. In panel A we see very

[14] Note that boys have higher math PSAT's than girls, even though they have lower grades in their math classes (and are somewhat less likely to take advanced math classes in high school). This pattern is also present in other samples, such as the sample of gifted students from the District studied in Card, Chyn and Giuliano (2024), and has interesting implications for college admission policies that down-weight or ignore standardized test scores.

[15] The gender gap in college completion is a national phenomenon. The 6-year degree completion rate for men who started at 4-year institutions in 2015 was 57.5%, versus 64% for women (Digest of Education Statistics, 2022, Table 326.1). The 3-year degree completion rate for men who started at 2-year institutions in 2015 was 29.6% versus 34.7% for women (Digest of Education Statistics, 2022, Table 326.2).

clear jumps in the probability of taking algebra in middle school, and taking and passing the class, of around 10-12 percentage points. In panel B we see smaller jumps in the probability of taking (or taking and passing) calculus, of around 5 percentage points. Figure 4 shows the same outcomes, focusing on successful completion of each course and cutting the sample by gender and FRL status. For both outcomes, the graphs show similar jumps for girls and boys and for FRL and non-FRL students.

Table 2 presents a series of estimated RD models for our pooled sample and by gender and FRL status. Each column shows estimates for a different outcome, while the rows pertain to different subgroups. For each sample or subsample we show estimates from models that include only four 5th-grade cohort dummies, and from models that include the full set of control variables used in the construction of our predicted probability of earning a bachelor's degree in STEM or business/economics (graphed in Figure 2A).[16] We cluster standard errors by 6th-grade middle school to allow for potential correlations in outcomes arising from school- or neighborhood-specific factors.

Column 1 presents the discontinuities in the *predicted* probability of a bachelor's degree in STEM or business/economics (the variable graphed in Figure 2A and Appendix Figure A2). These are all very small in magnitude, as expected given the patterns in the figures. Column 2 presents the estimated first-stage effects of passing the 380 point threshold on entering GEM in 6th grade (i.e., estimates of the parameter π in equation 1). Consistent with the plots of the first stage relationships, these range from 0.50 to 0.53 in magnitude, and are highly significant. Even for our smallest subsample of FRL participants (N=6,990) we have a t-statistic for the first stage of about 15.5 (or an F-statistic of 241), well above the threshold proposed by Lee et al. (2022) that allows an analyst to use conventional confidence intervals for 2SLS models with no adjustment for the imprecision of the first stage.

Columns 3-13 present 2SLS estimates of the effect of entry to GEM on course participation and course grades. As shown in the first row of column 3, the overall effect on the probability of taking early algebra is 20 ppt. Given the 61% probability of taking algebra early for the untreated compliers, this implies that 81% of the treated compliers enrolled in algebra by

[16] One of the controls is gifted status. A small share of gifted students is identified in 5th grade and one might be concerned that this is partly due to their performance on the math test. As a check, in Appendix Table A3, we present models in which we include an interaction between gifted status and the dummy for having a math score above 380. These models give results that are quite close to those without the extra control.

8th grade. The other 19% entered GEM in 6th grade but failed to earn an A or B grade and were moved back to the advanced track course in 7th grade, and also missed the GPA threshold for the 'second chance' to enter Algebra I in 8th grade (see Figure 1). Comparing results by gender we see that the treatment effect of GEM entry on early algebra is a little higher for boys than girls (+21 ppt vs +19 ppt). Taken together with the means of the untreated compliers, these impacts imply that only 13% of girls who entered GEM failed to enroll in algebra by 8th grade, compared to 23% of boys.

The results in column 4 show that the overall effect of GEM entry on enrolling in geometry in 8th grade is 43 ppt, with a significantly larger effect for girls (52 ppt) than boys (34 ppt). Combing these treatment effects with the means in Table 1 and the effects of GEM on algebra enrollment, it is straightforward to calculate the mean middle school enrollment and performance outcomes for the untreated and treated compliers. Figure 5a summarizes the distributions of an ordered version of these outcomes for the compliers of both genders. Given the rules for GEM we order middle school math attainment as: (1) never took algebra; (2) took algebra but did not pass; (3) passed algebra but did not take geometry; (4) took geometry but scored C or worse; (5) took geometry and scored B or better. The distributions are clearly shifted by entry to GEM, but even among the treated compliers only about one-half of girls enrolled in geometry and 40% earned a grade of B or better in the course, while for the boys only a third enrolled in geometry and a quarter earned a grade of B or better.

Column 5 presents estimates of the effect of GEM entry on the probability of taking calculus in high school. For the overall sample we see a significant effect of +9 ppt (t=3.6). Comparing impacts by subgroups, the effects are similar for girls and boys, though the regression-adjusted effect for boys is somewhat smaller (+7.4 ppt, t=2.5). The gains in calculus enrollment are similar for FRL participants (+9.7 ppt) and non-participants (+8.5 ppt).

Columns 6-9 present 2SLS estimates in which the outcome variable is the event of taking algebra in middle school (columns 6-7) or calculus in high school (columns 8-9) ***and*** achieving a certain grade. In interpreting these models it is important to recognize that entry to GEM can increase the share of students taking a given course (the extensive margin) and also affect the grades of students who would take the course regardless of GEM (an intensive margin effect on the "algebra always takers").[17] Thus the effect of GEM entry on the share of students achieving a

[17] As noted, the untreated GEM compliers have about a 60% enrollment rate in algebra.

certain grade (or better) in a given course can actually exceed the effect on the fraction taking the course. Indeed, the effect of GEM entry on the overall share of students taking algebra in middle school and earning a C- or better is +21.5 ppt, slightly larger than the 19.8 ppt effect on the share of students taking algebra. The 2 ppt gap implies that entry to GEM must have had some effect on grades for the students with scores near the GEM cutoff who would have taken algebra by 8th grade even if they narrowly missed the GEM threshold. Comparing the estimates in column 7 and column 3 we conclude that entry to GEM had intensive margin effects on algebra performance for all four subgroups.

The estimated effects of GEM on taking and passing calculus (column 9) are typically just a little smaller than the effects on taking calculus, though again some of the effect on grade outcomes could represent gains for the calculus always-takers. The estimated effects on earning a grade of B- or better (column 6) are about one-half as large.

Figure 5b summarizes the relative impacts on high school calculus course taking and grade outcomes. As in Figure 5a, we show the distributions of an ordered set of outcomes for the untreated and treated compliers: (1) never took calculus; (2) took but did not pass calculus; (3) earned a grade of B- or better in calculus; and (4) earned a grade of A- or better in calculus. While GEM increased calculus-taking and the share of compliers who pass or even earn a grade of B- or better in the course, it has little effect on the rate of very high achievement.

Finally, columns 10-13 show effects of GEM entry on math GPA's in 6th through 8th grades, and in high school. For the overall sample and all subgroups there are large negative impacts of entering GEM on 6th-grade math GPA's, on the order of −0.7 of a letter grade. There are also negative, but smaller, effects on 7th and 8th-grade math GPA's. These impacts presumably arise because teachers award similar distributions of grades regardless of class difficulty: thus an untreated complier will tend to get a higher grade in advanced track math (where she is close to the top of the distribution of math ability) than a treated complier in the GEM class (where she is closer to the bottom). Importantly, by high school these negative impacts have largely disappeared. Even for boys, the effect is small in magnitude and far from statistical significance (t=0.5).

We have also investigated the effects of GEM entry in 6th grade on a variety of other high school-related outcomes, including enrollment in *any* advanced high school math class (pre-calculus, calculus, or statistics); enrollment in AP or IB calculus; completion of AP or IB

calculus with a score of 3 or better (which is required for college credit); overall high school GPA; PSAT math and verbal scores; and whether the student took the SAT or ACT. The estimates for these outcomes are reported in Appendix Table 4.

In brief, we find effects on any advanced math course that are similar in size to the effects on calculus, with a positive boost for the overall sample and all 4 subgroups. For AP/IB calculus enrollment we find a positive effect of around 5 ppt for the overall sample (+5.1 ppt, t=2.7), with very similar effects by gender and FRL status. For completion of the AP Calculus exam with a score of 3 or better, we find marginally significant positive effects of around 3 ppt.[18] Looking at overall high school GPA's we find a relatively small effect in the overall sample (+0.002, t=0.1) with small but offsetting effects for girls (+0.044, t=1.6) and boys (−0.047, t=1.1).

Interestingly, we also find a significant and relatively large positive effect of GEM entry on PSAT math scores (+2.6 percentiles, t=2.7) with very similar effects for girls and boys but a smaller (and insignificant) estimated effect for FRL students. We find no effects of GEM on the probability that students take either the ACT or SAT.

d. Impacts of GEM Entry on College Outcomes

Next we turn to college-level outcomes. Panels A and B of Figures 6, 7, and 8 show the reduced-form relationships between 5th-grade reading scores and 3 main outcomes: the probability of finishing high school on time and entering college within 1 year (Figure 6); the probability of earning a bachelor's degree in any field within 6 years after the "normal" high school completion year (Figure 7); and the probability of earning a bachelor's degree in STEM or business-related fields within 6 years (Figure 8). In each case, panel A shows the reduced forms by gender, while panel B shows the reduced forms by FRL status. We discuss panels C and D below.

Looking first at on-time college entry in Figure 6, we see only small jumps at the 380 point threshold by gender or FRL status. Turning to Figure 7, we see some indication of a positive jump in the probability of obtaining a bachelor's degree within 6 years for girls, but no such effect for boys. The reduced-form plots by FRL status also show a small gain at the GEM threshold for FRL students. The reduced-form plots in Figure 8 for the probability of obtaining a

[18] This variable measures the score on the AB portion of the exam for students who took the BC exam.

STEM or business degree show a clear jump for girls, but no effect for boys, while the jumps for FRL and non-FRL participants are similar but modest in size.

Table 3 summarizes our estimates of the effect of GEM entry on college-level outcomes, using the same format as Table 2. The first stages for all the underlying models are the same as the ones reported in column 2 of Table 2. Column 1 presents estimates of the effect of GEM entry on the probability of graduating high school 7 years after 5th grade (i.e., on-time). None of the estimated effects are large or statistically significant. Column 2 presents models for on-time college entry (the same outcome shown in Figure 6). The overall effect (in row 1) is small but positive (+2 ppt, t=1.0), This average reflects a somewhat larger, marginally significant positive effect for girls (+3.5 ppt, t=1.6) and a zero effect for boys. The estimate for non-FRL students is relatively small (+1.5 ppt) while that for FRL students is a little larger in size but insignificant (+3.1 ppt, t=0.9).

It is useful to check whether the estimates on our main college-related outcomes are sensitive to choice of bandwidth. (Recall that we are using a symmetric bandwidth of 30 points in our main tables.) We investigate this issue in the lower panels of Figures 6, 7, and 8, where we show the 2SLS estimates by gender (panel C) and FRL status (panel D) as we vary the bandwidth from 10 points to 40 points, along with their 95% confidence intervals. For reference we also highlight using a solid-colored symbol the bandwidth choice selected by the CCT algorithm (using a symmetric uniform kernel and CCT's "regularization" procedure). For example in Figure 6C the CCT bandwidth choice is 30 for boys and 34 for girls; in Figure 6D the CCT choices are 26 and 38 for non-FRL and FRL students.

Looking at Figure 6, the pattern of estimates in panel C show that the estimated effects of GEM on on-time high school graduation for girls are about 3 or 4 ppts, regardless of bandwidth, whole the estimates for boys exhibit a bit more sensitivity, ranging from around -10 ppts at shorter bandwidths to zero or very small positive estimates at bandwidths of 30 or more. For non-FRL participants the estimated effects of GEM are quite similar and close to 0 for all bandwidths, while the estimates for FRL participants are a little more sensitive and tend to become more positive at higher bandwidths, but never statistically different from 0.

Returning to Table 3, column 3 shows estimates for models of the probability of enrolling at a selective college or university with a year of normal high school graduation time.[19] Here the overall effect is slightly positive, reflecting a relatively large positive effect for girls (+6.6 ppt, t=2.1) and a very small negative effect for boys (−0.9 ppt, t=0.3). The estimated effects of GEM on selective college enrollment for non-FRL participants is a little bigger than the overall average effect with that for FRL participants is close to zero, but both are statistically insignificant.

The estimates in columns 4 and 5 pertain to degree completion: column 4 shows effects of GEM on obtaining any degree in the 6 years after normal high school graduation (i.e., an associate's degree or bachelor's degree), while column 5 shows effects on bachelor's degrees only. Interestingly, the estimates in the two columns are quite similar. For the overall sample we estimate small positive effects on any type of degree (+3.1 ppts, t=1.6) and on bachelor's degrees (+2.9 ppt, t=1.4). These gains are attributable to girls, for whom we estimate a marginally significant 5.5 ppt effect on obtaining any degree (t=1.7) and a marginally significant 7.0 ppt effect on bachelor's degrees (t=2.0). For boys the estimated effects on both outcomes are very small, as are the estimated effects on non-FRL students. The effects for FRL participants are larger (comparable in size to the effects for girls) but relatively imprecise.

Figures 7C and 7D show the sensitivity of the estimated effects on bachelor's degree completion to bandwidth choice. For girls and non-FRL participants the estimates are very stable across bandwidths. There is more variability in the estimates for boys, which tend to be more negative at smaller bandwidths but settling close to 0 for bandwidths above 25. The estimates for FRL students are also a little more sensitive to bandwidth but never close to conventional significance. Notably, the estimates corresponding to the CCT bandwidth choices for boys and FRL students (29 and 26 respectively) are both very close to zero.

Columns 6-8 of Table 3 present estimated effects of entering GEM in 6th grade on the probability of obtaining a bachelor's degree in a STEM field (column 6), a degree in business or economics (column 7) or either of these two sets of fields (column 8). For the sample as a whole the estimates show small positive effects on both STEM and business related degrees (+1.9 ppt and +1.7 ppt) and a marginally significant positive effect on either STEM or business related

[19] The selective designation is based on the school's Carnegie Classification in the year the student enrolled. We note that 79% of the untreated compliers who enter college on time enter a selective college.

degrees (+3.7 ppt, t=1.8). All of these effects are concentrated among girls. The estimates for girls show a relatively large boost in STEM degree completion (+4.6 ppt, t=2.2) and in business or economics (+3.7 ppt, t=1.9), and a combined boost of 8.3 ppt (t=2.4) in STEM or business-related fields. Note that this 8.3 ppt effect is just a little bigger than the overall 7.0 ppt effect on obtaining any bachelor's degree, suggesting a relatively small "crowd out" effect on other fields. While not shown in the table, there is a small negative estimated effect on obtaining bachelor's degrees in heath related fields (-2.2 ppts, t=1.1). In contrast to these impacts for girls, we estimate only small effects, all far below conventional significance levels, on the fields of degree of boys. The effects for FRL and non-FRL participants are broadly similar to the effects for the overall sample, although there is greater "crowd out" of other majors among non-FRL students, while FRL students see an larger increase in their rate of bachelor's degree completion.

Figures 8C and 8D show the robustness of the estimated effects on degree completion in STEM or business related fields to choice of bandwidth. For girls, the estimated effects are relatively stable across bandwidths, pointing to an effect of around 8 ppt, while for boys the estimated effects are slightly more negative at low bandwidths but settle down very close to 0 for bandwidths above 25 points. The estimated effects on FRL and non-FRL students are also quite robust, and consistent with an impact of 4-5 ppt, roughly the average of the effects for girls and boys.

The impacts of GEM on major choice are summarized in Figure 9, which shows the mean shares of the treated and untreated compliers who obtain bachelor's degrees in STEM, business/economics, and all other fields. Looking first at the left panel, we see that among the untreated compliers, girls earn only about one-half as many degrees in STEM and business/economics as boys. Exposure to the GEM program closes that gap, while also expanding the share of girls who obtain any degree. By comparison, the field distributions of the treated and untreated complier boys are quite similar, as is the share with any degree. Thus, exposure to GEM allows girls to catch up to boys in the share with a bachelor's degree in STEM or business fields, while widening the gender gap in the probability of obtaining any bachelor's degree from 14 ppt among the untreated compliers to 22.5 ppt among the treated compliers.

The contrast between the distributions of degree outcomes by FRL status is not nearly so dramatic as the contrast by gender. For both FRL and non-FRL students, exposure to GEM has small positive effects on obtaining STEM or business-related degrees, and small positive effects

on obtaining any degree. However, the effect on bachelor's degree completion is somewhat larger for FRL students, implying that GEM exposure narrows the gap in degree completion between non-FRL and FRL compliers, from 17 ppts in the absence of GEM to 13.2 ppts with GEM.

e. Heterogeneity in the Effects of GEM by Predicted Probability of Success in GEM

So far we have contrasted the effects of GEM by gender and by FRL status, and found relatively large differences by gender. Motivated by the observation that less than one-half of the compliers who enter GEM successfully complete all three years of the program, in this section we develop a simple model to predict success in GEM and then compare results between groups classified by gender and high or low probability of success in GEM (defined as having a predicted probability above or below the gender-specific median). To preview the results, we find that within each gender, differences in the impact of GEM on college-level outcomes by predicted probability of success in GEM are relatively small. We conclude that the "essential heterogeneity" in the impacts of GEM (Heckman et al., 2006) is between genders, rather than in the match between the GEM curriculum and a student's abilities.

We predict success in GEM using a 2-equation probit model that accounts for selectivity into the program. The first equation describes the probability of entering GEM as a function of 5th grade math scores and all the control variables used to construct Figure 2A. We adopt the same functional form as our first stage model, allowing a change in the slope of the relationship between program entry and math scores at the GEM threshold, and including a dummy for having a score at or above the threshold. The second equation, estimated on the subset of students who actually enter GEM, describes the probability that a student successfully completes all three years of GEM (i.e., passes geometry by 8th grade). This model includes all the control variables, and the piece-wise linear function of 5th grade math scores, but excludes the dummy for having a score at or above the 380 point cutoff.[20] We fit the model separately by gender using our main RD analysis sample.

Estimation results are presented in Appendix Table A5. Most observed characteristics that are positively correlated with entry to GEM conditional on grade 5 math scores (e.g., ELL

[20] The identifying assumption is that the effect of 5th grade math scores on the probability of GEM completion is smooth at the cutoff – the same as the assumption in our RD model.

status or Asian ethnicity) are positively correlated with success in GEM. We also find that the error terms in the two equations are positively correlated ($\hat{\rho} = 0.22$ for girls, $\hat{\rho} = 0.47$ for boys); i.e., that the unobserved determinants of GEM entry and GEM success are positively correlated. These patterns suggest that the GEM selection process targets children who are more likely to complete the program. Even with this targeting, however, only a minority of marginally qualified entrants actually do so.

With the estimates from this model in hand, we then form the predicted probability of GEM success as a function of the observed covariates, and classify students of each gender into "high" (above median) and "low" (below median) probability of completing GEM. The upper panels of Appendix Figure A5 presents plots of the first-stage relationship between 5th-grade math scores and the probability of entering GEM in 6th grade by gender and predicted success group. All four groups show large jumps in GEM entry at the 380 point threshold, but the jumps are larger for high probability students, consistent with a targeting hypothesis. The lower panels of the figure show the reduced-form relationships between math scores and the probability of graduating with a degree in STEM or business-related fields within 6 years of the date of normal high school completion. For girls, both groups show jumps at 380 points, with a somewhat larger jump for the high-predicted-success girls, whereas for boys there is no indication of any shift at the GEM threshold.

Table 4 summarizes the main findings from our comparisons of the effects of GEM on high and low probability students, using the same format as Tables 2 and 3. The first column presents estimates of the discontinuity in the predicted probability of completing all 3 years of GEM: these are all very small and far from statistical significance. Column 2 presents the first stage effects of having a 5th grade math score at or above the GEM threshold on the probability of entry. Consistent with the patterns in Appendix Figure X, the jumps are about 13-14 ppt. larger for students with higher probability of successful completion of GEM. Even for low-probability students, however, there is still a 44-46 ppt jump at the test threshold.

Columns 3-5 present the estimated impacts of GEM entry on successful completion of algebra and geometry in middle school, and on taking and passing calculus in high school. The estimated impacts on algebra are a little smaller for high-probability girls; the estimates for the other three groups are similar and between 22 and 25 ppts. The effects on geometry are more

variable, in part because we have (essentially) classified groups based on this outcome.[21] High probability girls have about a 57 ppt impact. The effect for low-probability girls is about 41 ppt, slightly above the effect for high-probability boys (37 ppt). Low probability boys have only a 25% impact. Despite these differences in the effects on completing geometry, however, the estimated effects on passing calculus are similar for both groups of girls and for high-probability boys. The estimated effect on low-probability boys is smaller, though the 95% confidence interval extends from -0.05 to 0.10 so we cannot reject that all four groups have similar effects on calculus.

Columns 6-9 present the estimated effects of GEM entry on the probability that students complete high school on time and enter college the next year; that they graduate on time and enter a selective college, and that they complete a bachelor's degree within 6 years of on-time graduation. There is more heterogeneity within genders in these effects. The impacts on college entry and selective college entry for girls noted in Table 3 appear to be driven by low-probability girls, though interestingly the effects on 6-year degree completion are similar for low- and high-probability girls. GEM's impacts on on-time college entry are slightly positive for high-probability boys and slightly negative for low-probability boys, though the relative size of the impacts on degree completion are reversed.

Finally, columns 9-11 show the effects of GEM entry on the probability of majoring in STEM, business-related fields, or either. Here, the effects for high- and low-probability girls are positive and very similar in magnitude. For boys, on the other hand, the effects are all close to zero.

Our interpretation of the results in Table 4 is that the impacts of GEM are broadly similar within gender groups, particularly in degree completion and major choice. For girls, admission to GEM leads to a roughly 7 ppt effect on the probability of obtaining a bachelor's degree within 6 years of high school graduation, and positive effects on majoring in STEM and business-related fields that sum to about +8 ppt. For boys, there are no large effects on degree completion or major choices. Thus, differences in the probability of success in completing all three years of GEM do not matter very much in determining the size of the program's effects on college-related

[21]Recall that geometry participation is close to 0 for non-GEM students. Thus, the percent of GEM entrants who pass geometry is very close to the effect of GEM entry on passing geometry.

outcomes. Instead, it appears that the main source of heterogeneity is student gender. We return to a discussion of the possible sources of this heterogeneity in Section VI, below.

V. Spillover Effects on Untreated Compliers?

Our RD analysis relies on comparisons between students whose 5th-grade math scores are just above or just below the 380 point GEM threshold. A concern with this design is that the presence of GEM could have some effect on students who missed the threshold (i.e., the untreated compliers). The existing literature suggests three potential channels for such effects: (1) a negative peer effect, arising from the shift of the higher-scoring students to the GEM classes (Paloyo, 2020);[22] (2) a discouragement effect, arising from the stigma of being left out of the advanced math stream (Ellison and Swanson, 2011); (3) a "top of the class" effect (Denning et al., 2023) arising because the untreated compliers move from being around the 80th percentile of their classmates in math to being the highest scoring students in their classes. Peer and discouragement effects are a particular concern because they would lead to a positive bias in our RD estimate of the effect of GEM. Top of the class effects would work in the opposite direction.

The fact that the estimated effects of GEM on the longer run outcomes of boys are uniformly small could be interpreted as evidence against the presence of large peer effects, since boys who narrowly miss the GEM threshold experience the same changes in peer quality as girls with the same scores.[23] Similarly, one could argue that boys who miss the threshold experience the same potential discouragement effects as girls. To the extent that middle school girls are particularly sensitive to discouragement, however, this may be unpersuasive.[24]

To address these concerns, we use data from another large urban school district where we can follow students from 4th grade to the end of college. This district – which we refer to as "District 2" – has two main tracks in middle school math, with most of the students who score in the upper half of the statewide reading test in 5th grade participating in the upper track. It has no

[22] We note that many recent studies find limited or no evidence of peer effects. For example, Card and Giuliano (2016) find no effects of tracking 4th and 5th grade students into gifted/higher achiever (GHA) classrooms and regular classrooms on the students who narrowly miss the threshold for the GHA class. Likewise Abdulkadiroglu et al. (2014) find no evidence of peer effects on students who attend elite exam-based high schools.

[23] One might be concerned that the gender composition of the GEM classes is a lot different than that of the other math tracks. However, the differences are very small: 46% of 6th grade GEM students are female, versus 52% in advanced track 6th grade math, and 49% in the regular track class.

[24] Ellison and Swanson (2021) find evidence of larger discouragement effects for girls than boys from competing in math competitions and losing, and discuss earlier studies which often report this pattern. They interpret their findings in the context of a broader literature suggesting that females prefer to avoid competition.

GEM program or third track for very high scoring students. Nevertheless, it has a policy of pushing nearly all higher-scoring students into early algebra, whereas the District has a far more selective policy. The difference is illustrated in the two left-hand panels of Figure 10, where we show the fractions of girls (upper panel) and boys (lower panel) who enroll in Algebra in middle school in the District and in District 2, conditional on 5th grade math scores. To help interpret the graphs, we fit local linear models for the two districts using data ± 30 points from the GEM cutoff (allowing intercept and slope changes at the GEM cutoff). For students with scores close to the GEM cutoff, over 85**%** of students in District 2 enroll in Algebra in middle school – versus around 60% of students in the District with scores just below the cutoff, and 70% of students in the District with scores just above the cutoff.

As shown in the two middle panels, however, District 2 does not have a similar policy for early geometry. Among students with 5th grade math scores in the 380-400 range (i.e., 0-20 points above the GEM cutoff) less than 5% of girls or boys in District 2 enroll in geometry by 8th grade. In contrast, in the District this fraction is over 40% for girls, and around 30% for boys.

Given the differences in the middle school math sequence in District 2 – where the vast majority of students with 5th grade math scores on either side of the GEM threshold are taking *the same* advanced track courses – and in the District, where these students are separated into GEM and advanced tracks based on having a math score above or below 380 points, it is interesting to ask how the longer run outcomes of students with different 5th grade scores compare in the two districts. In particular, it is interesting to ask whether there is any evidence that students who have 5th grade math scores just below the GEM cutoff do worse in the District than in the District 2 – as might happen if there are significant negative peer effects or discouragement effects caused by moving the higher-scoring students to the GEM track. It is also useful to confirm that girls in the District who score just above the 380 point threshold show gains relative to girls in District 2 with similar scores. Finally, a cross-district comparison allows us to look at differences for students with scores well above the GEM threshold.

The right-hand panels of Figure 10 show the fractions of girls and boys with different 5th grade math scores who take calculus in high school in the two districts. For scores below 380 the fractions are quite similar in the two districts; for scores the GEM threshold, however, the fraction in the District is higher for both genders, consistent with the patterns in Figure 4C.

Interestingly, the gap diverges at higher scores, suggestive of larger effects of the GEM program on students who are further above the cutoff.[25]

A concern with these simple comparisons is that the distributions of student characteristics are different in the two districts, even conditional on 5th grade math scores. To assess this concern this we fit a model for the probability of graduating high school on time and completing a bachelor's degree in STEM or business/economics within 6 years, using data on students in the District with 5th-grade math scores below the GEM threshold. The predictors include a combination of 4th grade test scores and student characteristics. We then assigned the predicted probabilities from this model to all students in both districts and plotted the means by 5th grade math score. The results, shown in Appendix Figure A6, show that the mean fitted probabilities are very similar at each 5th grade math score. This suggests that the average characteristics in the two districts are very similar conditional on 5th grade math scores -- particularly the characteristics that affect degree completion and college major choice.

Figure 11 extends the cross-district comparisons to two key college-level outcomes: the probability of obtaining a bachelor's degree within 6 years of high school graduation (left hand panels); and the probability of obtaining a degree and majoring in STEM or business-related fields (right panels). For girls, we see that rates of bachelor's degree attainment among students with scores below 380 are a little higher in the District at every score, with a gap that is constant across scores, whereas the probabilities of completing STEM or business-related degrees are quite similar.[26] For scores above 380, however, we see a jump for girls in the District relative to girls in District 2, consistent with our reduced form plots (Figures 6 and 7). For boys, we see a similar pattern of means for the two districts at scores below 380, but no evidence of a jump at 380 – again consistent with our reduced form plots.

Most importantly, we see no indication that students in the District who score just below the GEM threshold have worse (or better) college-related outcomes then we might expect, given the outcomes for students with the same scores in District 2 and the differences between the

[25] In Appendix Figure A5, we present a comparison of other measures of mathematics achievement in high-school, including enrollment in AP/IB calculus and PSAT scores. The patterns are similar to those in the right-hand panels of Figure 10.

[26] It may be surprising that students in the District with scores below the GEM threshold have the same probability of majoring in STEM or business as those in District 2, given that a much higher share of this group in District 2 enrolls in algebra in 8th grade. We interpret this, and similar patterns in calculus-taking and other high school math outcomes, as evidence that pushing a large share of students into early algebra has very limited effects.

districts at scores far below the GEM threshold. In fact the fitted lines for students in the District with scores below the GEM threshold are essentially parallel to the fitted lines for students in District 2. We interpret this as evidence that there are no negative (or positive) spillover effects from the GEM program on students who narrowly miss qualifying for the program.

VI. Interpreting the Gender Differences in the Effects of GEM

<< preliminary/incomplete >>

Our findings suggest that entry to GEM has large effects on the probabilities of completing algebra and geometry in middle school for students of both genders – though the geometry gains are bigger for girls. It also has significant effects on the probabilities of taking calculus in high school and on math PSAT scores, of about the same magnitude for girls and boys. At the college level, however, there appear to be no effects on boys but relatively large effects on girls. Why do the impacts of GEM on middle and high school carry through to longer-run outcomes for girls but not for boys?

To begin it is useful to review several potential explanations that we believe do not hold up. First, we do not think the differences are related to the gender composition of middle school math classes. As noted, the share of girls is very similar in GEM classes, advanced track math classes and regular track classes. Second, we do not think they are an artefact of biases driven by negative effects of GEM on the untreated female compliers. Appendix Figure A5 and Figure 11 show no evidence that girls (or boys) with scores just below the GEM cutoff do worse than expected given patterns in District 2. Third, we do not think the gender differences are driven by differences in the "match" between student ability and the GEM curriculum, even though girls on average are more likely to complete GEM. The results in Table 4 show that even boys with higher predicted probabilities of completing GEM experience no gains in degree completion or STEM/business majoring, whereas girls with lower predicted probabilities gain about as much in these domains as girls with higher predicted probabilities.

One possible explanation is the gender gap in *non-cognitive* skills. Researchers including Jacob (2002) and Becker et al. (2010) have argued that the overall gender gap in college entry and degree completion is mainly due to a lack of non-cognitive skills among boys. A related body of work (e.g., Heckman and Rubinstein, 2001; Heckman et al. 2006) argues that cognitive and non-cognitive skills are complements in the education production function, and that gains in the cognitive domain have little impact for students who lack non-cognitive skills to exploit these

gains. To the extent that entry to GEM expands math skills, with little or no effect on non-cognitive skills, the main beneficiaries of GEM will be students whose math-related skills are low relative to their non-cognitive skills. Arguably this is more likely to include girls than boys.

VII. Summary and Conclusions

There is much disagreement about how to structure the math curriculum in American middle and high schools. On one hand, the science and engineering programs at many selective schools ask for evidence of mastery of advanced math in high school, especially calculus. This leads the parents of higher-achieving students to demand access to algebra and geometry in middle school so their children have a pathway to calculus by the end of high school. On the other hand, previous efforts to push "algebra for all" in middle school are perceived as having failed, especially for weaker students. A possible resolution to this conundrum is the availability of targeted acceleration programs like the GEM program we study in this paper.

GEM is a three-year program that compresses the standard 6th and 7th grade curriculum into a single year, leading to Algebra I in 7th grade and Geometry in 8th grade. Admission is limited to students who score above the 80th percentile on the statewide 5th-grade math test, and it has "off-ramps" after 6th and 7th grade that allow students to return to the conventional math track if they are struggling. In fact only about 40% of GEM entrants actually complete all three years, but 80% complete algebra by the end of 8th grade.

We use the admission threshold for GEM to conduct a regression discontinuity-based analysis of the effects of entering GEM on subsequent course taking and grades in middle and high school, and on post-secondary outcomes, including degree attainment and field of specialization. We have two main findings. First, the program has little or no effect on the college outcomes of boys, but relatively large effects on outcomes for girls, leading to relatively large increases in the probability of majoring in math-intensive fields (STEM and business /economics). Interestingly, the impacts are about the same within gender groups for students with higher and lower predicted probabilities of successfully completing all three years of GEM. Second, the impacts of GEM are about the same for economically disadvantaged students (who were receiving free or reduced price meals in 5th grade) and more advantaged students. Thus, GEM helps narrow gender gaps without widening disparities across income groups.

The main impacts of GEM for girls are summarized in Figure 12. We show the means of each of 7 outcomes for the *untreated compliers* – the set of students who form the control group

in a fuzzy RD design – as well as the causal impacts of entry to GEM on the *treated compliers* – the set of students who form the treatment group. Starting in middle school, GEM entry leads to a 22 ppt increase in the probability of enrolling in algebra by 8th grade, relative to a baseline of 60% for the untreated compliers. The rather high baseline illustrates the fact that most participants in GEM would have been in the advanced math sequence in the absence of GEM, which allows students who perform well in 7th grade to take algebra in 8th grade. Just over one-half of girls who enter GEM also enroll in geometry in 8th grade. By comparison, only 30% of boys make it all the way to the 3rd year of GEM.

In high school, participation in GEM leads to a 10 ppt increase in the fraction of girls who successfully complete calculus, and a 3.5 percentile increase in their PSAT math scores. Among students with math scores just below the GEM threshold, boys have higher PSAT math scores than girls. But for those with scores just above the threshold, GEM pushes girls' scores up to the level of boys.

At the college level, GEM leads to an insignificant (+4 ppt) gain in 5-year bachelor degree completion rates for girls but significant changes in their major choice. Most importantly, exposure to GEM leads to a 70% rise in the fractions of girls who graduate with bachelor's degrees in STEM and business or economics. Entry to GEM leads nearly 25% of girls to graduate in the relatively highly paid STEM and business-related fields – closing the gap with boys and potentially having large effects on the future careers of these young women.

A widespread concern with tracking programs like GEM is that they could harm lower-performing students who miss out on the advanced track. These concerns have led some school districts to eliminate math tracking in middle school, despite the resulting problems in meeting the "hidden curriculum" requirements for selective engineering and science programs. To address this concern, we use data from a second large school district in the same state, which has regular and advanced tracks in middle school math but not a third upper track like GEM for the highest performers in math. We find no evidence that students of either gender with scores just below the GEM threshold perform worse in the District than in this other school district.

Overall, we conclude that the introduction of a targeted accelerated math curriculum like GEM has the potential to help close the gender gap in the selection of STEM majors and higher-paid college majors in general, without hindering the academic performance of lower-performing students.

References

Abadie, Alberto. 2002. "Bootstrap tests for distributional treatment effects in instrumental variable models." *Journal of the American Statistical Association* 97 (457): 284–92.

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–96.

Altonji, Joseph G., Peter Arcidiacano and A. Maurel. 2016. The determinants of field choice in college and graduate school: Determinants and wage effects." In Eric A. Hanushek, Stephen Machin and Ludger Woessmann, editors. *Handbook of the Economics of Education* Volume 5. Elsevier: New York: 305-396.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82 (6): 2295–2326.

Card, David and Laura Giuliano. 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review* 106(10): 2783–2816.

Card, David, Eric Chyn and Laura Giuliano. 2024. "Can gifted education help higher-ability boys from disadvantaged backgrounds?" UC Santa Cruz Department of Economics Working Paper, September 2024.

Card, David and A. Abigail Payne. 2021. "High school choices and the gender gap in STEM." *Economic Inquiry* 59(1): 9-28.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2015. "The aftermath of accelerating algebra: Evidence from district policy initiatives." *Journal of Human Resources* 50 (1): 159-188.

Cohodes, Sarah R. 2020. "The Long-Run Impacts of Specialized Programming for High-Achieving Students." *American Economic Journal: Economic Policy* 12(1): 127-166.

Denning Jeffrey T., Richard Murphy and Felix Weinhardt. 2023. "Class Rank and Long-Run Outcomes." *Review of Economics and Statistics* 105(6): 1426-1441.

DesJardins, Stephen L., Dennis A. Ahlburg, and Brian P. McCall. 2002. "A Temporal Investigation of Factors Related to Timely Degree Completion." *Journal of Higher Education* 73 (5): 555-581.

Domina, Thurston, Andree McEachin, Andrew Penner and Emily Penner. 2015. "Aiming high and falling short: California's eighth-grade algebra-for-all effort." *Educational Evaluation and Policy Analysis* 37 (3): 275–295.

Dougherty, Shaun M., Joshua S. Goodman, Darryl V. Hill, Erica G. Litke and Lindsay C. Page. 2017. "Objective course placement and college readiness: Evidence from targeted middle school math acceleration." *Economics of Education Review* 58: 141–161.

Ellison, Glenn, and Ashley Swanson. 2021. "Dynamics of the Gender Gap in High Math Achievement." *Journal of Human Resources* 58(5): 1679-1711.

Hahn, Jinyong, Petra Todd Wilbert van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* 69 (1): 201-209.

Huffaker, Christopher. 2023. "Cambridge schools are divided over middle school algebra." *Boston Globe* July 14, 2023.

Huffaker, Elizabeth, Sarah Novicoff and Thomas S. Dee. 2024. "Ahead of the game? Course-taking patterns under a math pathways reform." Stanford University Graduate School of Education Working Paper. June 2024.

Lee, David S. 2008. "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics* 142: 675–697.

Lee, David S., Justin McCrary, Marcelo J. Moreira and Jack Porter. 2022. "Valid t-ratio inference for IV." *American Economic Review*, 112 (10): 3260–90.

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

McNally, Sandra 2020. "Gender differences in tertiary education: What explains STEM participation?" IZA Policy Paper No. 165. October 2020.

National Mathematics Advisory Panel. 2008. Foundations for success: The final report of the National Mathematics Advisory Panel. U.S. Department of Education: Washington, DC 20008.

Peters, Scott J. and James A. Carter. 2023. "The where and why of accelerated middle-school math." *Gifted Child Quarterly* 67(3) 235–243.

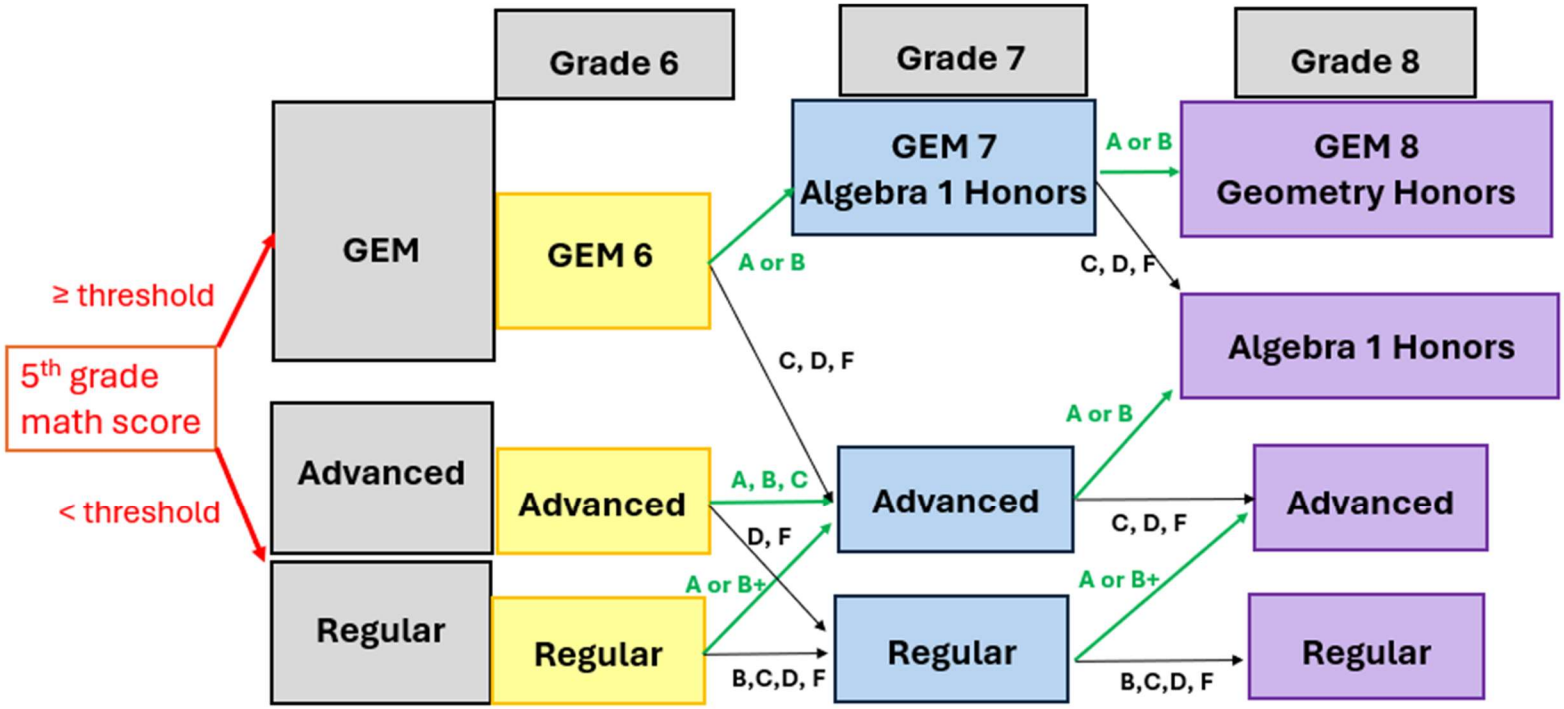
Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, editors, *Handbook of the Economics of Education*. Volume 3. New York: Elsevier.

Speer, Jamin D. 2017. "The gender gap in college major: Revisiting the role of pre-college factors." *Labour Economics* 44: 69-88.

United State Department of Education National Center for Education Statistics. 2022. *Digest of Education Statistics.* Available at https://nces.ed.gov/programs/digest/. Accessed September 9 2024.

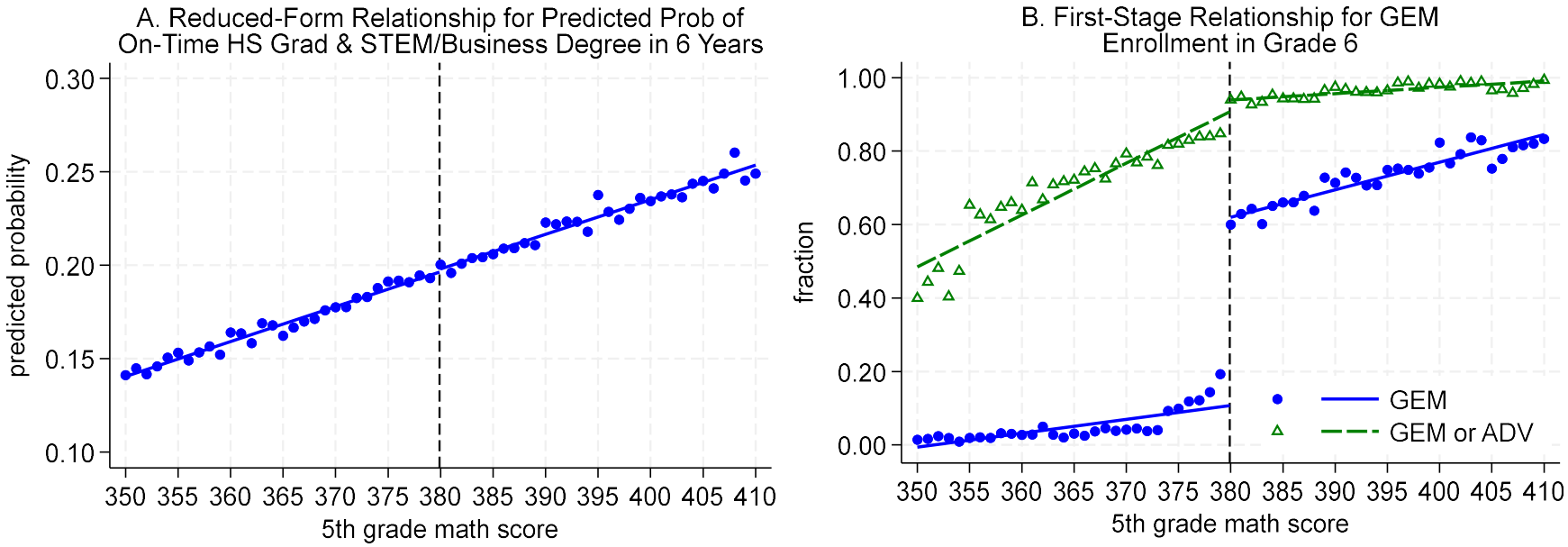
Yoder, Steven. 2024. "Eliminating advanced math tracks often prompts outrage. Some districts buck the trend." *The Hechinger Report* April 17, 2024.

Figure 1. Math Pathways in Middle School



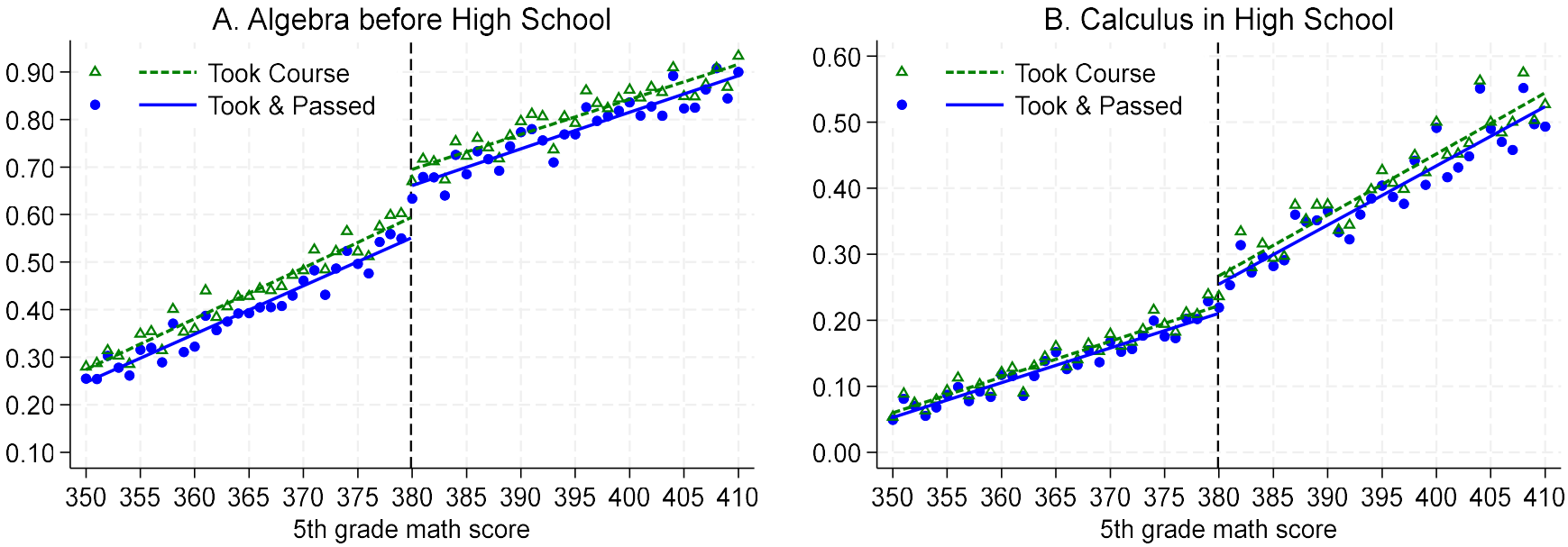
Notes: This figure illustrates various math pathways available to students in all District middle schools.

Figure 2. Predicted Outcome and GEM Enrollment by Fifth-Grade Math Score, Pooled Sample



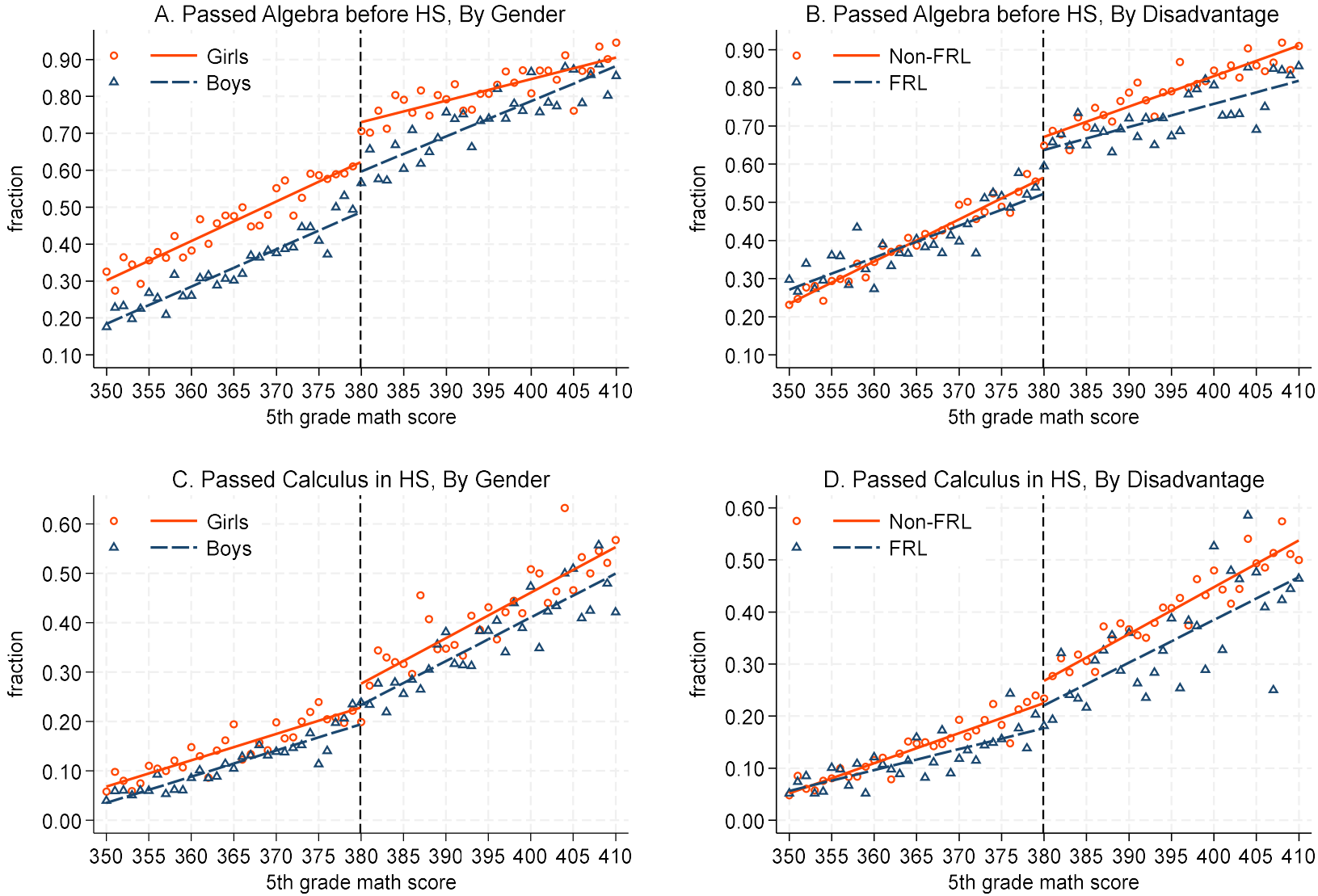
Notes: This figure plots sample means of the dependent variable at each value of 5th-grade math score (the running variable) along with fitted values from local linear RD models with a symmetric bandwidth of 30. In Panel A, the dependent variable is an index of pre-determined student characteristics constructed as the predicted probability of graduating from high school on time and earning a bachelor's degree in STEM or Business/Economics within 6 years of high-school graduation. The prediction model is fit using students with 5th-grade math scores below the GEM eligibility threshold of 380 and includes 2nd-order polynomials in 4th-grade math test scores, 4th-grade reading test scores, and student age; indicators for student gender, five race/ethnicity categories, FRL status, English Language Learner status, and cohort; and the interactions of gender with 4th-grade test scores and FRL status. In panel B, the dependent variables are an indicator for participating in the GEM program in 6th grade (the treatment variable, in blue circles) and an indicator for participating in *either* GEM *or* advanced-track math (in green triangles).

Figure 3. Reduced-Form Relationship for High School Outcomes, Pooled Sample



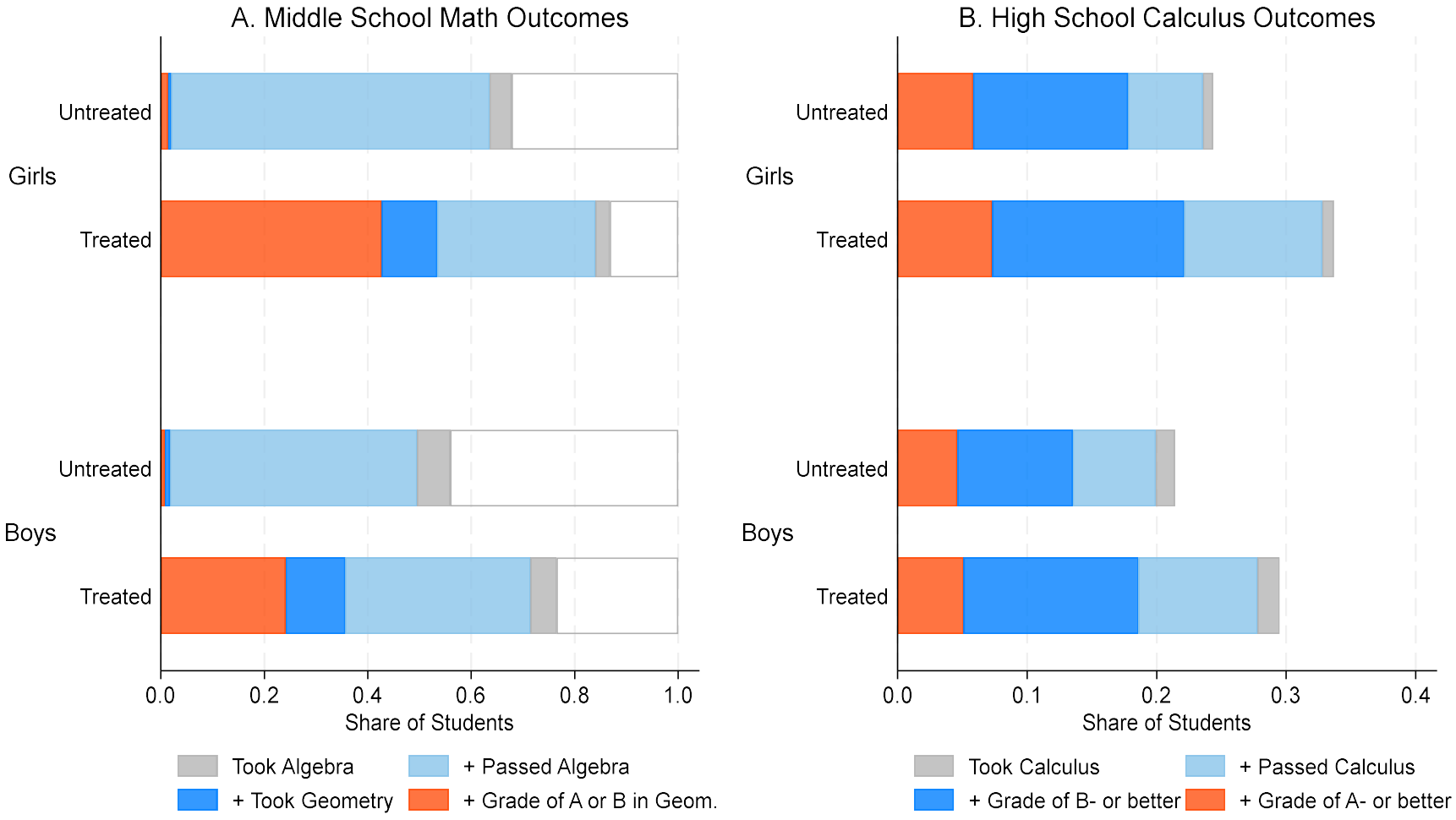
Notes: This figure plots sample means of the dependent variable at each value of fifth-grade math score (the running variable) along with fitted values from local linear RD models with a symmetric bandwidth of 30. In Panel A, the dependent variables are indicators for having enrolled algebra in middle school (grade 7 or 8) and having completed it with a passing grade (C- or better). In Panel B, the dependent variables are having enrolled in calculus in high school (grades 9-12) and having completed it with a passing grade.

Figure 4. Reduced-Form Relationship for Successful Completion of Early Algebra / HS Calculus, by Gender and Disadvantage



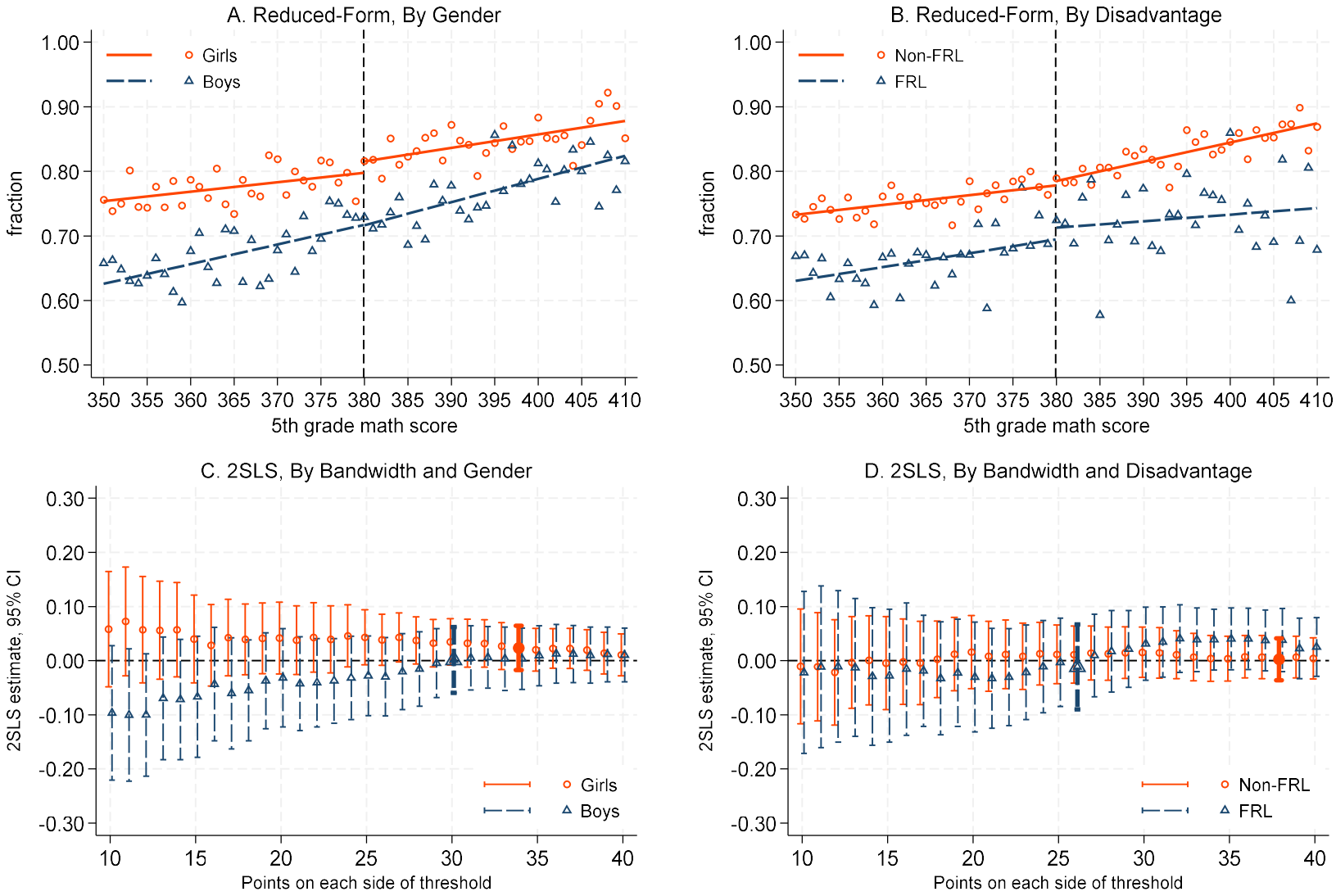
Notes: This figure plots sample means of the dependent variable at each value of 5th-grade math score (the running variable), along with fitted values from local linear RD models with a symmetric bandwidth of 25, separately by gender (Panels A and C) and by FRL status (Panels B and D). In Panels A and B, the dependent variable is an indicator for taking algebra in middle school (grade 7 or 8) and earning a passing grade (C- or better). In Panels C and D, the dependent variable is an indicator for taking calculus in high school and earning a passing grade (C- or better).

Figure 5. Outcome Distributions for Treated and Untreated Compliers, By Gender



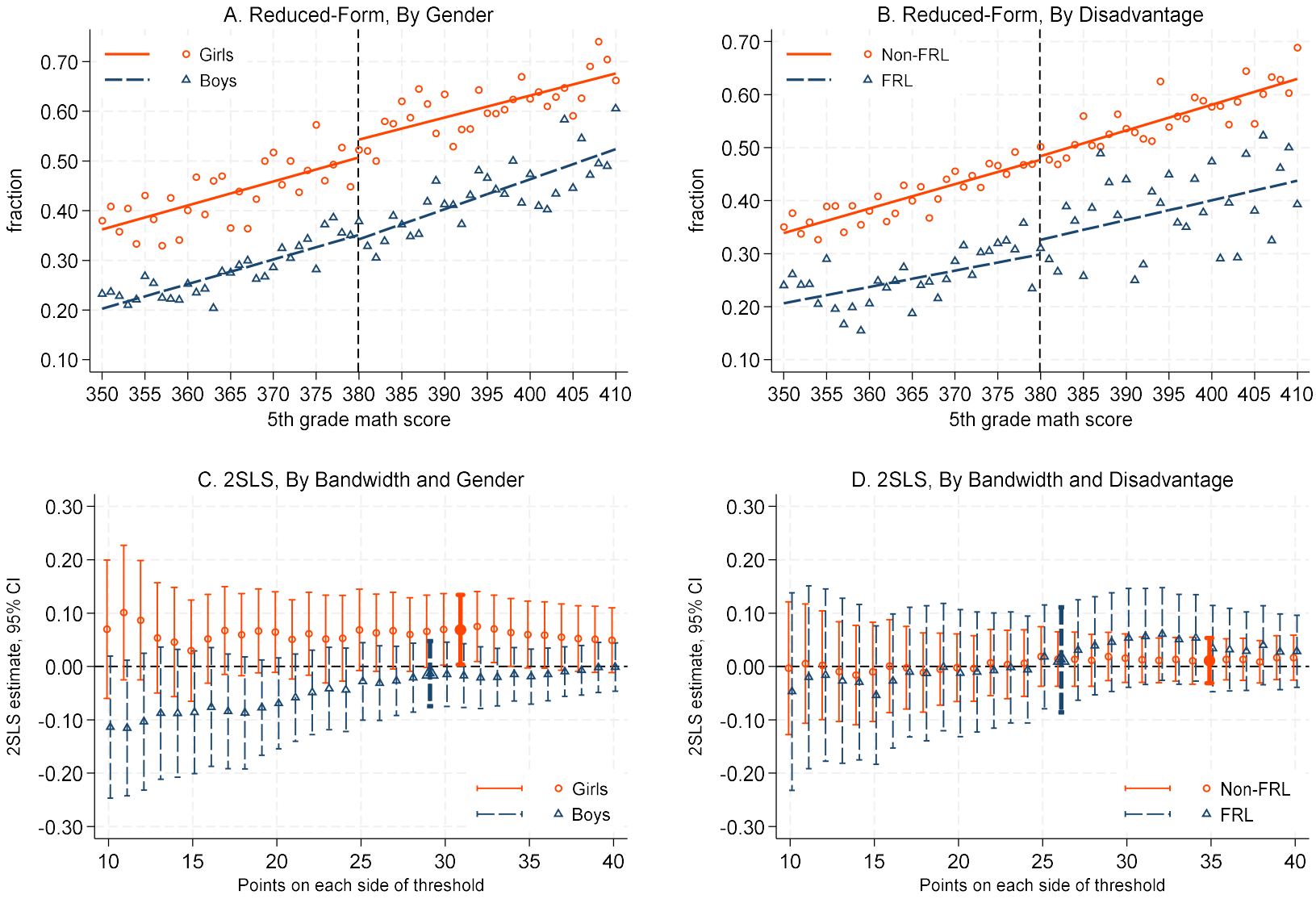
Notes: This figure shows the estimated shares of treated (GEM participant) and untreated (non-participant) compliers who achieve various math course outcomes. The estimates are constructed by applying the approach of Abadie (2002) to an RD setting (see text for details). In panel A, taking algebra is defined as enrolling in algebra in either 7th grade or 8th grade. In both panels, passing Algebra (Calculus) is defined as earning a C or better. For students who took Algebra twice, this variable is based on the grade the second time they took it.

Figure 6. Graduate High School on Time and Enter College within 1 Year, by Gender and Disadvantage



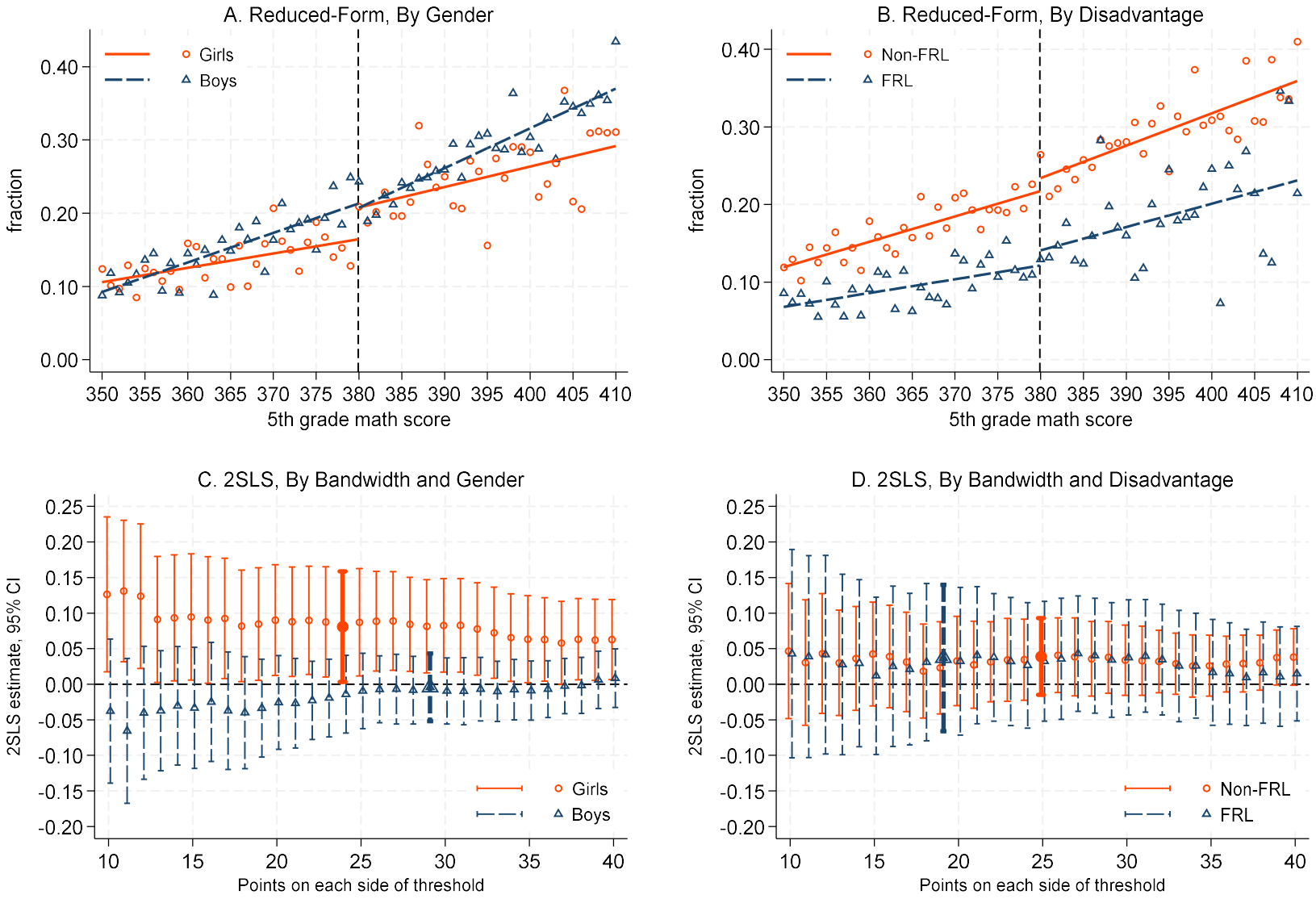
Notes: Panels A and B plot sample means of an indicator for graduating high school on time and enrolling in college within 1 year at each value of 5th-grade math score (the running variable), along with fitted values from local linear RD models with a symmetric bandwidth of 30, separately by gender (Panels A) and by FRL status (Panels B). Panels C and D plot 2SLS estimates and 95% confidence intervals, for the treatment effect of GEM entry on the dependent variable, for symmetric bandwidths between 10 and 40. Estimates that use the CCT symmetric bandwidth are highlighted with solid-filled markers and thicker lines.

Figure 7. Earn a Bachelors Degree in Any Field Within 6 Years, by Gender and Disadvantage



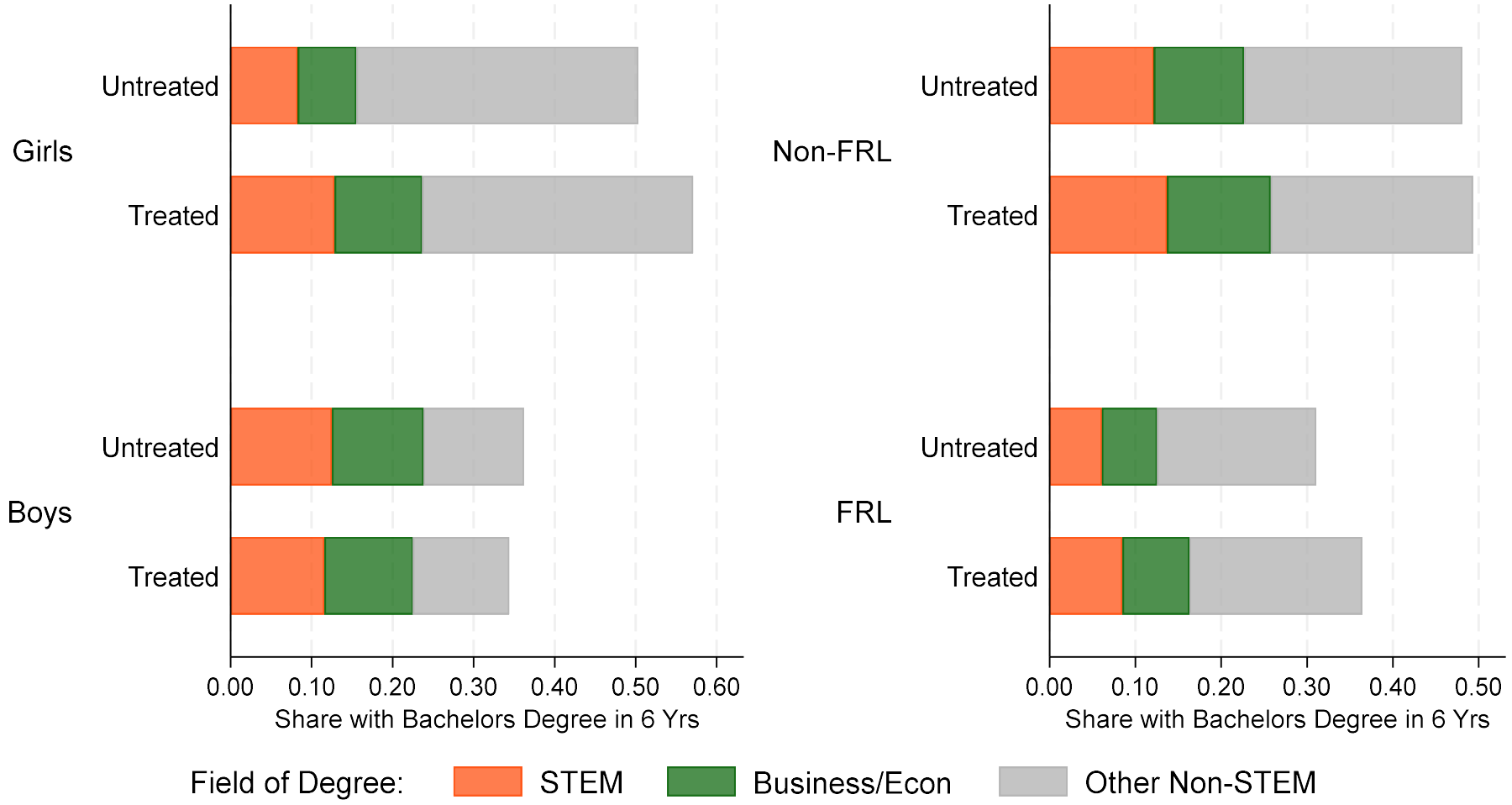
Notes: Panels A and B plot sample means of an indicator for graduating high school on time and earning any college degree within 6 years at each value of 5th-grade math score (the running variable), along with fitted values from local linear RD models with a symmetric bandwidth of 30, separately by gender (Panels A) and by FRL status (Panels B). Panels C and D plot 2SLS estimates and 95% confidence intervals, for the treatment effect of GEM entry on the dependent variable, for symmetric bandwidths between 10 and 40. Estimates that use the CCT symmetric bandwidth are highlighted with solid-filled markers and thicker lines.

Figure 8. Earn a Bachelors Degree in STEM or Business/Econ within 6 Years, Gender and Disadvantage



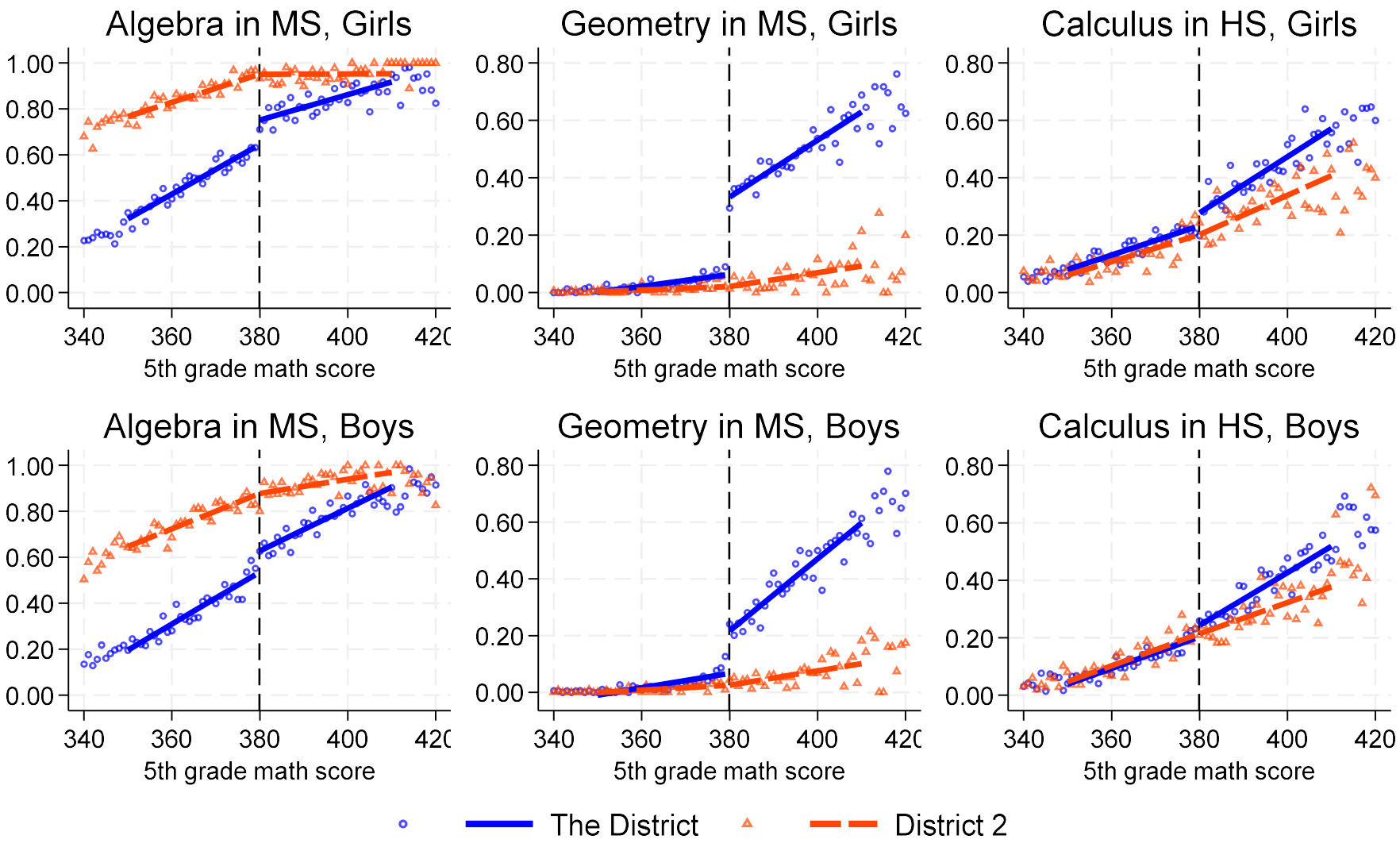
Notes: Panels A and B plot sample means of an indicator for graduating high school on time and earning any college degree within 6 years at each value of 5th-grade math score (the running variable), along with fitted values from local linear RD models with a symmetric bandwidth of 30, separately by gender (Panels A) and by FRL status (Panels B). Panels C and D plot 2SLS estimates and 95% confidence intervals, for the treatment effect of GEM entry on the dependent variable, for symmetric bandwidths between 10 and 40. Estimates that use the CCT symmetric bandwidth are highlighted with solid-filled markers and thicker lines.

Figure 9. Shares of Treated and Untreated Compliers with Bachelors Degrees, by Field



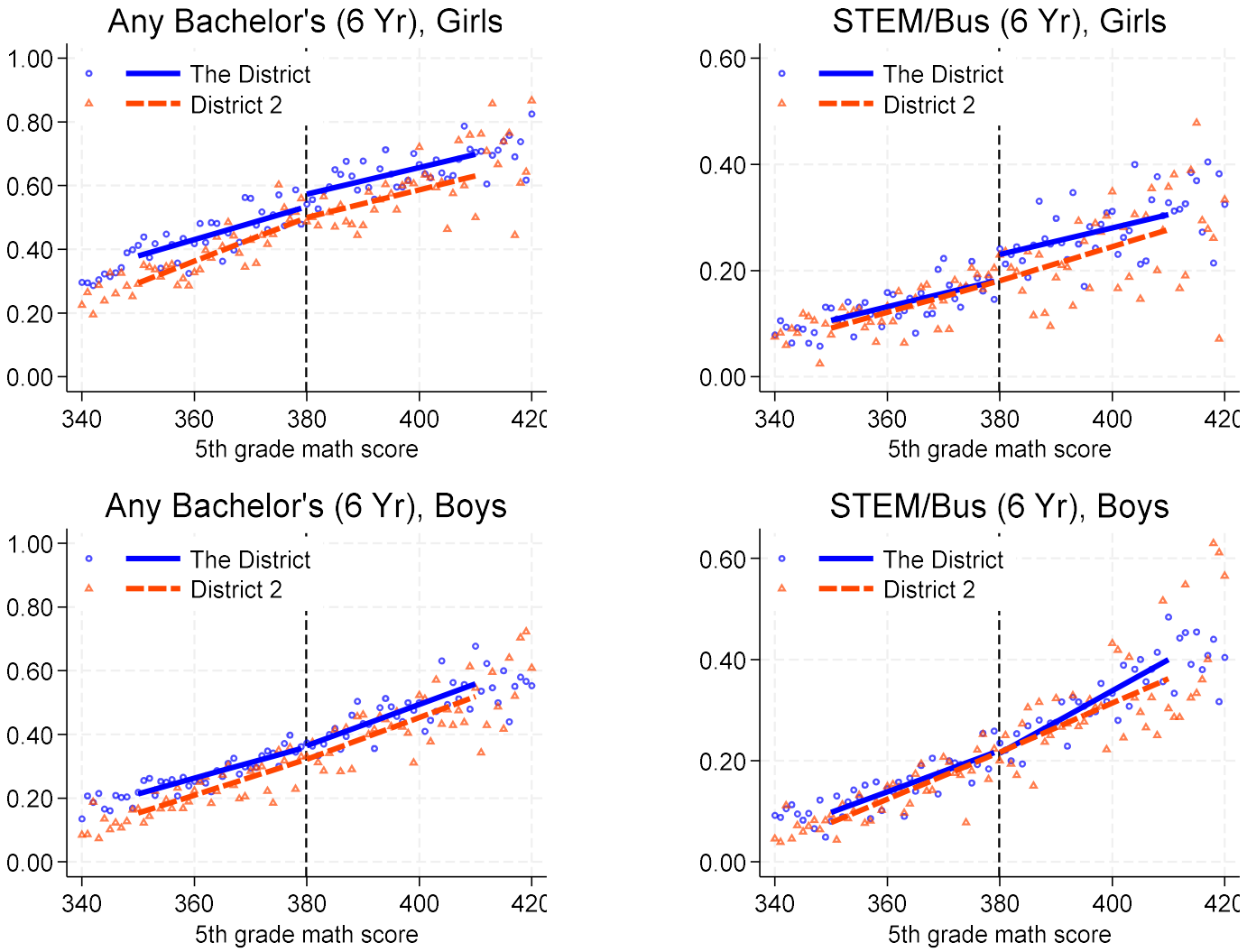
Notes: This figure shows the estimated shares of treated (GEM participant) and untreated (non-participant) compliers who earned a bachelors degree within 6 years of on-time high school graduation, separately by major field. The estimates are constructed by applying the approach of Abadie (2002) to an RD setting (see text for details).

Figure 10. Math Course Enrollments, Between-District Comparisons by Gender



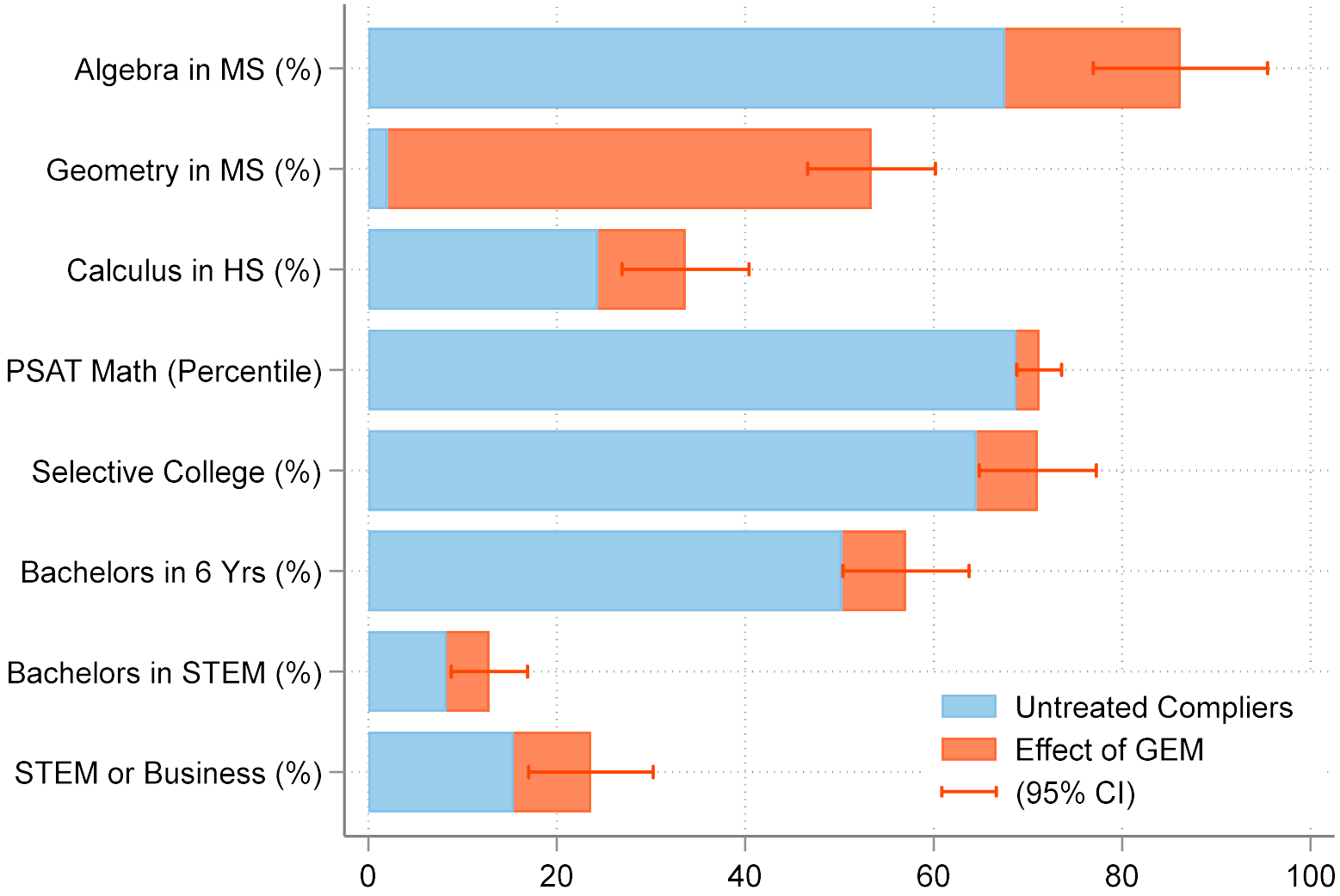
Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure 11. College Outcomes, Between-District Comparisons by Gender



Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure 12. Summary of Effects of GEM Entry on Girls



Notes: This figure shows estimated means of key outcomes for female untreated compliers (in blue) along with 2SLS estimates and 95% confidence intervals of the corresponding treatment effects for female treated compliers (in orange). The estimates are constructed by applying the approach of Abadie (2002) to an RD setting (see text for details).

Table 1. Estimated Means for Main Analysis Sample and for Untreated Compliers

	Main Analysis Sample	Untreated Compliers (Not in GEM)				
A. Baseline Characteristics		**All**	**Girls**	**Boys**	**Non-FRL**	**FRL**
Demographic characteristics						
Female	0.49	0.48	1.00	0.00	0.46	0.52
FRL	0.30	0.29	0.32	0.27	0.00	1.00
White	0.44	0.42	0.39	0.45	0.52	0.18
Black	0.23	0.22	0.27	0.17	0.13	0.44
Hispanic	0.25	0.26	0.23	0.29	0.25	0.28
Gifted	0.10	0.07	0.06	0.07	0.06	0.09
Baseline achievement						
grade 4 math score (standardized)	0.68	0.79	0.76	0.82	0.82	0.72
grade 4 reading score (standardized)	0.57	0.69	0.74	0.64	0.72	0.61
School peer characteristics						
avg grade 4 math & reading scores	0.14	0.14	0.13	0.15	0.20	0.00
fraction FRL	0.40	0.41	0.42	0.40	0.35	0.56
fraction GEM eligible	0.22	0.22	0.21	0.22	0.23	0.18
Predicted prob graduate high school on time & earn bachelors in STEM or Bus/Econ in 6 yrs	0.19	0.20	0.19	0.21	0.23	0.13
B. Mean intermediate outcomes or potential outcomes						
Math track in 6th grade						
GEM (treatment)	0.29	0.00	0.00	0.00	0.00	0.00
advanced	0.49	0.94	0.96	0.92	0.96	0.91
regular	0.22	0.06	0.04	0.08	0.04	0.09
Middle school course enrollments & grades						
algebra by 8th grade	0.56	0.61	0.68	0.56	0.64	0.56
completed algebra with B- or better by 8th grade	0.40	0.43	0.52	0.35	0.45	0.40
completed algebra with C- or better by 8th grade	0.53	0.56	0.64	0.50	0.59	0.51
geometry in 8th grade	0.18	0.02	0.02	0.02	0.02	0.02
6th grade math GPA	3.11	3.43	3.57	3.30	3.49	3.27
7th grade math GPA	3.01	3.17	3.36	3.01	3.21	3.09
8th grade GPA	2.89	2.96	3.10	2.83	3.00	2.86
High school course enrollments & grades						
took any advanced math (statistics, calculus, etc.)	0.64	0.69	0.74	0.64	0.73	0.61
took any calculus	0.23	0.23	0.24	0.21	0.24	0.19
completed any calculus with B- or better	0.16	0.16	0.18	0.14	0.17	0.13
completed any calculus with C- or better	0.22	0.22	0.24	0.20	0.23	0.18
took AP/IB calculus	0.15	0.14	0.14	0.14	0.15	0.11
took AP calculus exam + scored 3 or better	0.08	0.06	0.06	0.07	0.07	0.05
high school math GPA (all grades, all math courses)	3.22	3.27	3.40	3.15	3.29	3.21
College boards						
PSAT math percentile	64.8	69.5	68.8	70.3	70.8	66.4
PSAT reading percentile	59.8	62.6	64.5	60.8	64.3	58.6
Took SAT or ACT	0.88	0.89	0.93	0.85	0.90	0.85

Note: Table continues; see notes on next page.

Table 1, Cont'd. Estimated Means for Main Analysis Sample and for Untreated Compliers

	Main Analysis Sample	Untreated Compliers (Not in GEM)				
		All	Girls	Boys	Non-FRL	FRL
C. Mean post-secondary outcomes or potential outcome						
College enrollment						
graduate high school on time	0.95	0.97	0.97	0.96	0.98	0.94
+ enter any college within 1 year	0.75	0.76	0.79	0.74	0.79	0.70
+ enter selective college within 1 year	0.57	0.60	0.65	0.57	0.65	0.50
College degree earned within 6 years						
+ any college degree	0.54	0.55	0.61	0.49	0.59	0.45
+ bachelors (BA or BS)	0.41	0.43	0.50	0.36	0.48	0.31
+ BA/BS in STEM	0.09	0.10	0.08	0.13	0.12	0.06
+ BA/BS in Business/Econ	0.09	0.09	0.07	0.11	0.10	0.06
+ BA/BS in STEM or Bus/Econ	0.19	0.19	0.15	0.24	0.22	0.12

Notes: The first column shows means of variables for all students in the main analysis sample (with math scores between 350 and 410). The remaining columns show estimated mean characteristics of all the compliers (Panel A), or estimated mean potential outcomes for compliers if they do not enter GEM (Panels B & C). Complier characteristics and potential outcomes are estimated by applying the method of Abadie (2002) to an RD setting (see text for details).

Table 2. Estimated Treatment Effects on Math Course Enrollment and Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Predicted STEM/Bus Degree in 6 Years	First Stage (GEM Entry)	Course Enrollments			Took Algebra in MS & Earned:		Took Calculus in HS & Earned:		Grade Point Avg in Math Courses			
			Algebra in MS	Geometry in MS	Calculus in HS	B- or better	C- or better	B- or better	C- or better	6th Grade	7th Grade	8th Grade	All HS Math
1. Pooled	**0.002**	**0.513****	**0.198****	**0.427****	**0.090****	**0.149****	**0.215****	**0.049***	**0.088****	**-0.666****	**-0.372****	**-0.047**	**-0.019**
(N = 23,225)	(0.002)	(0.028)	(0.040)	(0.032)	(0.025)	(0.030)	(0.035)	(0.022)	(0.024)	(0.078)	(0.053)	(0.054)	(0.039)
with controls			0.190**	0.419**	0.082**	0.137**	0.206**	0.040+	0.080**	-0.684**	-0.387**	-0.062	-0.028
			(0.039)	(0.032)	(0.024)	(0.031)	(0.035)	(0.022)	(0.024)	(0.077)	(0.051)	(0.056)	(0.039)
2. Girls	**0.004**	**0.526****	**0.187****	**0.514****	**0.094****	**0.154****	**0.205****	**0.044**	**0.093****	**-0.642****	**-0.439****	**-0.043**	**-0.016**
(N =11,488)	(0.002)	(0.028)	(0.047)	(0.035)	(0.034)	(0.047)	(0.045)	(0.031)	(0.035)	(0.082)	(0.059)	(0.066)	(0.046)
with controls			0.184**	0.510**	0.091**	0.147**	0.201**	0.040	0.089*	-0.653**	-0.450**	-0.054	-0.021
			(0.046)	(0.035)	(0.034)	(0.047)	(0.044)	(0.031)	(0.034)	(0.081)	(0.058)	(0.066)	(0.045)
3. Boys	**0.001**	**0.500****	**0.207****	**0.340****	**0.085****	**0.142****	**0.222****	**0.053+**	**0.082****	**-0.697****	**-0.327****	**-0.064**	**-0.031**
(N =11,737)	(0.003)	(0.032)	(0.044)	(0.039)	(0.031)	(0.034)	(0.039)	(0.027)	(0.031)	(0.088)	(0.066)	(0.078)	(0.065)
with controls			0.198**	0.329**	0.074*	0.131**	0.214**	0.042	0.071*	-0.708**	-0.330**	-0.068	-0.038
			(0.043)	(0.039)	(0.030)	(0.033)	(0.037)	(0.026)	(0.030)	(0.085)	(0.065)	(0.078)	(0.064)
4. Non-FRL	**0.002**	**0.513****	**0.187****	**0.447****	**0.085****	**0.164****	**0.208****	**0.059***	**0.086****	**-0.652****	**-0.327****	**-0.067**	**-0.017**
(N = 16,235)	(0.002)	(0.031)	(0.050)	(0.039)	(0.029)	(0.039)	(0.043)	(0.025)	(0.027)	(0.087)	(0.056)	(0.061)	(0.045)
with controls			0.179**	0.440**	0.077**	0.151**	0.198**	0.050*	0.078**	-0.670**	-0.344**	-0.085	-0.029
			(0.049)	(0.039)	(0.028)	(0.039)	(0.042)	(0.024)	(0.027)	(0.088)	(0.053)	(0.063)	(0.045)
5. FRL	**-0.000**	**0.512****	**0.221****	**0.373****	**0.097***	**0.105***	**0.227****	**0.020**	**0.086***	**-0.707****	**-0.485****	**-0.006**	**-0.029**
(N = 6,990)	(0.002)	(0.033)	(0.041)	(0.040)	(0.039)	(0.040)	(0.038)	(0.032)	(0.038)	(0.101)	(0.103)	(0.081)	(0.071)
with controls			0.214**	0.366**	0.092*	0.098*	0.222**	0.014	0.080*	-0.715**	-0.482**	0.001	-0.025
			(0.045)	(0.038)	(0.038)	(0.045)	(0.042)	(0.031)	(0.037)	(0.093)	(0.103)	(0.082)	(0.069)

Notes: Table entries are 2SLS estimates for the effect of entry into GEM in 6th grade (i.e., they are estimates of β in equation 3). The baseline models include 4 dummy variables for the student's cohort but no other controls. The models with controls include dummies for student gender, FRL and ELL status, race, and gifted status; age; and 2nd-order polynomials in 4th-grade math and reading scores. Standard errors clustered on the middle school where the student was enrolled in 6th grade.

Table 3. Estimated Treatment Effects on Post-Secondary Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	On-Time High School Grad	On-Time College Entry	Enroll in Selective College	Any College Degree in 6 Years	Any Bachelors Degree in 6 Years	Bachelors degree within 6 years in (field): STEM	Bus/Econ	STEM or Bus/Econ
1. Pooled	**-0.006**	**0.020**	**0.030**	**0.031**	**0.029**	**0.019**	**0.017**	**0.037+**
(N = 23,225)	(0.009)	(0.020)	(0.024)	(0.020)	(0.021)	(0.013)	(0.015)	(0.021)
with controls	-0.007	0.014	0.019	0.022	0.019	0.018	0.015	0.034
	(0.008)	(0.019)	(0.023)	(0.020)	(0.019)	(0.013)	(0.015)	(0.021)
2. Girls	**-0.004**	**0.035**	**0.066***	**0.055+**	**0.070***	**0.046***	**0.037+**	**0.083***
(N =11,488)	(0.009)	(0.022)	(0.032)	(0.032)	(0.034)	(0.021)	(0.020)	(0.034)
with controls	-0.004	0.031	0.056+	0.047	0.059+	0.046*	0.035+	0.082*
	(0.009)	(0.022)	(0.028)	(0.031)	(0.032)	(0.020)	(0.020)	(0.032)
3. Boys	**-0.010**	**0.001**	**-0.009**	**0.003**	**-0.015**	**-0.008**	**-0.003**	**-0.009**
(N =11,737)	(0.016)	(0.031)	(0.034)	(0.027)	(0.031)	(0.018)	(0.017)	(0.024)
with controls	-0.009	-0.003	-0.016	-0.001	-0.021	-0.011	-0.004	-0.013
	(0.015)	(0.030)	(0.032)	(0.027)	(0.030)	(0.018)	(0.017)	(0.024)
4. Non-FRL	**-0.002**	**0.015**	**0.038**	**0.014**	**0.016**	**0.017**	**0.016**	**0.034**
(N = 16,235)	(0.010)	(0.024)	(0.027)	(0.028)	(0.023)	(0.016)	(0.018)	(0.025)
with controls	-0.003	0.008	0.027	0.003	0.002	0.016	0.015	0.031
	(0.010)	(0.023)	(0.027)	(0.027)	(0.022)	(0.016)	(0.017)	(0.025)
5. FRL	**-0.016**	**0.031**	**0.002**	**0.065**	**0.054**	**0.024**	**0.013**	**0.038**
(N = 6,990)	(0.020)	(0.034)	(0.045)	(0.042)	(0.047)	(0.034)	(0.026)	(0.041)
with controls	-0.014	0.034	0.003	0.071+	0.059	0.025	0.014	0.040
	(0.019)	(0.032)	(0.042)	(0.040)	(0.046)	(0.035)	(0.026)	(0.042)

Notes: Table entries are 2SLS estimates for the effect of entry into GEM in 6th grade (i.e., they are estimates of β in equation 3). The baseline models include 4 dummy variables for the student's cohort but no other controls. The models with controls include dummies for student gender, FRL and ELL status, race, and gifted status; age; and 2nd-order polynomials in 4th-grade math and reading scores. Standard errors clustered on the middle school where the student was enrolled in 6th grade.

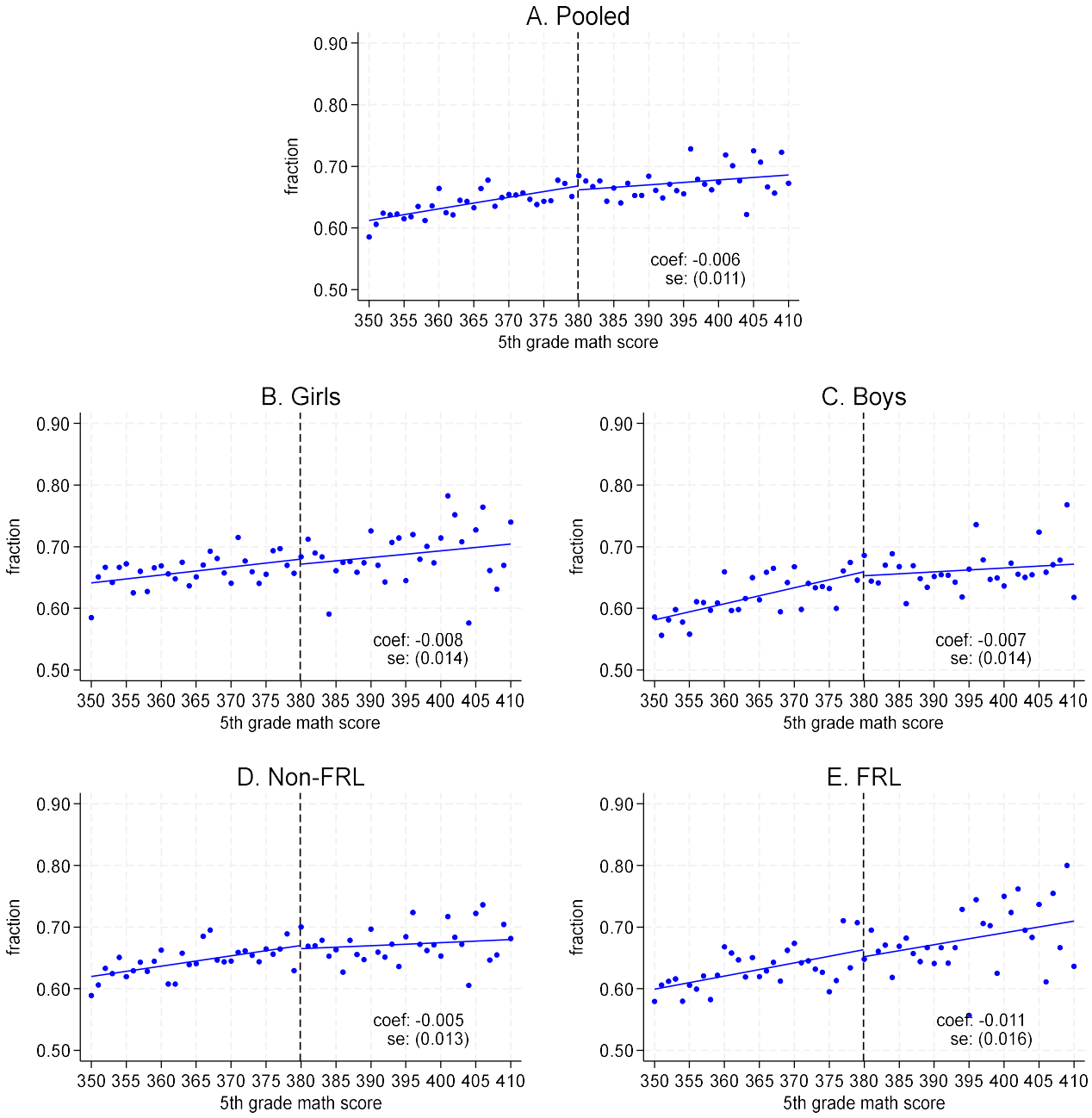
Table 4. Heterogeneous Treatment Effects, by Gender and Probability of Success in GEM

			2SLS Treatment Effect Estimates for Key Outcomes								
			Math Course Outcomes						Bachelor's Degree within 6 years in (field):		
	Predicted STEM/Bus Degree in 6 Years	First Stage (GEM in 6th Grade)	Passed Algebra in MS	Passed Geometry in MS	Pass Calculus in HS	On-Time College Entry	Enroll in Selective College	Bachelors Degree in 6 Years	STEM	Bus/Econ	STEM or Bus/Econ
Girls											
High P(Success)	**0.004**	**0.590****	**0.167****	**0.574****	**0.079+**	**0.007**	**0.016**	**0.057**	**0.050**	**0.033**	**0.083***
(N=5,753)	(0.003)	(0.029)	(0.047)	(0.041)	(0.044)	(0.027)	(0.032)	(0.045)	(0.032)	(0.026)	(0.041)
with controls			0.166**	0.572**	0.076+	0.005	0.005	0.048	0.051	0.032	0.083*
			(0.047)	(0.040)	(0.044)	(0.028)	(0.032)	(0.043)	(0.031)	(0.025)	(0.040)
Low P(Success)	**0.002**	**0.461****	**0.254****	**0.407****	**0.110***	**0.070+**	**0.130***	**0.084**	**0.041**	**0.041**	**0.082**
(N=5,730)	(0.002)	(0.033)	(0.064)	(0.040)	(0.045)	(0.039)	(0.059)	(0.055)	(0.035)	(0.028)	(0.050)
with controls			0.246**	0.401**	0.104*	0.061	0.120*	0.070	0.039	0.040	0.079
			(0.065)	(0.041)	(0.045)	(0.038)	(0.055)	(0.051)	(0.034)	(0.026)	(0.047)
Boys											
High P(Success)	**0.004**	**0.564****	**0.217****	**0.377****	**0.118****	**0.022**	**0.001**	**-0.032**	**-0.020**	**0.023**	**0.003**
(N=5,867)	(0.004)	(0.036)	(0.045)	(0.045)	(0.041)	(0.038)	(0.047)	(0.043)	(0.025)	(0.022)	(0.036)
with controls			0.204**	0.367**	0.107*	0.014	-0.008	-0.039	-0.019	0.022	0.003
			(0.046)	(0.045)	(0.041)	(0.037)	(0.046)	(0.042)	(0.025)	(0.022)	(0.037)
Low P(Success)	**-0.004**	**0.435****	**0.222****	**0.254****	**0.027**	**-0.032**	**-0.035**	**-0.002**	**0.003**	**-0.037**	**-0.032**
(N=5,861)	(0.003)	(0.034)	(0.056)	(0.043)	(0.037)	(0.058)	(0.060)	(0.052)	(0.035)	(0.034)	(0.045)
with controls			0.230**	0.255**	0.030	-0.021	-0.020	0.004	0.003	-0.035	-0.030
			(0.056)	(0.043)	(0.037)	(0.059)	(0.061)	(0.053)	(0.035)	(0.034)	(0.045)

Notes: Table entries are 2SLS estimates for the effect of entry into GEM in 6th grade (i.e., they are estimates of β in equation 3). The baseline models include four dummy variables for the student's cohort but no other controls. Models are estimated separately for boys and girls whose predicted probability of successfully completing the GEM program is above or below their gender-specific medians. See text (section IV.e) for details. The models with controls include dummies for student race, ELL status, and gifted status; age; and 2nd-order polynomials in 4th-grade math and reading scores. Standard errors clustered on the middle school where the student was enrolled in 6th grade.

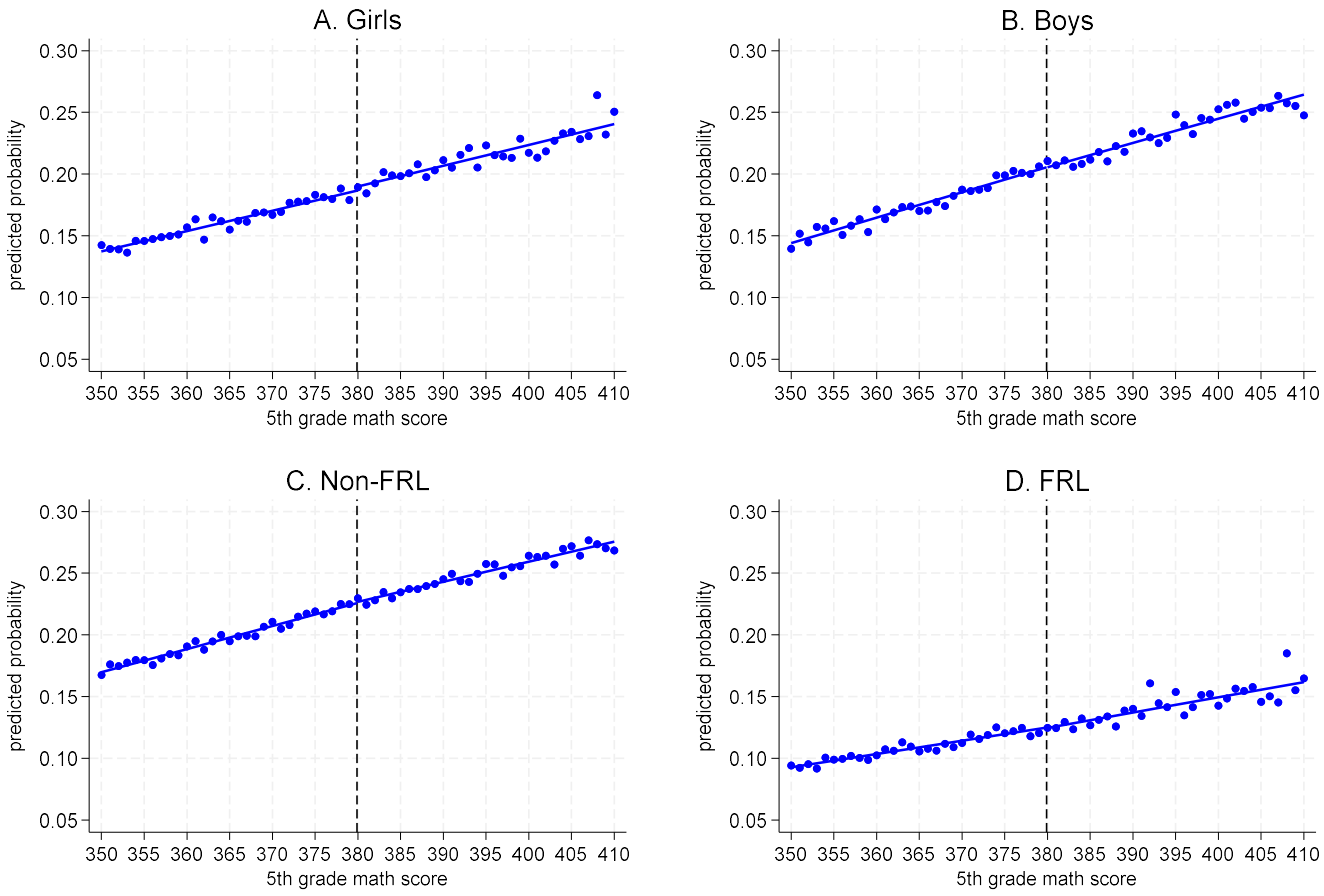
Appendix A: Additional Figures and Tables

Figure A1: Reduced-Form Relationships for Meeting Sample Criteria



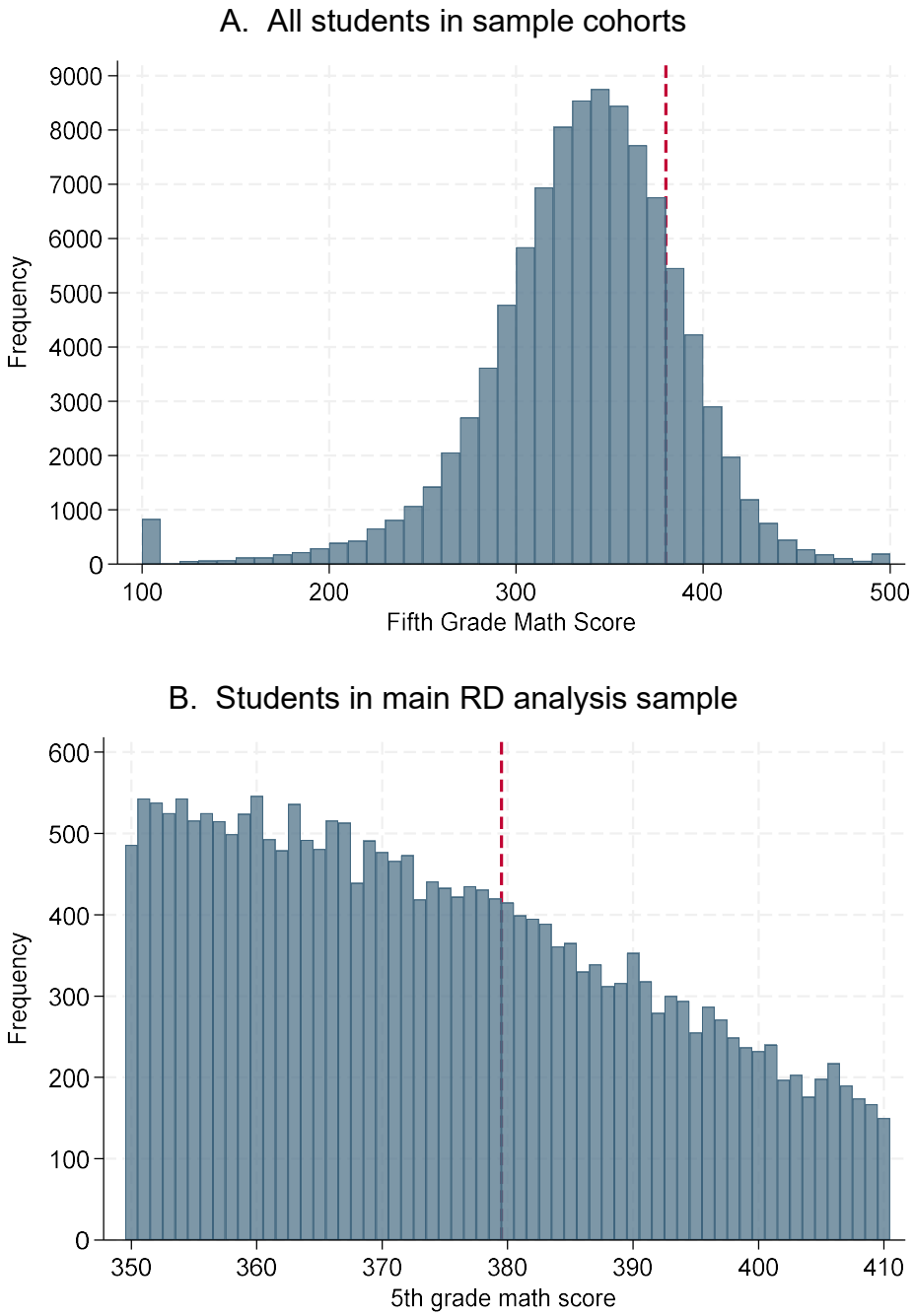
Notes: This figure plots means of an indicator for meeting the inclusion criteria for the main analysis sample (see Appendix Table A2), using all students who enrolled in the District for 5th grade between 2004 and 2008. It also shows fitted values from a local linear RD model with a symmetric bandwidth of 30. Panel A pools all demographic groups while Panels B-D show the relationships separately by gender and by FRL status. Each panel also reports the estimated discontinuity at the GEM threshold of 380, along with standard errors clustered on elementary school (in parentheses).

Figure A2: Reduced-Form Relationships for the Predicted Probability of Completing a Bachelor's Degree in STEM/Business/Economics within 6 Years of On-Time High School Graduation, by Gender and Disadvantage



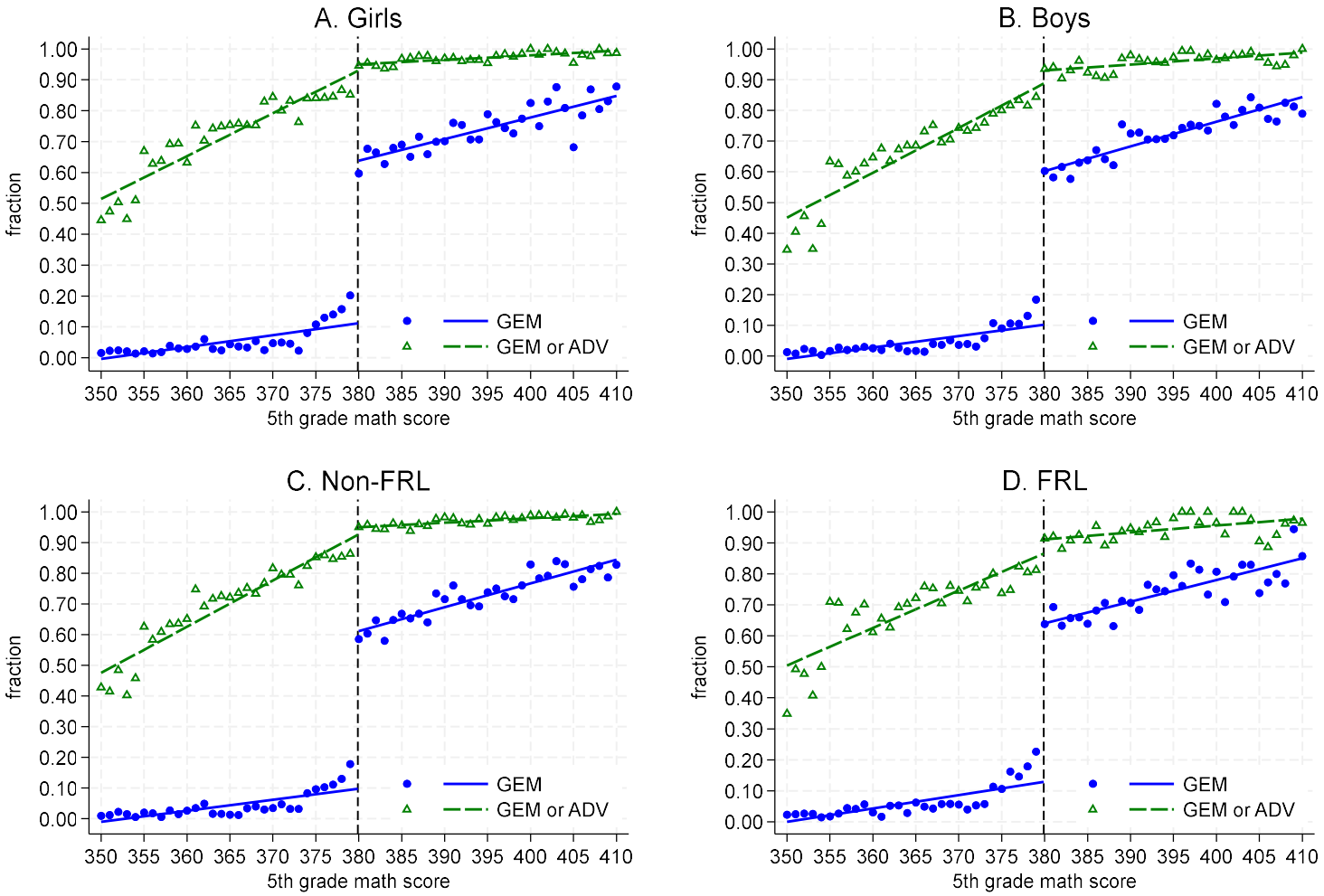
Notes: This figure plots sample means, along with fitted values from local linear RD models with symmetric bandwidths of +/- 30 points, of an index of observable pre-treatment characteristics. The index is constructed as the predicted probability of graduating high school on time and completing a bachelors degree (in any field) within 6 years, from a model fit to sample observations with 5th-grade math scores below the threshold. The model includes: second-order polynomials in 4th-grade math test scores, 4th-grade reading test scores, and student age; indicators for student gender, five race/ethnicity categories, FRL status, English Language Learner status, and cohort; and the interactions of gender with 4th-grade test scores and FRL status.

Figure A3: Histograms of Running Variable



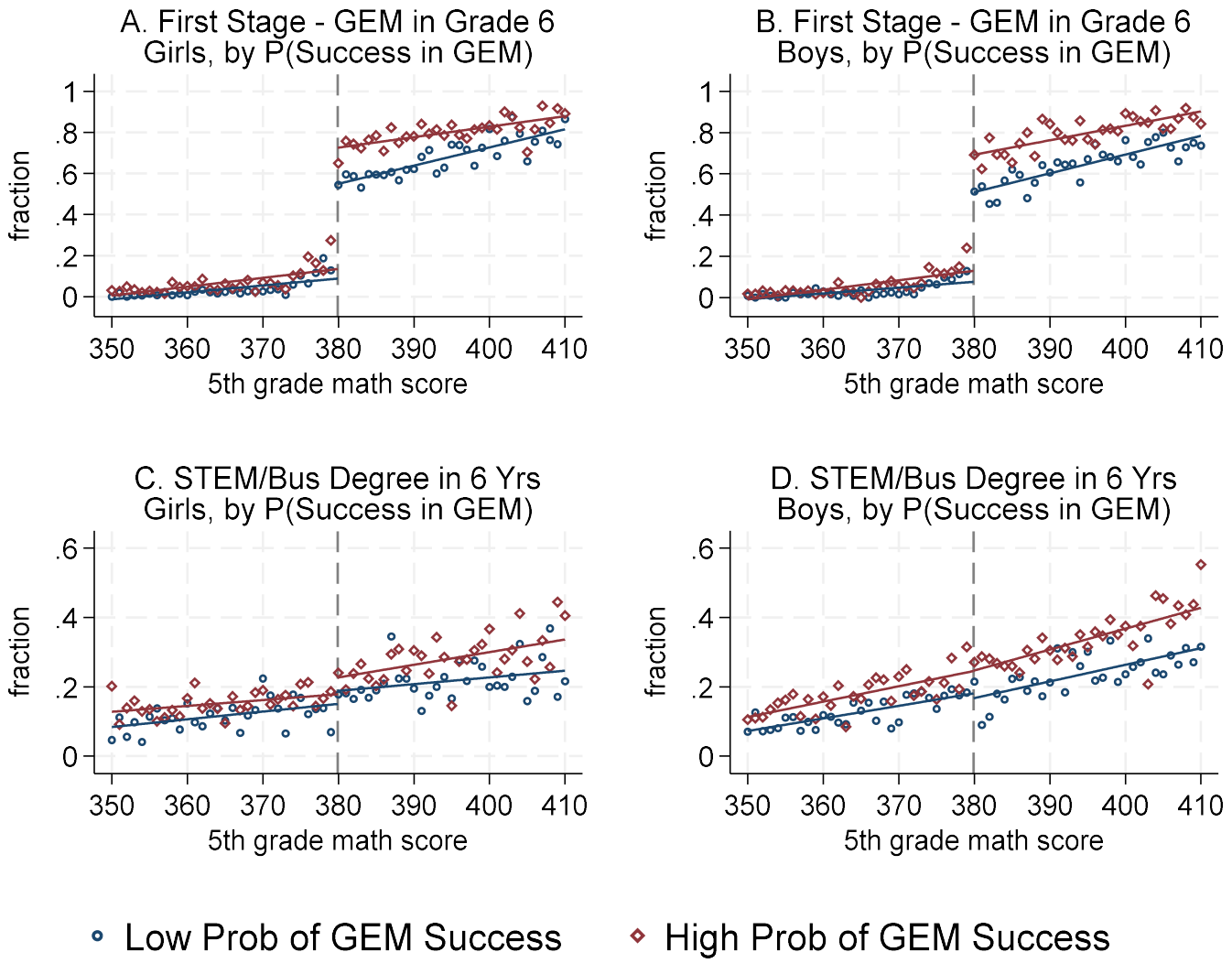
Notes: This figure shows the frequency distribution of 5th-grade math scores for all students in our sample cohorts (Panel A) and for students in the main RD analysis sample (Panel B). The sample consists of all students in the District who completed 5th grade in between 2003 and 2007, were not identified as gifted, had non-missing 4th-grade test scores, were enrolled in 6th grade in the following year (between 2004 and 2008), and were enrolled in (or graduated from) a District high school six years later (between 2010 and 2014). The RD analysis sample is further restricted to scores within +/- 30 points of the threshold for eligibility to participate in the GEM program in 6th grade. The dashed vertical lines in each panel indicate the threshold.

Figure A4: First-Stage Relationships for GEM Enrollment, by Gender and Disadvantage



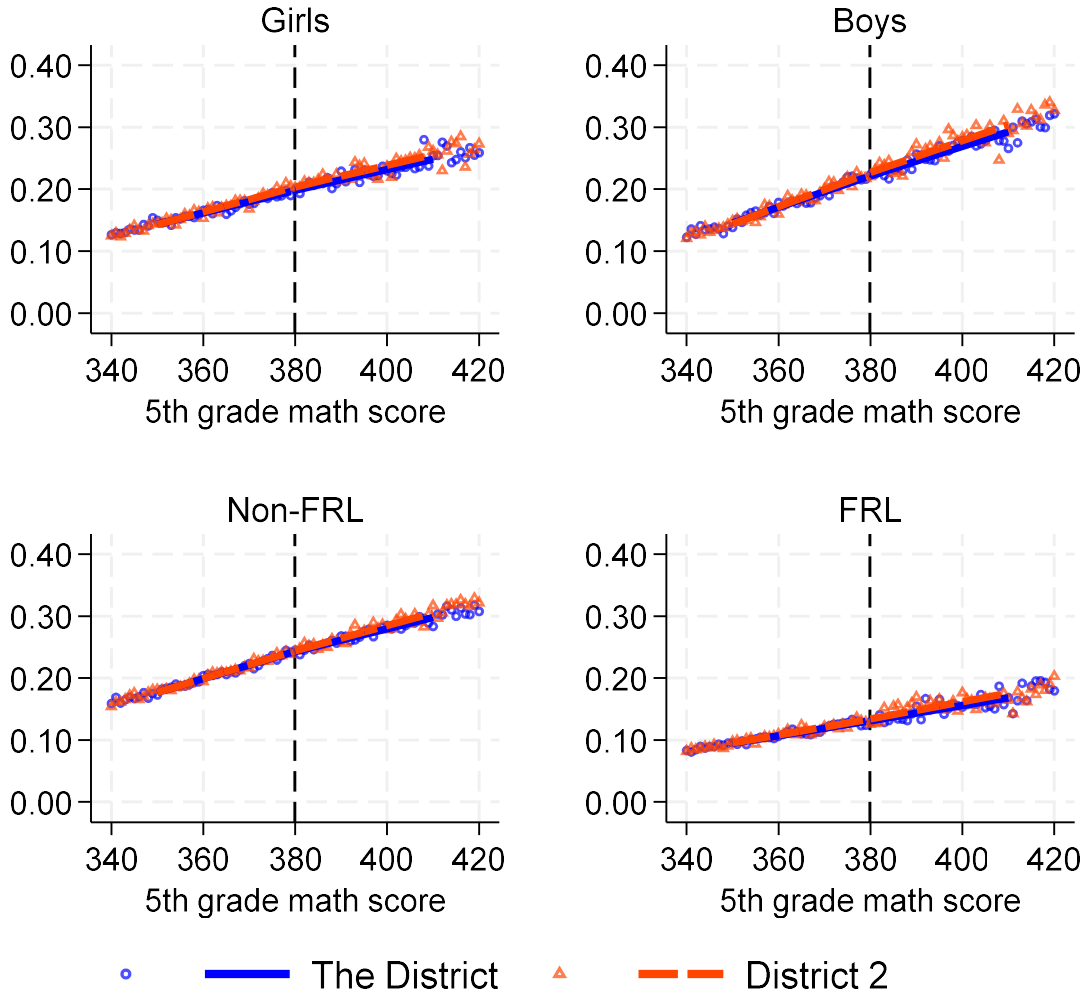
Notes: This figure plots sample means of an indicator for participating in the GEM program in sixth grade (the treatment variable, in blue circles) and an indicator for participating in either GEM or advanced-track math (in green triangles) at each value of fifth-grade math score (the running variable), along with fitted values from local linear RD models with a symmetric bandwidth of 30.

Figure A5. Reduced-Form Relationships by Gender and Probability of Successfully Completing GEM



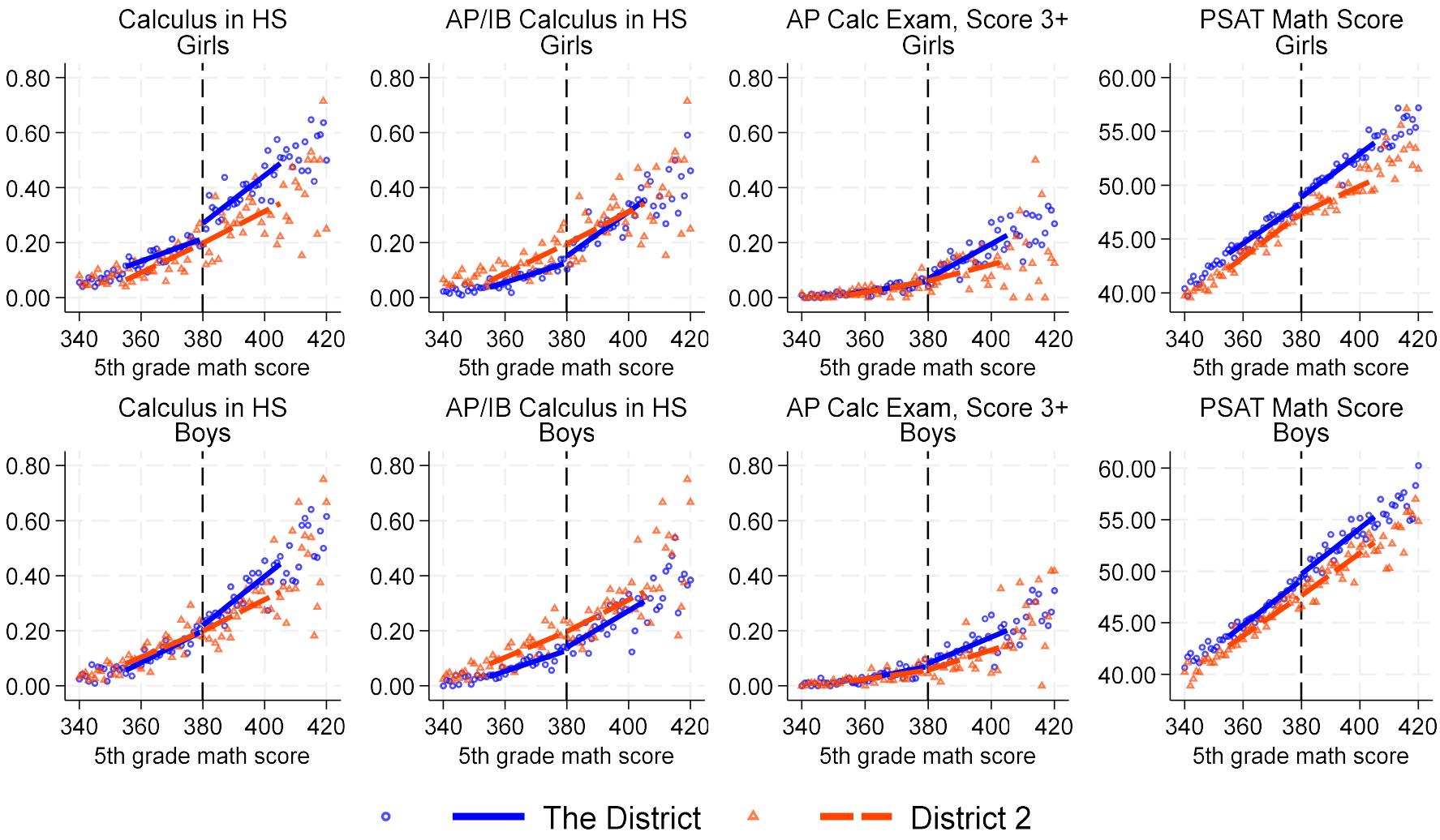
Notes: This figure plots sample means and fitted values from local linear RD models with a symmetric bandwidth of 30 points at each value of fifth-grade math score (the running variable). Panels A and B, the dependent variable is an indicator for participating in the GEM program in sixth grade (the treatment variable); in panels C and D it is an indicator for earning a bachelors degree in STEM or a Business-related field within 6-years of on-time graduation. In each panel, means and estimated models are shown separately for girls or boys whose predicted probability of successfully completing the GEM track – defined as completing geometry in 8th grade with a passing grade – is above or below the gender-specific median. See text (section IV.e) for details on the model used to estimate predicted success.

Figure A6. Predicted On-Time Graduation and Bachelor's Degree within 6 Years, Between-District Comparisons by Gender and FRL Status



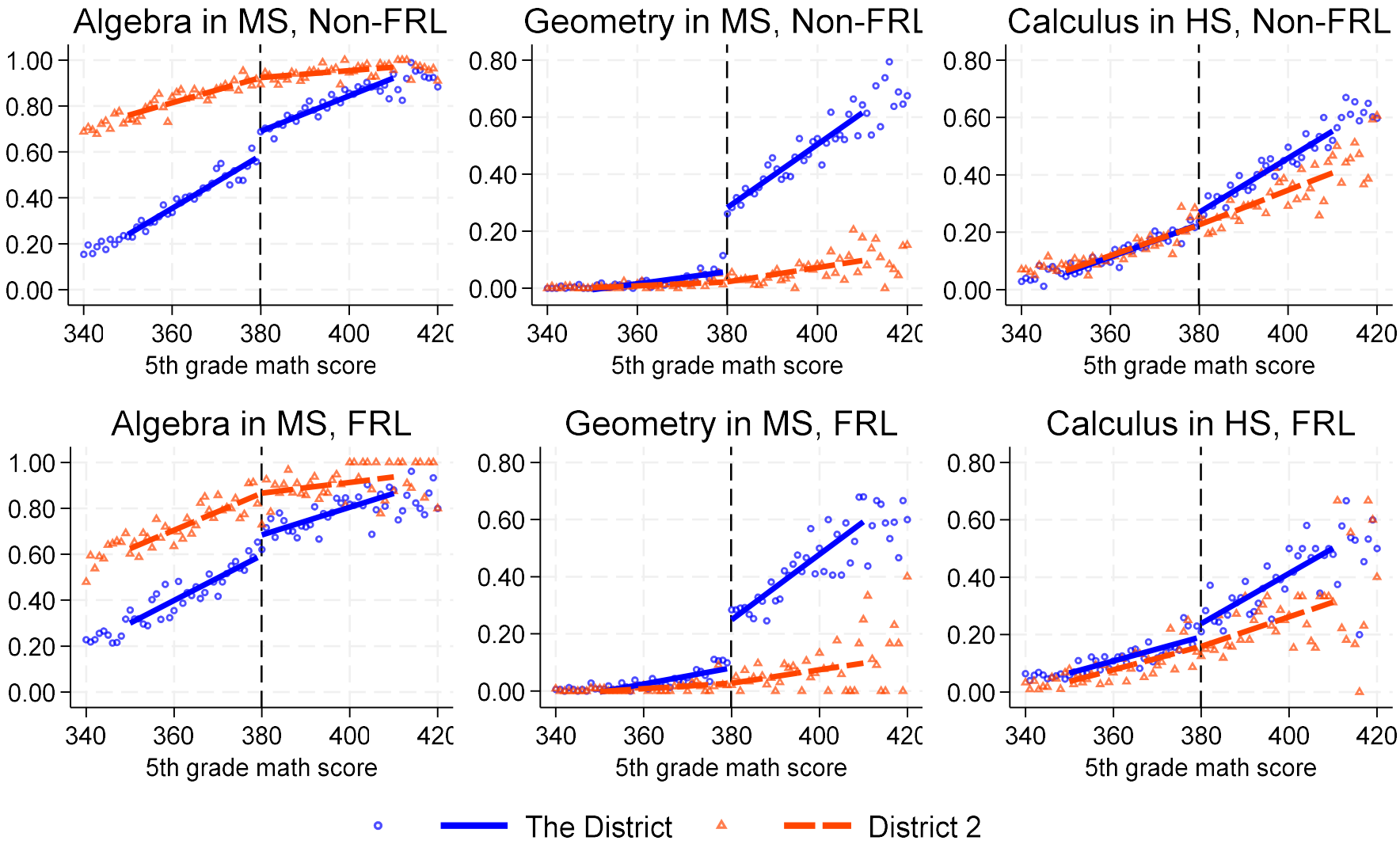
Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380. The dependent variable is an index of observable pre-treatment characteristics constructed as the predicted probability of graduating high school on time and completing a bachelor's degree in STEM or Business/Economics within 6 years. The prediction models are fit to the District RD sample using observations with 5th-grade math scores below the threshold and include: second-order polynomials in 4th-grade math test scores, 4th-grade reading test scores, and student age; indicators for student gender, five race/ethnicity categories, FRL status, English Language Learner status, and cohort; and the interactions of gender with 4th-grade test scores and FRL status.

Figure A7. More High-School Outcomes, Between-District Comparison by Gender



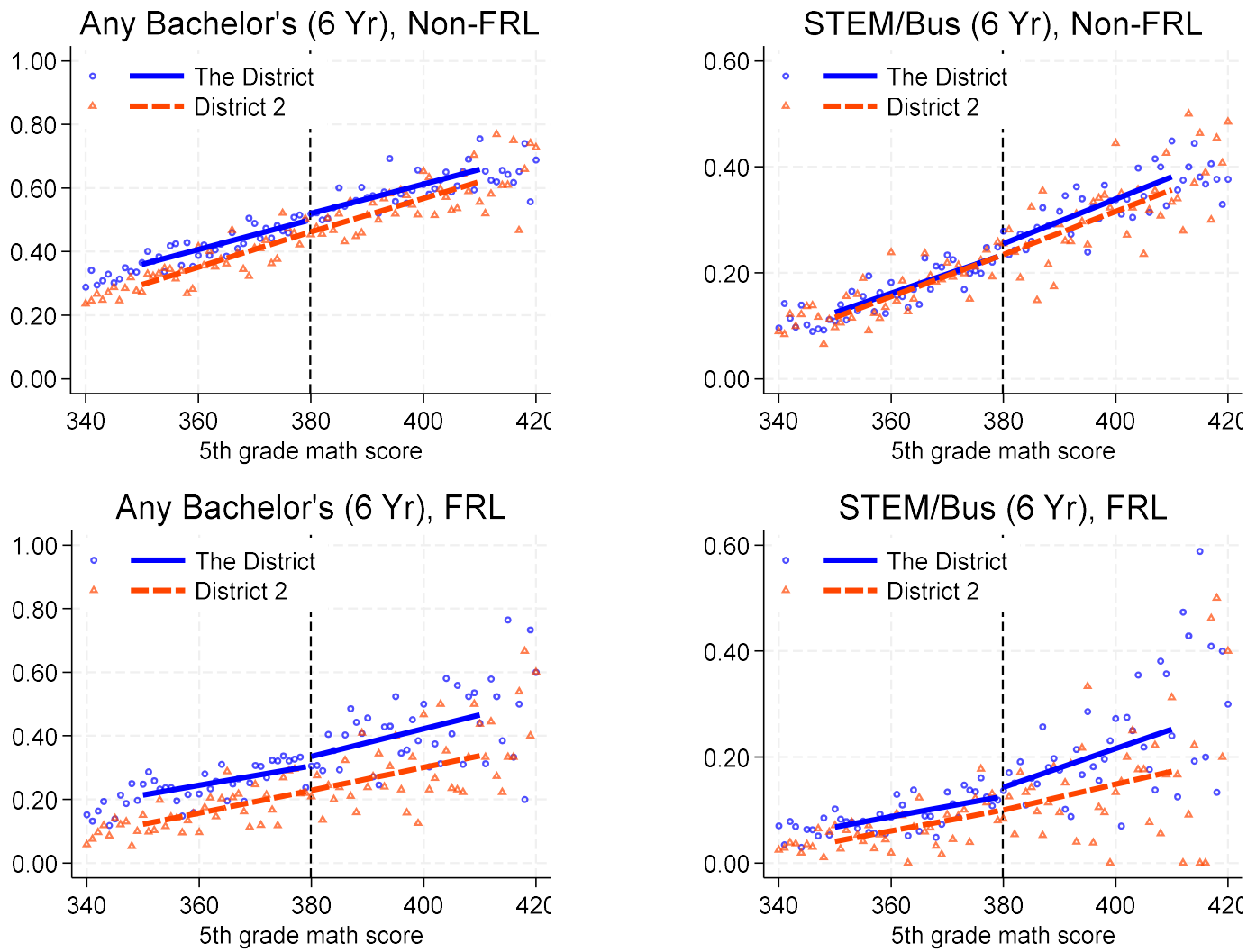
Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure A8. Math Course Enrollments, Between-District Comparison by FRL Status



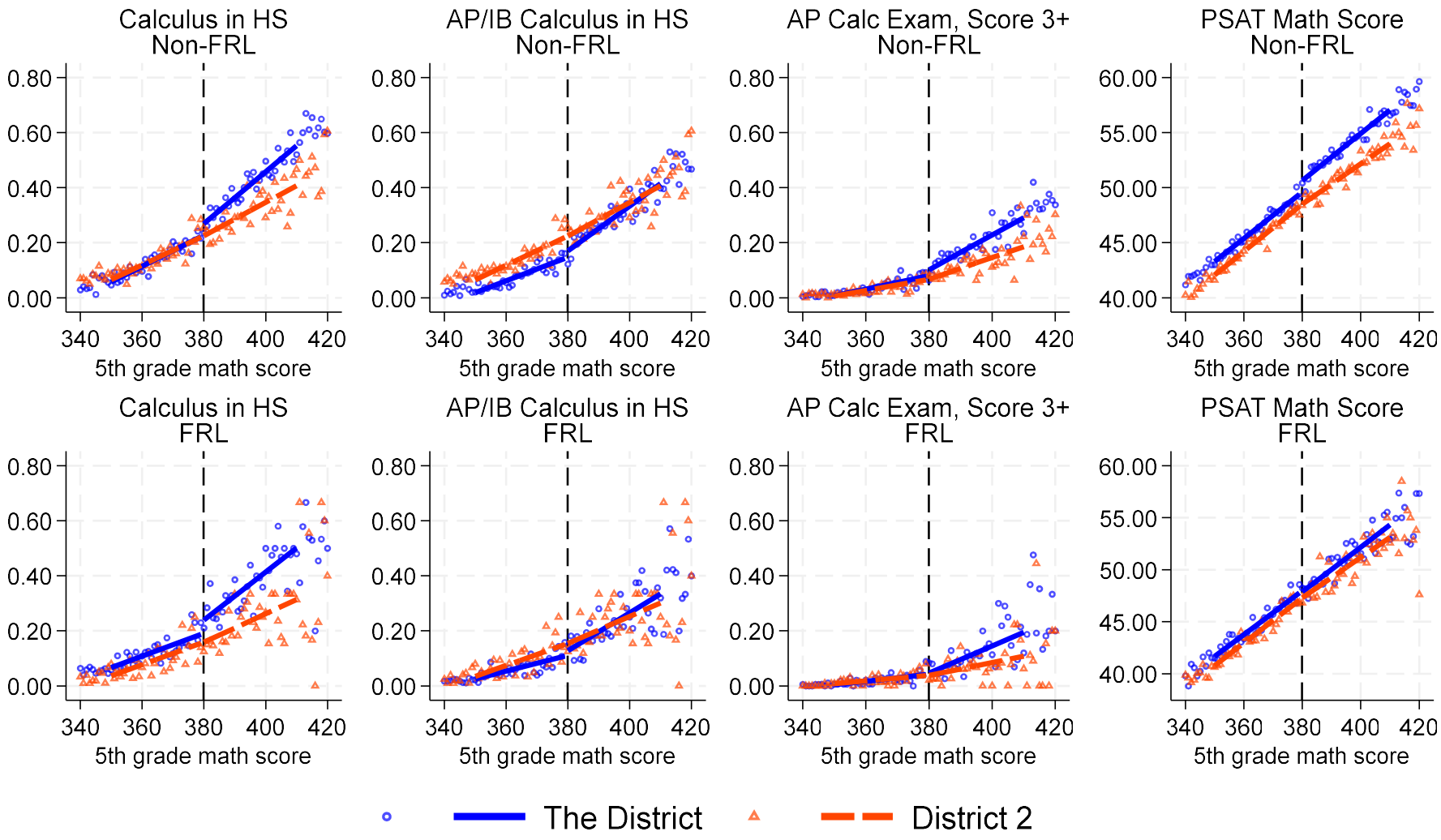
Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure A9. College Outcomes, Between-District Comparisons by FRL Status



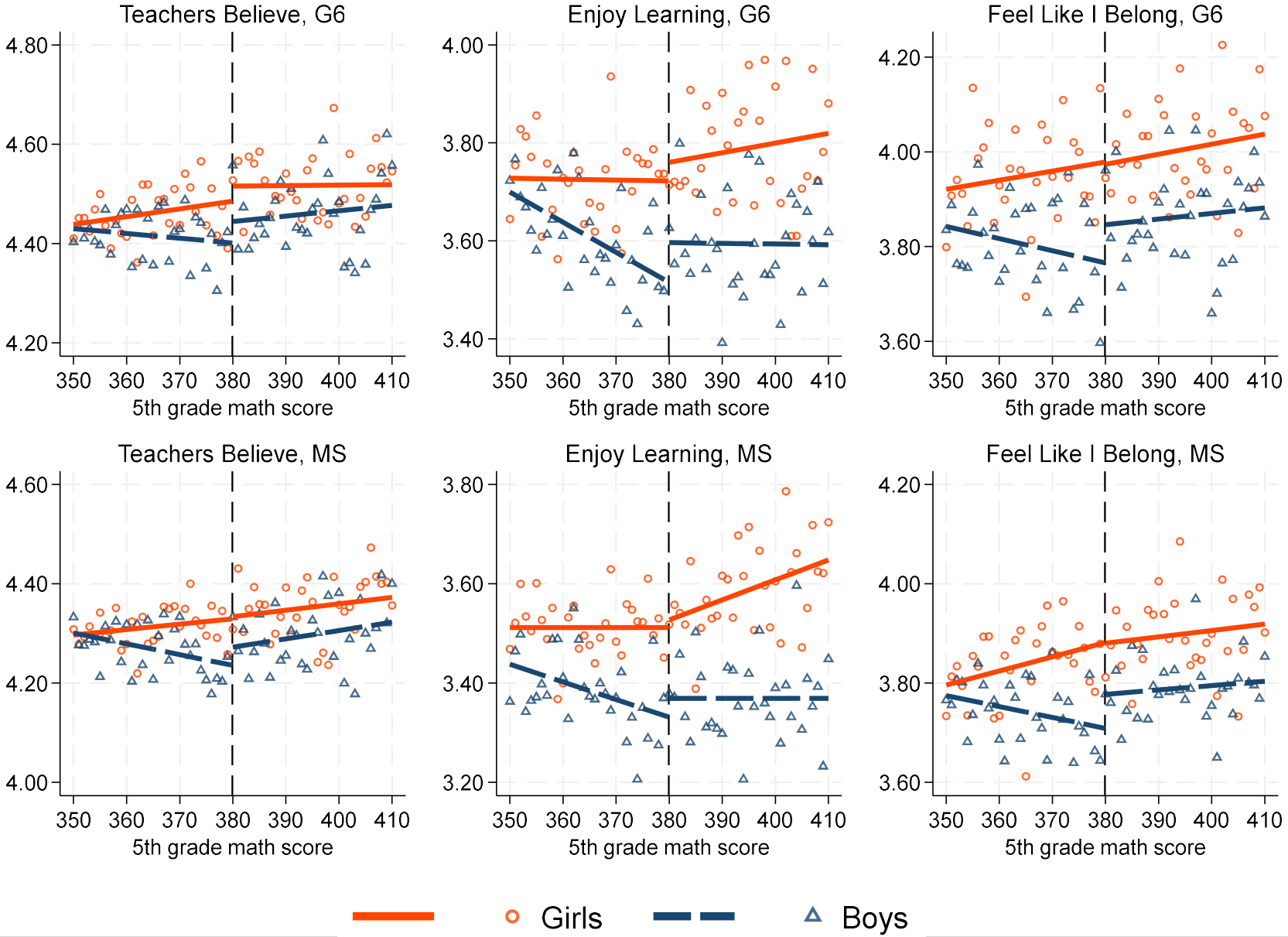
Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure A10. More High-School Outcomes, Between-District Comparison by FRL Status



Notes: This figure plots sample means of each variable at each value of the 5th-grade math score, separately for the District and a comparison district (District 2) that did not offer an accelerated (GEM-like) track for high performers but enrolled a larger share of lower performers in algebra in 8th grade. The solid blue lines represent fitted values from local linear RD models with a symmetric bandwidth of 30 fit to District data, while the orange dashed lines are fitted values from models fit to District 2 data. The District 2 sample includes all students in that district who meet the criteria used to select our main RD sample (or the District 2 analogues). The District 2 models use a 30-point symmetric bandwidth and allow for different slopes above and below 380.

Figure A11. Reduced-Form Relationships for Survey-Based Outcomes in Middle School, by Gender



Notes: Dependent variables are student response to survey questions asking them whether their "teachers believe [they] can succeed," whether they "enjoy learning" at their school, and whether they "feel like they belong" at their school. Responses to both questions are measured using a 5-point Likert scale where 1 represents "strongly disagree" and 5 represents "strongly agree." The top row shows mean responses in the spring of 6th grade (near the end of the first year of potential GEM participation). The bottom row shows mean responses from grades 6, 7, and 8.

Appendix Table A1: Estimated female earnings gap with and without controls for field of degree

	Estimated female effect:	
	No controls for age (1)	Including age controls (2)
1. No other controls	-0.328	-0.328
	(0.003)	(0.003)
2. Controls for education (2 dummies) and class of main job (7 dummies)	-0.312	-0.312
	(0.003)	(0.003)
3. Controls for education, job class and STEM/business or economics (2 dummies)	-0.269	-0.270
	(0.003)	(0.003)
4. Controls for education, job class and field of degree (172 dummies)	-0.247	-0.247
	(0.003)	(0.003)
Share of gender effect in row 2 attributable to:		
- STEM/business	13.8%	13.5%
- detailed field of degree	20.8%	20.8%

Notes: table entries are estimated coefficients of indicator for female in models for log annual wage and salary earnings. Sample includes native born people age 30-35 in 2010-2018 American Community Survey (ACS) with positive wage/salary earnings and BA degree or higher (n=556,741). Models in row 1 include no other controls (except age and age*female in column 2). Models in row 2 add controls for education (2 dummies) and class of main job (e.g., private, state and local goverment, etc.). Models in row 3 add 2 dummies for STEM fields and for fields in business or economics. Models in row 4 add dummies for the full set of 4-digit field of study variables in ACS. All models in column 2 also include a linear age term and female interacted with (age-32.5), so the coefficient for the main effect of female gender represents the wage gap at age 32.5 (the mid-point of the age range in the sample).

Appendix Table A2. Sample Criteria & Sample Means

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Main RD Analysis Sample				
	All in Grade 5	Observed in Grades 4-6	Followed Through HS Grad	Pooled	Girls	Boys	Non-FRL	FRL
Demographic characteristics								
female	0.49	0.49	0.51	0.49	1.00	0.00	0.49	0.51
FRL	0.43	0.43	0.40	0.30	0.31	0.29	0.00	1.00
White	0.35	0.35	0.35	0.44	0.43	0.45	0.56	0.17
Black	0.34	0.34	0.34	0.23	0.25	0.21	0.13	0.45
Hispanic	0.24	0.24	0.24	0.25	0.24	0.26	0.22	0.31
Gifted	0.06	0.06	0.07	0.10	0.10	0.11	0.12	0.08
Baseline achievement								
grade 4 math score (standardized)	0.08	0.08	0.17	0.68	0.64	0.71	0.73	0.54
grade 4 reading score (standardized)	0.08	0.08	0.16	0.57	0.65	0.50	0.65	0.39
School peer characteristics (6th grade)								
avg grade 4 math & reading scores	0.07	0.07	0.09	0.14	0.14	0.14	0.21	-0.02
fraction FRL	0.46	0.46	0.44	0.40	0.41	0.40	0.33	0.56
fraction GEM eligible	0.20	0.20	0.20	0.22	0.22	0.22	0.24	0.17
Math track in 6th grade								
GEM (treatment)	0.16	0.16	0.18	0.29	0.29	0.30	0.31	0.26
advanced math	0.27	0.27	0.29	0.49	0.51	0.47	0.49	0.49
regular math	0.57	0.57	0.53	0.22	0.20	0.23	0.20	0.25
Number of observations	98,709	79,494	55,402	23,225	11,488	11,737	16,235	6,990

Notes: The sample in column 1 includes five cohorts of students who entered 5th grade in 2003-2007. Column 2 is restricted to students who were observed in 4th grade (and have 4th-grade test scores) and were also observed in 6th grade the following year (2004-2008). Column 3 is limited to students who were enrolled in (or graduated from) a District high school six years later (between 2010 and 2014). Finally, columns 4-8 report means for the analysis samples for the RD models, which include students who meet the criteria in column 5 and also had 5th-grade math score within 30 points of the GEM eligiblity threshold (our preferred RD bandwidth).

Appendix Table A3. More High School Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Took Any Advanced Math Course	**Took AP/IB Calculus Course**	**Took Calc AP Exam + Scored >= 3**	**High School GPA (Overall)**	**PSAT Math**	**PSAT Verbal**	**Took SAT/ACT**
1. Pooled	**0.067****	**0.051***	**0.030***	**0.002**	**2.650****	**2.008**	**0.009**
(N = 23,225)	(0.021)	(0.019)	(0.014)	(0.022)	(0.983)	(1.243)	(0.018)
2. Girls	**0.058+**	**0.052+**	**0.027**	**0.044**	**2.443+**	**2.846+**	**0.008**
(N =11,488)	(0.035)	(0.028)	(0.022)	(0.027)	(1.230)	(1.597)	(0.018)
3. Boys	**0.071****	**0.050***	**0.032+**	**-0.047**	**2.838***	**1.117**	**0.006**
(N =11,737)	(0.024)	(0.021)	(0.018)	(0.043)	(1.324)	(1.749)	(0.028)
4. Non-FRL	**0.037**	**0.049***	**0.034***	**-0.009**	**3.386****	**2.382+**	**0.000**
(N = 16,235)	(0.023)	(0.023)	(0.016)	(0.025)	(1.161)	(1.307)	(0.017)
5. FRL	**0.137****	**0.054+**	**0.019**	**0.019**	**0.342**	**0.583**	**0.025**
(N = 6,990)	(0.039)	(0.030)	(0.028)	(0.045)	(1.510)	(2.474)	(0.044)

Notes: Table entries are 2SLS estimates for the effect of entry into GEM in 6th grade (i.e., they are estimates of β in equation 3). The models include four dummy variables for the student's cohort but no other controls. Standard errors clustered on the middle school where the student was enrolled in 6th grade.

Appendix Table A5: Estimated Model of Selection into GEM and Probability of Completing GEM by Passing Geometry

	Girls				Boys			
	Selected into GEM		Passed Geometry		Selected into GEM		Passed Geometry	
	Coeff	Std Err	Coeff	Std Err	Coeff	Std Err	Coeff	Std Err
Gr 4 math test score (σ's)	0.252	0.078	0.660	0.119	0.200	0.080	0.521	0.115
Gr 4 reading test score (σ's)	0.205	0.058	0.118	0.086	0.162	0.047	0.136	0.071
Square of Gr 4 math test score (σ's)	-0.022	0.032	-0.128	0.043	-0.009	0.031	-0.069	0.039
Square of Gr 4 reading test score (σ's)	-0.090	0.025	-0.005	0.035	-0.051	0.024	0.040	0.035
Age in Gr 5	-0.457	1.721	3.581	2.514	-1.491	1.084	-2.777	1.600
Age in Gr 5 squared	0.025	0.082	-0.174	0.119	0.067	0.051	0.124	0.075
Gifted	0.231	0.051	0.067	0.062	0.396	0.050	0.130	0.056
FRL	0.092	0.041	-0.199	0.056	0.135	0.041	-0.156	0.054
White	-0.249	0.093	-0.177	0.124	-0.240	0.091	-0.223	0.112
Hispanic	-0.099	0.097	-0.209	0.129	-0.197	0.095	-0.156	0.117
Black	0.232	0.097	0.171	0.130	0.054	0.097	-0.047	0.120
Asian	0.097	0.114	0.371	0.151	0.123	0.113	0.380	0.137
English Language Learner	0.078	0.112	0.321	0.158	-0.028	0.098	0.295	0.141
Gr 5 math test score	0.035	0.003	0.013	0.010	0.037	0.004	0.029	0.010
Gr 5 reading test score	0.003	0.001	0.002	0.001	0.004	0.001	0.003	0.001
Gr 5 math score ´ 1[Gr 5 math score≥380]	-0.018	0.004	0.002	0.009	-0.021	0.004	-0.007	0.010
1[Gr 5 math score≥380]	1.513	0.061	--	--	1.465	0.061	--	--
Cohort dummies	yes		yes		yes		yes	
Correlation between selection into GEM and passed geometry if in GEM	0.223	0.107			0.471	0.094		

Note: table presents estimated coefficients from bivariate probit model of selection into GEM and probability of passing geometry if admitted to GEM. Model is fit separately to data for girls and boys, using main RD analysis sample. See text (section IV.e) for details.