ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-265

Sibling Effects on High School Exam Taking and Performance

Oded Gurantz
University of Missouri

Michael Hurwitz
College Board

Jonathan Smith
Georgia State University

Younger siblings take more advanced high school course end of year exams when their older siblings perform better in those same exams. Using a regression discontinuity and data from millions of siblings who take Advanced Placement (AP) exams, we show that younger siblings with older siblings who marginally "pass" an AP exam are more likely to take at least one AP exam, increase the total number of AP exams, and are more likely to take the same exam as their sibling. The largest impacts are found among sisters, but we do not see differential effects in coursework where females are underrepresented.

VERSION: July 2020

Suggested citation: Gurantz, Oded, Michael Hurwitz, and Jonathan Smith. (2020). Sibling Effects on High School Exam Taking and Performance. (EdWorkingPaper: 20-265). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/m1m3-8g75

Sibling Effects on High School Exam Taking and Performance

Oded Gurantz, University of Missouri

Michael Hurwitz, College Board

Jonathan Smith, Georgia State University

July 2020

Abstract: Younger siblings take more advanced high school course end of year exams when their older siblings perform better in those same exams. Using a regression discontinuity and data from millions of siblings who take Advanced Placement (AP) exams, we show that younger siblings with older siblings who marginally "pass" an AP exam are more likely to take at least one AP exam, increase the total number of AP exams, and are more likely to take the same exam as their sibling. The largest impacts are found among sisters, but we do not see differential effects in coursework where females are underrepresented.

Acknowledgements: The results do not reflect the views of the College Board. Thanks to participants at the Association for Education, Finance, and Policy, Stanford's Center for Education Policy Analysis, and the University of Bristol Workshop on Assessment and Feedback for questions and comments.

1. Introduction

High school courses and curriculum have long-lasting impacts on outcomes such as high school graduation, college major, and earnings (Altonji, Blom, & Meghir, 2012; Goodman, forthcoming; Joensen & Nielsen, 2009, 2018; Rose & Betts, 2004). Disparities in coursework appear as early as primary school and continue through college, typically falling on gender, racial, or income lines (Card & Giuliano, 2016a, 2016b; Conger, Long, & Iatarola, 2009). Given the disparities in high school coursework it may prove useful to understand how students select (or are selected) into particular courses, as simply expanding course taking opportunities does not necessarily lead to meaningful changes in participation rates, reductions in disparities, or improvements in longer-term outcomes (Darolia, Koedel, Main, Ndashimye, & Yan, 2018; Grissom & Redding, 2016).

This paper estimates the causal impact of older siblings' Advanced Placement (AP) exam performance on younger sibling AP course taking using over 8 million sibling-by-AP exam observations in the high school graduating cohorts of 2004 to 2014. High school students can choose from over 30 different AP courses that offer advanced, college-level material, which culminate in an end-of-year subject-specific exam. Students who perform well on an AP exam can earn college credit, which can increase the likelihood students take more advanced college courses, decreases time to degree, and impacts the choice of college major (Avery, Gurantz, Hurwitz, & Smith, 2018; Gurantz, forthcoming; Smith, Hurwitz, & Avery, 2017).

There are several reasons to believe that older siblings may influence younger sibling course participation. First, siblings are an extreme version of peers and we know that peers influence schooling outcomes in a variety of contexts, including the whether to attend school and which schools to attend (Bennett & Bergman, 2019; Bursztyn & Jensen, 2015; Dobbie & Fryer, 2014). Second, a growing body of work focuses on the importance of peers that "look like me," referring to the impact of observing a same group peer in contexts where that group is underrepresented (Dee, 2005; Ellis & Gershenson, 2016; Gershenson, Hart, Hyman, Lindsay, & Papageorge, 2018). Women may be more likely to perform better or enter competitive environments when exposed to female role models (Buser, Niederle, & Oosterbeek, 2014; Scott E. Carrell, Page, & West, 2010; Kofoed & mcGovney, 2019; Lim & Meer, 2017; Niederle & Vesterlund, 2007). Recent research finds that women take more college economics courses when introduced to a successful women economics major during their introductory economics courses (Porter & Serra, 2019). Finally, family members, including both parents and children, learn from one another and influence one another's choices and decisions (Aguirre & Matta, 2018; Altmejd et al., 2020; Black, Breining, et al., 2017; Black, Grönqvist, & Öckert, 2017; Heissel, forthcoming; Qureshi, 2018).

To overcome the typical peer and sibling reflection problem, we exploit the fact that students who take AP exams only observe their integer score between 1 and 5 but we, the researchers, observe each exam's underlying continuous score. Using a regression discontinuity design, we compare the AP exam choices of two nearly identical younger siblings who only differ in that one older sibling's continuous AP exam score just barely earned a higher integer score than the other older sibling. This comparison of two younger siblings who are on average identical but for the integer score of their older siblings is in the spirit of random assignment, which is difficult to achieve with peers and nearly impossible to achieve with siblings.[1]

Our primary finding is that older siblings who just barely earned a higher integer score on an AP exam have substantial impacts on the AP participation of their immediate younger sibling.

[1] Research has leveraged adoption to produce estimates of random assignment of children to parents (Lindquist, Sol, & Praag, 2015; Sacerdote, 2007).

At the 2/3 threshold, commonly thought of as passing an AP exam, younger siblings are 2.2 percentage points (6.8 percent) more likely to take the exact same AP as their older sibling. Older siblings have impacts on both the intensive and extensive margin of AP participation, as younger siblings are 1.1 percentage points (1.6 percent) more likely to attempt at least one AP exam, and among AP exam takers, participate in 0.05 (1.3 percent) more exams overall. At higher thresholds we also find positive impacts on taking the same AP exam, though these effects are largely younger students substituting towards (or away from) exams in which their older sibling performed better (or worse). We also find strong evidence that these induced students perform well in the exams, as those who take the same AP exam as their older sibling pass this course approximately 80 percent of the time. Results also point to no negative spillover effects that cause a decline in overall performance.

We find the largest peer effects among sisters, though there are positive and statistically significant results within mixed-sibling pairs. Within sister pairs, the younger sister is 2.6 percentage points (6.3 percent) more likely to take the same exam when her older sister receives a higher integer score, compared to a marginally significant 0.9 percentage points (2.4 percent) for brothers. Most importantly, we do not find strong evidence that these impacts are mediated by the types of courses in which the sibling receives a higher score, such as areas where females are traditionally underrepresented. Younger sisters appear just as likely to respond to the higher score in classes where they are typically overrepresented relative to males (e.g., English Literature) as ones where they are underrepresented (e.g., Physics).

Several additional analyses allow us to investigate whether role models and parents contribute to the results. First, we find no meaningful differences by race/ethnicity or parental education, both of which are subgroups that are typically underrepresented in courses such as

STEM and economics. We also find no differential impact by age gaps, one measure of the "closeness" of a sibling. These subgroup results do not support the idea that role models play more of a role in any particular subgroup. Second, we find evidence consistent with these impacts being passed through siblings and not parents. Parent effects are typically difficult to rule out, since they play an important role in sibling dynamics (Becker & Tomes, 1976; Pitt, Rosenzweig, & Hassan, 1990; Yi, Heckman, Zhang, & Conti, 2015). Yet in families with more than three children, the oldest sibling's AP score has an effect on the middle sibling but not the youngest sibling. If parents were learning from the older sibling's performance we would expect that they translate this information to all younger siblings, which we do not observe.

This paper helps identify one key determinant of course taking: signals of sibling performance in advanced coursework. By focusing on the role of siblings in the choice of courses, we show the importance of social and family networks in the formation of skills through course choice.[2] Joensen and Nielsen (2018) show that when older siblings are exposed to a pilot program that offers an advanced math course, younger siblings are themselves more likely to take math-science courses. Typically, researchers have evaluated impacts through differences in "school curriculum", whether by policy reforms or idiosyncratic variation by geography and time. This is analogous to how presenting students with different choice sets effects decision-making, whereas this paper, to the best of our knowledge, is the first to measure how younger siblings course choice changes in response to the observed academic performance measures of two older siblings who have made identical educational investment choices. The better we understand student decision

[2] A related line of inquiry investigates ways in which older siblings impact younger siblings' academic performance rather than course choice. For example, Qureshi (2018) exploits random assignment of teacher quality to an older sibling, Karbownik and Ozek (2019) exploit the school starting age of the older sibling, and Nicoletti and Rabe (2018) consider how older sibling academic achievement influences younger sibling academic achievement, as measured by mandatory test scores.

making, the more opportunities exist to generate sound policy to achieve one's goals, such as reducing the gender or racial gaps in STEM and economics course work, college majors, and careers. At the broadest level, this paper shows the importance of accounting for spillover effects in educational contexts and how these relationships may also open up opportunities to change student decision making.

2. Background and Data

2.1. Advanced Placement Background

High school students can enroll in Advanced Placement (AP) courses, which offer college-level academic material across a variety of subject areas. There is no requirement that students in AP courses take the AP exam offered at the end of the school year, though just over 2.7 million high school students took almost 5 million AP exams across 38 different subject areas in 2017.[3] For context, there are approximately 15 million high school students across the U.S., just under 4 million per grade, and most AP exams are taken by 11th and 12th graders. Although students may take AP exams for many reasons, the primary justification is to earn college-credit that would allow the student to have one fewer requirement, sometimes in the number of courses and sometimes in the subject matter. Most students who earn credit choose to skip the course, with students passing AP STEM exams typically increasing the number of STEM courses they take as a result (Gurantz, forthcoming). AP credit also increases the likelihood a student earns their degree on-time (i.e., within four years) (Smith et al., 2017). Yet the decision to take AP exams includes tradeoffs. In one context, students randomly induced to take AP exams rarely passed the exams

[3] Exams take place over a two-week period in May with only one administration per subject per year, scores are released several months later, and only 0.3 percent of students ever retake an AP exam. Exam statistics from https://secure-media.collegeboard.org/digitalServices/pdf/research/2017/Program-Summary-Report-2017.pdf

but received lower course grades that were not necessarily adjusted for by college admissions offers (Conger, Kennedy, Long, & McGhee, 2019; Conger, Long, & McGhee, 2020).

Whether a student is offered the opportunity to skip a course varies by college. AP scores are reported to students and colleges on a 1 through 5 scale, where 1 translates into "no recommendation" and 5 translates into "extremely well-qualified". Each university independently determines what AP exam score they will accept in order to receive credit for the course, though a small fraction of colleges do not accept any AP scores or may require students within that major to retake that course. The most common benchmark to earn credit is earning a score of 3, which is often referred to as "passing" an AP exam; although many colleges require a 4 or 5, credit is almost never offered at lower levels.

The integer scores derive from a continuous, "raw" score based on performance on multiple choice and free-response sections. Because the AP exams are criterion-based, cut scores are established based on earning a pre-determined number of points that predict college-performance at varying levels and not on relative performance.[4] The exams are designed so students earning a score of 3 on one test administration should have an identical mastery of material as students earning a 3 on a separate administration.

2.2. College Board Data

We make use of student-level College Board data on all PSAT/NMSQT (PSAT), SAT, and AP exam takers from the 2004 through 2014 cohorts. This paper only considers the 17 most popular subject exams during the time period under study (see Appendix Table 1 for details on these exams). As previously described, AP exams are given integer scores between 1 and 5, but

[4] Continuous raw scores range from 0 to 180 points, though there is considerable variation in the scoring range and maximum across exams.

we also have access to the underlying continuous score that is mapped into the integer score. These scores are never reported publicly or to the student.[5] We can observe the raw continuous AP score only for the 2004 through 2009 cohorts, but integer data for 2004 through 2014; the cohorts with available raw scores were previously studied in Smith et al. (2017) and Avery et al. (2018). We identify the sharp boundaries which separate students into the AP integer categories, which are later used in our research design. We observe all exam scores on all attempts but cannot identify students who participated in AP coursework but did not take the exam.

For each student we have basic self-reported demographic data, such as gender, race/ethnicity, parental income, and parental education. We also observe the high school each student attends as well as their names and home addresses, which we use to identify siblings (described below). We also observe performance on both the PSAT and SAT exams. The PSAT is commonly considered a practice exam for the SAT and a college readiness assessment taken by approximately three million students every year, mostly high school sophomores and juniors.[6] The SAT is one of two leading college admissions exams that is relied upon, with varying weight, in the admissions process. It is commonly taken near the end of students' junior year in high school or in the fall of their senior year. Prior to 2006, the SAT consisted of a math and a critical reading section, each ranging from 200 to 800 points, in increments of 10. A writing section was added after 2005 with the same scoring such that the total score across the three sections ranges from 600 to 2400.

2.3. Analytic Dataset

[5] We are only able to access data on raw scores for exams taken during the 2003-04 through 2008-09 school years. Therefore some AP test takers, particularly in the 2004 and 2005 cohorts, will not have raw scores that can be mapped to their scaled scores taken in sophomore or junior year of high school. The few exams without an accompanying raw score are removed from our analyses.

[6] The PSAT consists of three multiple choice sections -- math, critical reading, and writing. Each section is scored between 20 and 80 for a total score range of 60 to 240. The PSAT is offered once a year in October and most frequently offered at students' high schools.

We begin with data on all AP exam takers from the 2004 through 2009 graduating cohorts. We then look for their younger siblings among the PSAT, SAT, and AP exam takers from the 2005 through 2014 cohorts. Similar to previous work using these same data, we identify siblings as those who share the student's last name and home address (Goodman, Hurwitz, Smith, & Fox, 2015). For home address, we use city, state, and the first five characters of the street, including street number. Our method of identifying siblings is unlikely to result in matches that are considered false positives, given the low likelihood that two students close in age to share a last name and home address. The method may, however, fail to match siblings if their families have changed home address between exam administrations, or if their last names are different due to being half-siblings or from other naming differences.

Our analysis primarily focuses on the impacts of the older sibling's performance on the younger sibling closest in age, unless otherwise noted. Approximately 83% of the sibling sample consists of families with exactly two siblings, and larger families almost always contain exactly three siblings. One reason for this is simply the time period of data that we have available, which might limit our ability to observe larger families with wide age gaps between siblings, as well as our matching process which limits false positives. Finally, we make two small refinements to the dataset. First, we remove all families with twins from the analysis because we are particularly interested in dynamic responses that require age gaps, and ultimately this constitutes a relatively small group that simplifies the analysis. Second, we only include siblings where both siblings have taken either the PSAT or SAT exam and the older sibling has taken an AP exam. This allows us to look at how younger siblings select into AP exams without potential endogenous selection that is caused by the AP process itself. In practice this last restriction removes relatively few students

– about five percent of the sample – as most AP takers have also taken at least one of these exams (and we shortly show that selection into the sample is unrelated to older sibling exam scores).

Table 1 provides descriptive statistics of the younger siblings, along with our analytical sample that includes all students whose older sibling was within five continuous score points of each AP integer threshold. AP exam takers predominately identify as white and Asian, with only 12% of the full sample considered African-American or Hispanic, and 75% of the sample describes their parents as having earned a bachelor's degree or higher. Representation of traditionally underrepresented ethnic groups declines from 19% at the 1/2 threshold to 6% at the 4/5 threshold. Gender representation is the one characteristic that appears well-balanced, at least on average, at all points of the AP distribution, at roughly 49% across all AP thresholds.

Table 1 also shows that the academic performance of the younger sibling is highly correlated with the AP performance of the older sibling. Younger sibling SAT scores rise from 1079 to 1234 for students whose older sibling was near the 1/2 and 4/5 thresholds, respectively. About 74% of the younger siblings in our sample take at least one AP exam, and the probability of taking AP and total number of AP exams are all increasing with the older sibling's integer scores. For example, 27% of younger siblings took an AP exam when their older sibling scored near the 1/2 threshold, compared to 43% when the older sibling scored near the 4/5 threshold; total AP exams taken rose from 2.6 to 4.1 between these two thresholds as well.

Our analytical sample of younger siblings weighs slightly more towards students who are traditionally overrepresented in AP exams, and higher education more broadly. For example, the universe of AP takers during this same time frame was 69% white, 16% African-American and Hispanic, 43% male, 36% reported family income over $100,000 (among those who reported), whereas the younger sibling sample in this analysis is 71% white, 12% African-American and

Hispanic, 49% male, and 43% in the high-income bracket among those reporting income (summary statistics for the universe of AP takers is available in other papers, such as Avery et al. (2018)). Similarly, if we examine all younger siblings in the College Board data on SAT test takers, irrespective of AP participation, they are 68% white, 18% African-American and Hispanic, 46% male, and 27% in the high-income bracket (see Goodman et al. (2015)). Nonetheless, most raw differences in demographic characteristics are relatively small, suggesting our sample is likely fairly representative of sibling behaviors broadly, or at least among those siblings who participate in college preparatory activities.

3. Methodology

We estimate causal impacts of older sibling AP performance on younger sibling's AP participation using a regression discontinuity design. We closely follow the empirical design in Smith et al. (2017), who use discontinuities in assignment to AP integer scores as an instrument. We diverge from previous work as we have information from two individuals in each observation; the older sibling provides the running variables (i.e., continuous AP score) that gives us our instrument and treatment variables, and the younger sibling provide the outcomes of interest.

More formally, each sibling pair is denoted by i and individuals within the sibling pair are distinguished by i_s where $s \in (y, o)$. The younger sibling in sibling pair i is denoted by i_y and the older sibling is denoted i_o. The older sibling takes AP exam j and receives continuous score $C_{i_o j}$. The four distinct integer thresholds (e.g., 1/2 or 4/5) are indexed by (n) and each exam has a distinct set of thresholds t_j^n. We generate the forcing variable as follows:

$$Dist_{i_o jn} = C_{i_o j} - t_j^n$$

which captures the distance for older sibling i's score on exam j from the threshold n. We then define four dichotomous variables if the older sibling is above/below each integer threshold:

$$Threshold_{i_o jn} = \begin{matrix} 1 \text{ if } Dist_{i_o jn} \geq 0 \\ 0 \text{ if } Dist_{i_o jn} < 0 \end{matrix}$$

After generating these variables, our basic empirical framework is shown by the regression discontinuity equation presented below:

$$Outcome_{i_y} = \alpha_0 + \alpha_1 \cdot Threshold_{i_o jn} + \alpha_2 \cdot Dist_{i_o jn} + \cdots$$

$$\ldots + \alpha_3 \cdot Threshold_{i_o jn} \times Dist_{i_o jn} + \Gamma X_i + N + J_t + T + \varepsilon_{i_y}$$

where $Outcome_{i_y}$ is the younger sibling outcome, such as number of APs taken. X_i is a vector of controls for the sibling pair i and can include characteristics of the younger sibling, older sibling, and their relationship, such as years in between them. Our primary analyses exclude X_i since it is not central to the research design.

There are two complications in our research design worth noting. First, each sibling could be assigned to multiple thresholds, as someone who receives an integer score of 1 is still theoretically below the 4/5 threshold, albeit quite far away. To simplify we include all students when focused on just one AP threshold, but when running pooled results we take the straightforward approach of assigning each student to the closest threshold based on the continuous AP score metric (i.e., if someone scores a 3 and is 5 points from the 2/3 threshold and 3 points to the 3/4 threshold, they are included in the 3/4 estimation). Given the relatively large spacing between thresholds relative to the bandwidths used, neither approach complicates the analysis. As such, we control for the closest threshold we are using for the older sibling, denoted by the integer threshold fixed effects *N*, when estimated pooled results. Specifications that only use a subsample of data around one threshold exclude this variable. We also include AP subject interacted with exam year fixed effects J_t (e.g., separate fixed effects for AP Biology offered in 2004 through 2009) and cohort fixed effects *T*, which represents the older sibling's graduation cohort to account for older siblings taking the exam in different years (e.g., 11^{th} vs 12^{th} grade).

The second potential concern is that the "sibling-by-exam" construction of our dataset (i.e., an older sibling taking 3 AP exams results in 3 separate younger sibling observations) overweights individuals who took multiple AP exams and violates independence of observations assumptions. Appendix Table 6 deals with this issue a number of ways, through subsample restrictions and weighting schemes, and is discussed later.

Our primary interest is in α_1, which is interpreted as the causal effect of an older sibling receiving a relatively higher integer score on the younger sibling's outcome. Our primary results rely on a linear regression using a rectangular kernel with a bandwidth of 5 points, with standard errors clustered at the level of the older sibling. Appendix tables report results using triangular kernels, alternative bandwidths, functional forms, and covariate. Alternate standard error adjustments, including robust standard errors and clustering by the running variable, produce similar results and are omitted for brevity, though appendix tables showing optimal bandwidth results also provide robust bias-corrected estimates (Calonico, Cattaneo, & Titiunik, 2014). The one exception to our robust estimates comes from "honest" confidence intervals as suggested by Kolesár and Rothe (2018). Here the results are highly sensitive to the choice of the parameter K (a smoothing parameter which is a bound on the second derivative imposed by the researcher); when applying their recommended heuristic we generally get results that reject the null of no effect, but small changes to this parameter lead to larger confidence intervals that eliminate statistical significance. We later also examine impacts on the non-closest sibling (e.g., first sibling paired with third).

3.1. Validity of Research Design

As in any regression discontinuity design, we must make sure there is no manipulation around the discontinuity. In this context, that means we should not see older siblings (or the graders of the exam) manipulating their continuous scores that eventually map into the well-known integer scores. Students are unable to directly manipulate their position as scoring is a complicated formula that applies weights unknown to students to different questions and both weights and questions vary from year to year, along with the threshold locations. Thresholds are also pre-determined by psychometricians prior to a separate set of graders who score the exam. As such, it is unlikely we see manipulation or sorting on the behalf of the graders.

We formally test manipulation with the standard covariate balancing and density tests. In this context, we also have a sample selection concern that implicates the validity of our design. We are concerned that if AP scores have a causal impact on younger sibling's AP participation, then the younger siblings may only appear endogenously in our data if the older sibling obtains above a certain threshold on an AP exam. As we have a larger dataset of PSAT and SAT takers to draw from, we can formally assess the issue.

Starting with all AP exam takers in 2004 to 2009 and the universe of PSAT, SAT, and AP takers in 2005 to 2014, we test (and reject) that an AP exam taker is discontinuously more or less likely to be matched to a younger sibling after attaining a higher integer score. Appendix Table 2 shows balancing tests by estimating equation (1), using various younger sibling related covariates as outcome measures, including having a younger sibling, total number of siblings, and indicators for various younger sibling family sizes. Results are presented separately for each of the four integer thresholds and all thresholds stacked. All estimates are null, with point estimates on dummy variables generally around 0.1 percentage points. To put this in perspective, estimates on whether the older sibling is more or less likely to have a younger sibling by crossing the AP threshold range from -0.3% to 0.5%, with almost every estimate in the table indicating less than a one percent change. Appendix Table 3 provides covariate balance tests for just the younger sibling sample, and again finds no significant differences coincident with the threshold. Appendix Figure 1 shows density of AP scores, binned by one-tenth of the continuous AP score, with no evidence of discrete bunching at the threshold. Density tests calculated using rddensity in Stata 16.1 fail to reject smoothness, with estimated p-values at the four AP thresholds as 0.30, 0.23, 0.99, 0.36. We take this as strong evidence that our identification strategy is not threatened by sample selection.[7]

4. Main Results

Table 2 shows that having an older sibling receive a higher AP integer score has significant impacts on younger sibling's participation in AP exams. The first row provides results from a stacked model that uses all four integer thresholds. We find that younger siblings are 1.4 percentage points (3.8 percent) more likely to imitate their older sibling by taking the same AP exam when their older sibling earns a higher AP integer score. Younger siblings also take 0.029

[7] Before restricting to the younger sibling subsample, we begin with the full sample of roughly 8 million AP exam takers from the high school cohorts of 2004 through 2009. Balancing tests, along with density figures that show no manipulation near the thresholds, are previously reported in Avery et al. (2018) and not reproduced here.

more AP exams overall (0.9 percent). In the aggregate this is not driven by students being more likely to take at least one AP exam (i.e., the extensive margin), with a statistically insignificant 0.27 percentage point effect. The increase primarily occurs through the intensive margin, where younger siblings take approximately 0.027 more AP exams (conditional on taking at least one exam).

Disaggregating by each of the four AP integer thresholds, Table 2 highlights that the largest changes are found when older siblings marginally improve from a score of 2 to 3, the threshold commonly considered a passing exam score. Graphical representation of these results is found in Figure 1. Younger siblings whose older siblings just barely score a 3 over a 2 are 2.2 percentage points (6.8 percent) more likely to take the same AP exam as their older sibling, and take 0.083 more exams overall (2.9 percent).[8] This comes from increases at both the extensive and intensive margins of AP participation. Younger siblings are 1.1 percentage points more likely to take at least one AP exam (1.6 percent) and, among those taking AP, increase their total participation by 0.052 aggregate exams (1.3 percent).

At higher integer thresholds we find evidence that younger siblings imitate their older sibling by shifting into the same AP exams, but no meaningful changes in overall AP participation. Students whose older sibling scores above the 3/4 and 4/5 thresholds are 1.5 and 0.9 percentage points more likely to take the same AP exam as their older sibling, respectively. These impacts are smaller in absolute magnitude and percentage terms than impacts at the 2/3 threshold. These positive impacts appear due in large part to substitution effects as there are no significant impacts on overall AP taking at either of these thresholds. In the aggregate, younger siblings at higher

[8] If we focus on just the increase in AP exams taken without the matched AP exam at the 2/3 threshold, the treatment effect is 0.065 additional AP exams with a standard error of 0.020, showing that there are spillover effects on AP taking generally. Results continue to be close to zero and statistically insignificant at the other thresholds.

thresholds are more likely to imitate their older siblings by shifting the portfolio of AP exams, rather than being inspired to attempt additional AP exams.

As might be expected, these causal estimates are smaller than what would be observed in naïve regressions that simply examine correlations after controlling for academic and background characteristics. Although our causal estimate suggests earning an AP score of 3 causally increase the likelihood the younger sibling takes the same exam by 2.2 percentage points, a simple multivariate regression analysis suggests an older sibling score of 3 increases matched exam taking by 3.6 percentage points. We find a similar 40-50% overstatement of the causal relationship in the naïve regressions at the higher thresholds - the matched exam increase is causally estimated as 1.6 and 0.9 percentage points at the higher 3/4 and 4/5 thresholds, the naïve estimates are similarly large at 3.1 and 2.7 percentage points in the regression analysis. Comparing the relationship between older sibling integer scores and younger sibling's total AP exams taken, the naïve estimates from shifting from 2 to 3, 3 to 4, and 4 to 5 on the integer scale are 0.04, 0.09, and 0.13 AP exams, relative to causal estimates of 0.09, 0.00, an -0.01 (regressions not shown).

4.1. Robustness Tests

We follow the literature and test the robustness of our results by varying regression bandwidth, utilizing quadratic rather than linear slopes, alternating between rectangular and triangular kernels, and adding student-level covariates. Appendix Table 4 shows that our results are consistent across these specifications and are not driven by any arbitrary functional form decisions. This table only shows results for the 2/3 threshold, where the magnitudes are largest, but other thresholds yield similar results. Appendix Table 5 re-estimates Table 2 using optimal bandwidth techniques as performed by the CCT methodology using rdrobust in Stata 16.1, resulting in larger bandwidths and similar results.

Another potential concern is that the "sibling-by-exam" construction of our dataset (i.e., an older sibling taking 3 AP exams results in 3 separate observations) overweights individuals who took multiple AP exams and violates independence of observations assumptions. Appendix Table 6 deals with this issue three ways: the first row uses only older siblings who took exactly one AP exam; the second row applies weights of $1/N$ to each individual, where N is the number of AP exams taken by the older sibling, in order to downweight individuals contributing multiple exams to the estimate; and the third row randomly selects only one of the older sibling's AP exams to contribute to the analysis. Results are insensitive to these robustness checks. The final row applies one more test, where we remove the three earliest older sibling cohorts (2004 to 2006), in case we have falsely labeled certain students as "older siblings" when they are more likely to be themselves younger siblings. Estimates are again unchanged.

4.2. AP Exam Performance

Although our primary intention is to examine younger sibling participation in AP exams, we can also examine whether older sibling's experiences impact the academic performance of these students. One concern is that the type of younger sibling induced to take AP by a shift in the signal their older sibling receives could be one that is unlikely to perform well in the AP course. Examining whether these induced students perform well on their AP exams helps us understand the potential policy implications of these signals and their impact on student behaviors.

Table 3 estimates impacts on total exams where a younger sibling scores a 3, 4, or 5 – typically considered "passing" – relative to scoring a 1 or 2, and then disaggregates these results between the matched AP exam and all other possible AP exams. When an older sibling earns a 3 it increases the total number of AP exams passed by 0.066 ($p<0.01$) relative to an increase of total AP exams not passed of 0.0173 (not statistically significant). This suggests the induced student passes roughly 80% of new AP exams taken. This total effect consists of two parts. First, the induced student mimics their older sibling by taking the same AP exam, increasing the likelihood they pass the exam by 1.4 percentage points (7.4%) and do not pass by 0.8 percentage points (5.8%). Second, the higher score has positive spillover effects, increasing the total number of other AP exams passed by 0.05 (3.1%) and not passed by 0.01 (1.0%).

At higher AP thresholds we find that the higher score increases the likelihood the younger sibling takes and passes the same AP exam, but this appears to come as a substitution effect away from other AP exams they would have passed. For example, earning a 4 over a 3 increases the likelihood the younger sibling passes that same AP exam by 0.013 but decreases the total number of other AP exams passed by 0.12, leading to a net effect of 0.001 more AP exams passed. Impacts on earning a lower score of 1 or 2 are all extremely small and insignificant, and results where the older sibling earns a 5 over a 4 exams are all identical if a bit smaller in magnitude.

Students may be more likely to pass AP exams via two methods: (1) their older sibling's performance induces them to take AP, thus changing the extensive margin, or (2) their older sibling's performance directly improves the younger sibling's performance (e.g., motivates them to study harder). Our treatment effect combines both of these mechanisms, but assuming that impacts are mostly through extensive margin effects suggests that the majority of students induced to take these exams are performing well, implying some underinvestment in AP courses nationwide for students at these specific margins.

5. Gender – Gaps, Impacts, and Mechanisms

5.1. AP Participation Gaps by Gender

Gender gaps exist across many educational outcomes and Advanced Placement exams are no exception. Figure 2 shows the female-to-male ratio in AP exam taking by subject, with numerical results provided in Appendix Table 1.[9] Females outnumber males most in AP Psychology exam, at 1.8 to 1, with the English Language and Literature exams close behind. At the opposite end of the spectrum are a number of male-dominated AP disciplines, including Physics, and Calculus BC. Yet the male-female divide is not evenly split into the areas commonly referred to as STEM and non-STEM. Females have high levels of participation in Biology and Environmental Science, both considered STEM. Generally we find that females are well represented in the non-STEM disciplines with one notable exception – Economics. For every male who takes a macroeconomics and microeconomics AP exam, only 0.82 and 0.73 females, respectively, take those exams.

5.2. Impacts by Gender and Sibling Gender

Table 4 examines differences in AP taking by sibling gender and shows that impacts are consistently larger for females than males. The top row of Table 4 shows when the older sibling is female and earns a relatively higher integer score, the younger sibling is 1.9 percentage points (5 percent) more likely to take the same AP exam, compared to 1.2 percentage points for males (3.3 percent), with a test of differences finding p=0.09. Similarly, when the younger sibling is female, regardless of older sibling gender, the estimate on matched AP exam is over twice as large compared to males (2.1 versus 1.0 percentage points or 5.4 and 2.9 percent, respectively, with p=0.00). We find similar differences when examining impacts on total AP exams taken (Table 4, column 2).

[9] Appendix Table 1 shows male-female representation in both the national sample (calculated using public available data) and our analytical sibling sample; there are no meaningful differences between the two data sources.

The bottom of Table 4 shows the differential peer effects by the interacted gender composition of sibling pairs. For shorthand we describe a sibling pair by referencing the older sibling first, so an "F-M" pairing would indicate a female older sibling and a male younger sibling. Overall, the largest match effects are found in F-F sibling pairs, with the younger sister 2.6 percentage points (6.3 percent) more likely to imitate their older sister, compared to roughly 1.2 and 1.5 percentage points (3.6 and 4.2 percent) for mixed gender siblings, and 0.9 percentage points (2.4 percent) for M-M pairs. Having an older sister also increases total AP participation by roughly 0.07 exams (roughly 2 percent) regardless of the gender of the younger sibling, with no impacts when the older sibling is male. Thus younger siblings seeing their older sister perform better are motivated to try that specific AP exams and increase their total AP portfolio. Graphical results for match and total AP effects are shown in Figure 3, with bandwidth robustness shown in Appendix Table 7 and CCT optimal bandwidth results in Appendix Table 8.

Appendix Table 9 shows the previous results for gender pairs disaggregated by threshold. Estimates for each sibling group is generally positive, but the primary difference appears to be that younger sisters are proportionately more responsive to their older sister at all thresholds. For example, younger sisters are 2.9 and 1.8 percentage points more likely to imitate their older sister at the 2/3, 3/4, and 4/5 thresholds, respectively. In contrast, younger siblings in mixed gender pairs imitate their older sibling at rates of roughly 2.0 and 0.7 at these same two thresholds, or a magnitude about half of the previous rate.

5.3. Underrepresentation and Gender

A commonly posed question is whether an individual is more responsive academically when they observe someone of similar background characteristics perform well in a subject in which they have been traditionally underrepresented (e.g., females in STEM) (Kofoed & mcGovney, 2019; Lim & Meer, 2017). Evidence for this in our context would be if, for example, a younger sister is more responsive to her older sister's performance in AP Physics – where females are less likely to engage – than AP Psychology, where female participation is common. One benefit of our context is that we can reverse this scenario and examine whether males are more responsive to their older brother's performance in AP Psychology and less responsive to AP Physics. To examine this issue we explore how the gender composition of siblings interact with the female-male participation ratio in the AP exam subject.

Overall our results do not support an interpretation that representation within the AP exam is driving the results. Table 5 examines the importance of gender representation by replicating Table 4 but interacting the female-to-male ratio of the AP exam subject with the AP threshold effect. For purposes of interpreting the estimates, the F-M ratio ranges 1.4 points (from 1.79 for Psychology to 0.36 for Physics) and is centered at the sample mean (previously shown in Table 4). Examining impacts on taking the identical AP exam (Table 5, column 1), we find only weak evidence of interaction effects, ranging from -0.10 to 0.04 percentage points. To put this in perspective, the most literal interpretation would imply that impacts on taking the identical AP exam for F-F siblings would be roughly 2.2 percentage points in AP Psychology and 3.6 percentage points in AP Physics. Yet the interaction effects are not only statistically weak, but relatively consistent in direction and magnitude across all gender groups. That is, younger brothers who observe their older sister perform well are also more responsive to AP Physics than AP Psychology. Thus the interaction may indicate that younger siblings are more responsive to sibling performance in more difficult exams or those in STEM fields, but not the typical representation of the class per se. The results and interpretation are the same for total AP exams taken (Table 5, column 2).

We explore the sensitivity of the previous results to different definitions of underrepresentation, but results are unchanged. Appendix Table 10 disaggregates impacts by STEM and non-STEM exams and finds no differential results when looking at sibling gender composition.[10] We also examine other approaches, such as moving away from linearity assumptions within the female-to-male ratio and examining results separately for strongly female- or male-dominant exams, or using the percent of exam takers who are female rather than the ratio (e.g., 64% in Psychology rather than the 1.79 value), in case there are non-linearities in this transformation that impact our estimates (results omitted for brevity). Overall the results suggest no meaningful differences in effects across these exams and are indifferent to the various male-female representations used.

Overall, the evidence seems clear that younger sisters are most responsive when they have older sisters who perform relatively well on AP exams, regardless of exam subject. The fact that we do not see larger impacts in subjects that are male dominated is somewhat of a rejection of the traditionally viewed "role-model" mechanism as it relates to gender homophily. In contrast to most previous studies, we are estimating impacts derived from observing sibling performance, rather than whether an older sibling chooses to participate or not within a discipline. Regardless, the results do show that one way younger females participate in advanced coursework, particularly those that are male dominated, is when their older sister performs well in the course.

[10] The results in Appendix Table 10 present some suggestive but generally mild evidence that younger females may be slightly more responsive to observing their older sister score higher on STEM exams, relative to female-female sibling pairs who observe higher non-STEM performance. P-values testing difference of coefficients are generally in the range of 0.08 to 0.23. We thank the reviewer for suggesting these tests.

6. Mechanism - Siblings or Parents?

A key question is whether the impacts we observe are due to direct transmission of peer effects between siblings, or instead driven by changes in parental behaviors. Direct transmission would indicate that younger siblings change their decisions based on conversations between siblings or by simply observing their outcomes and updating their perceived probability of success in a subject. On the other hand, older sibling's performance may cause parents to update their beliefs on the expected benefits and instruct the younger sibling to take the same AP exam. Distinguishing between the two mechanisms gives us a sense of how information travels through families. This path of information may implicate many more contexts than AP exam choice, such as college and major choice or even non-educational settings, like sports and activities participation. In addition, there may be similar settings where a parent or siblings are not involved. For example, a student may learn of the success of an older friend or classmate without a parent ever knowing. These types of peer effects are likely to be subtle yet pervasive.

Our first piece of evidence relies on previous results and points to the direct sibling to sibling mechanism. Our gender-specific results are somewhat difficult to rationalize if the parents are involved. While parents may have gender preferences, the fact that the F-F result is stronger than both the F-M and M-F results suggest that siblings, not parents, are driving the results. A potential threat to the argument is that gender composition of siblings are correlated with unobserved cross-family differences and that drives the differential result. However, when we split the sample by student ethnicity or parental education, as in Appendix Table 11 we do not see differential effects. Both black and Hispanic students and parents with no college education are less likely to encounter AP in their schools and may have less information about AP in general. The fact that they do not respond differentially to older siblings scoring relatively higher on AP compared to whites, Asians, and parents with college education suggests that parents aren't learning and imparting knowledge to their children but rather siblings from all backgrounds are imparting information to one another. These results reinforce the idea that there is something specific to gender composition of the siblings, not about parents or students with a knowledge deficit learning from older siblings.

The above evidence is suggestive that the knowledge is transferred through siblings but there are alternative explanations we cannot rule out. For example, parents may try to figure out whether their (younger) daughter should take an AP course, and they think that the older sister's signal is more informative about this than the signal of an older brother. And this may be even more true for brothers.

6.1. Multiple Sibling Households

We have the opportunity to learn something about the mechanism involved by exploiting three sibling households. So far, our results show the impacts of the older sibling most proximate in age to the younger sibling (e.g., our oldest observed sibling on the second oldest observed sibling). Here we examine whether there are direct impacts of the oldest sibling's performance on siblings less proximate in age, by restricting to households where we observe exactly three siblings. (Relatively few families have more than three children and their inclusion does not change results). In Table 6 we estimate effects at the 2/3 threshold. Although direct impacts of the oldest sibling on the 2^{nd} sibling of matching the AP exam are 2.0 percentage points (row 2), similar to the main effect shown in Table 2, the direct effect of the oldest sibling on the 3rd sibling is just 0.7 percentage points and statistically insignificant. Estimates on other outcomes also show quantitatively similar results between Table 2 and Table 6 when focused on just the second sibling, although statistical significance is often lacking due the smaller sample size of this group, whereas point estimates for the impacts of the oldest sibling score on the third sibling are all much smaller and essentially zero. We can also measure the impact of the second sibling directly on the third sibling, though these results are much more speculative as our sample at this stage is lacking in power. In these results we use the same running variable of continuous AP score but for the sample of second siblings, thus needing to restrict to these younger siblings who took an AP exam between 2004 and 2009. We find the second sibling's higher AP performance leads to a 1.3 percentage point increase in the third sibling taking the matched AP exam, relatively similar to our initial result, though larger standard errors preclude statistical significance; expanding to a 10 point bandwidth rather than 5, in order to increase power, leads to a 1.9 percentage point effect significant a $p<0.05$ (estimates not shown for brevity). Results on other main outcomes are all small and statistically insignificant, though too noisy to say much more at this stage.

Combined, these results show that younger siblings are more likely to be impacted by the AP exam score of an older sibling that is closest in age than their other siblings.[11] If our results were driven by parents updating their beliefs on the benefits of an AP exam this would suggest that they, in essence, forget when applying this information to the youngest siblings. Alternatively, if the middle sibling follows the oldest sibling but then performs poorly, parents may downgrade their assessment of the exam for the third child. But as we have already shown, we see that younger siblings induced to take an AP exam perform well.

7. Conclusion

[11] At the bottom of Appendix Table 11, we find that there are no heterogeneous effects by distance in age (e.g., one year versus three years apart).

We show that younger siblings are attuned to the academic performance of their older siblings, using their performance as guideposts to determine the extent they participate in challenging academic disciplines. We examine this topic in the context of AP exams, taken by millions of students across the country, which allows us to examine a large and diverse group of students in a key policy context. While we cannot know the entirety of the downstream effects, prior research suggests two important points. First, exposure to curricula can change longer-term outcomes (Bian, Leslie, & Cimpian, 2017; Fricke, Grogger, & Steinmayr, 2015; Jackson, 2014), suggesting that course choice is an important outcome for younger siblings. Second, we find positive impacts on exam taking translate into positive impacts on passing an AP exam, which we know can impact on-time college graduation and major choice (Avery et al., 2018; Smith et al., 2017). Our results differ from recent experimental work that randomized students into AP Biology and Chemistry courses, where large increases in participation did not translate into any measurable increase in students passing these exams (Conger et al., 2020).[12] The samples in these two studies are substantially different, with ours consisting of students who traditionally participate in AP at high rates, relative to their study of low-income schools offering STEM courses for the first time. Care should be taken about who gets promoted into AP courses, though prior work has found that universal screening is likely to benefit traditionally underrepresented groups (Card & Giuliano, 2016a, 2016b).

In contrast to other studies, our person of primary interest (younger sibling) is merely exposed to a signal of performance, not a distinct educational intervention, which may explain our

[12] Conger et al. (2020) also find that students assigned to AP have a lower probability of enrolling in a selective college, likely in part due to their poor experience in these courses. This study is underpowered to detect changes in college-going rates; for example, at the 2/3 threshold younger siblings are roughly 1 percentage point more likely to take AP, and even assuming this increases college-going by 10 percentage points our standard errors are incapable of detecting point estimates of 0.001. Indeed we find null results very close to zero on college going or college selectivity at the 2/3 threshold.

relatively modest effect sizes. The causal impact of receiving an integer score of a 3 relative to a 2 increases a younger sibling's likelihood of taking a matched exam by 7% and total AP exams taken by almost 3%; viewed another way, the increase in total AP exams taken is a 0.03 standard deviation effect size (the standard deviation at this threshold is roughly 3.0 exams). Older siblings typically have stronger academic performance, with birth order effects on cognitive and non-cognitive skills measured at roughly 0.1 to 0.2 standard deviations, though results vary across studies (Black, Devereux, & Salvanes, 2011; Black, Grönqvist, & Öckert, 2018). Classroom peers tend to have a strong influence on their schoolmates (Black, Devereux, & Salvanes, 2013; S. E. Carrell, Hoekstra, & Kuka, 2018; Lavy, Paserman, & Schlosser, 2012), and older siblings in particular exert strong influence on younger siblings, with prior sibling spillover studies tending to produce generally smaller, though still substantial, impacts on exam performance between 0.02 to 0.16 standard deviations (Karbownik & Ozek, 2019; Landersø, Nielsen, & Simonsen, forthcoming; Nicoletti & Rabe, 2018; Qureshi, 2018). Of course these estimates are second-order relative to more direct educational interventions, such as test score gains in charter school settings (ranging from 0.2 to 0.4 standard deviations as in Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011) and later studies) or from class size reductions (Chetty et al., 2011), though not all studies in these areas find positive impacts (e.g., Angrist, Lavy, Leder-Luis, and Shany (2019)).

In a paper with academic context most similar to ours, Joensen and Nielsen (2018) find that older sibling participation in a math-science courses increases younger sibling participation by two to three percentage points, an estimate relatively similar to our observed effect of matched AP exams at the 2/3 threshold. Given that students are bombarded with constant signals of academic performance, from the insignificant to the consequential (e.g., Avery et al. (2018)), this suggests that cumulatively these signals may have meaningful impacts on subsequent college and career choices.

We also show heterogeneous impacts of sibling effects that differ from previous research, demonstrating the importance of context, differences in questions and methods, and collective bodies of literature. In families with more than two siblings, we only find direct impacts on the younger sibling closest in age, suggesting that younger siblings are more attuned to the academic performance of the sibling that immediately precedes them, or the signal is washed out by the intermediate sibling. This is in contrast to Dahl, Løken, and Mogstad (2014), who finds spillover effects that magnify over time in the family or at work. We also find the largest impacts between sisters, as opposed to Joensen and Nielsen (2018) who find stronger transmission between brothers. On the one hand, the topic area of our study and Joensen and Nielsen (2018) are quite similar, given that they also focus on more advanced curriculum in high school. On the other hand, they differ dramatically both in the population studied (e.g., Danish students in the 1980s versus American students in the 2000s) and the associated instrument (e.g., they measure differential exposure to advanced curriculum, whereas we compare two students with similar exposure who are academically equivalent, but who receive different signals of their performance). These findings highlight the importance of using multiple studies to draw inferences about sibling's effects, rather than relying on a single study in isolation.

Finally, these results have particular relevance for education policy. Much prior research has focused on underrepresented role models as serving to close educational gaps (Egalite & Kisida, 2018; Lim & Meer, 2017), with a particular focus on females and underrepresented minorities in STEM fields and economics (Arcidiacono, Hotz, & Kang, 2012; Buckles, 2019). We show that siblings are determinants of course-taking, but that this is largely true for siblings of all groups and all fields. These results imply that expanding coursework for all students is unlikely to close gaps. Instead, targeted expansions at students who are underrepresented have the added benefit of the spillover effects from good performance. These results also show us that parents are not the only educational influencers in a child’s life outside the school building and it is worth educators, researchers, and policy makers considering new policies and interventions to harness the unique relationship of siblings.

References

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters And Pilots. *The Quarterly Journal of Economics, 126*(2), 699-748. doi:10.1093/qje/qjr017

Aguirre, J., & Matta, J. (2018). *Walking in Your Footsteps: Sibling Spillovers in Higher Education Choices*. SSRN Working Paper 3255875.

Altmejd, A., Barrios-Fernandez, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., . . . Smith, J. (2020). *O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries*. Annenberg Institute EdWorkingPaper 20-230. Providence, RI.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics, 4*(1), 185-223. doi:doi:10.1146/annurev-economics-080511-110908

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2019). Maimonides Rule Redux. *American Economic Review: Insights, 1*(3), 309-324.

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics, 166*(1), 3-16. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407611001151

Avery, C. N., Gurantz, O., Hurwitz, M., & Smith, J. (2018). Shifting College Majors in Response to Advanced Placement Exam Scores. *Journal of Human Resources, 53*(4), 918-956. Retrieved from http://jhr.uwpress.org/content/53/4/918.abstract

Becker, G. S., & Tomes, N. (1976). Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy, 84*(4), 143-162.

Bennett, M., & Bergman, P. (2019). *Better Together? Social Networks in Truancy and the Targeting of Treatment*. New York, NY.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children’s interests. *Science, 355*(6323), 389-391. doi:10.1126/science.aah6524

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., . . . Simonsen, M. (2017). *Sibling Spillovers*. NBER Working Paper No. 23062.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Older and Wiser? Birth Order and IQ of Young Men. *CESifo Economic Studies, Oxford University Press, 57*(1), 103-120.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics, 31*(1), 119-153. doi:10.1086/666872

Black, S. E., Grönqvist, E., & Öckert, B. (2017). *Born to Lead? The Effect of Birth Order on Non-Cognitive Abilities*. NBER Working Paper No. 23393.

Black, S. E., Grönqvist, E., & Öckert, B. (2018). Born to Lead? The Effect of Birth Order on Noncognitive Abilities. *Review of Economics and Statistics, 100*(2), 274-286.

Buckles, K. (2019). Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage. *Journal of Economic Perspectives, 33*(1), 43-60. doi:doi: 10.1257/jep.33.1.43

Bursztyn, L., & Jensen, R. (2015). How Does Peer Pressure Affect Educational Investments? *The Quarterly Journal of Economics, 130*(3), 1329-1367. doi:10.1093/qje/qjv021

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics, 129*(3), 1409-1447. doi:10.1093/qje/qju009

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Non-parametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica, 82*(6), 2295-2326.

Card, D., & Giuliano, L. (2016a). Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review, 106*(10), 2783-2816. doi:doi: 10.1257/aer.20150484

Card, D., & Giuliano, L. (2016b). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678-13683. Retrieved from http://www.pnas.org/content/pnas/113/48/13678.full.pdf

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The Long-Run Effects of Disruptive Peers. *American Economic Review, 108*(11), 3377-3415. doi:10.1257/aer.20160763

Carrell, Scott E., Page, M. E., & West, James E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics, 125*(3), 1101-1144. doi:10.1162/qjec.2010.125.3.1101

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *Quarterly Journal of Economics, 126*(4), 1593-1660.

Conger, D., Kennedy, A. I., Long, M. C., & McGhee, R. (2019). The Effect of Advanced Placement Science on Students' Skills, Confidence and Stress. *Journal of Human Resources*. doi:10.3368/jhr.56.1.0118-9298R3

Conger, D., Long, M. C., & Iatarola, P. (2009). Explaining race, poverty, and gender disparities in advanced course-taking. *Journal of Policy Analysis and Management, 28*(4), 555-576. doi:10.1002/pam.20455

Conger, D., Long, M. C., & McGhee, R. (2020). *Advanced Placement and Initial College Enrollment: Evidence from an Experiment* Association for Education, Finance, and Policy 2020 Conference.

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer Effects in Program Participation. *American Economic Review, 104*(7), 2049-2074. doi:doi: 10.1257/aer.104.7.2049

Darolia, R., Koedel, C., Main, J. B., Ndashimye, F., & Yan, J. (2018). *High School Course Access and Postsecondary STEM Enrollment and Attainment*. National Center for Analysis of Longitudinal Data in Education Research. Washington DC.

Dee, T. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2), 158-165.

Dobbie, W., & Fryer, R. G., Jr. (2014). The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. *American Economic Journal: Applied Economics, 6*(3), 58-75. doi:doi: 10.1257/app.6.3.58

Egalite, A. J., & Kisida, B. (2018). The Effects of Teacher Match on Students' Academic Perceptions and Attitudes. *Educational Evaluation and Policy Analysis, 40*(1), 59-81. doi:10.3102/0162373717714056

Ellis, J. R., & Gershenson, S. (2016). *LATE for the meeting: Gender, peer advising, and college success*. IZA Discussion Paper No. 9956.

Fricke, H., Grogger, J., & Steinmayr, A. (2015). *Does exposure to economics brig new majors to the field? Evidence from a natural experiment*. NBER Working Paper 21130. Cambridge, MA.

Gershenson, S., Hart, C. M. D., Hyman, J. M., Lindsay, C. A., & Papageorge, N. (2018). *The Long-Run Impacts of Same-Race Teachers*. NBER Working Paper No. 25254. Cambridge, MA.

Goodman, J. S. (forthcoming). The Labor of Division: Returns to Compulsory High School Math Coursework. *Journal of Labor Economics*.

Goodman, J. S., Hurwitz, M., Smith, J., & Fox, J. (2015). The relationship between siblings' college choices: Evidence from one million SAT-taking families. *Economics of Education Review, 48*, 75-85. doi:https://doi.org/10.1016/j.econedurev.2015.05.006

Grissom, J. A., & Redding, C. (2016). Discretion and Disproportionality:Explaining the Underrepresentation of High-Achieving Students of Color in Gifted Programs. *AERA Open, 2*(1), 2332858415622175. doi:10.1177/2332858415622175

Gurantz, O. (forthcoming). How college credit in high school impacts postsecondary course-taking: the role of AP exams. *Education Finance and Policy*.

Heissel, J. A. (forthcoming). Teen fertility and siblings' outcomes: Evidence of family spillovers using matched samples. *Journal of Human Resources*.

Jackson, K. (2014). Do College-Preparatory Programs Improve Long-Term Outcomes? *Economic Inquiry, 52*(1), 72-99. doi:10.1111/ecin.12040

Joensen, J. S., & Nielsen, H. S. (2009). Is there a Causal Effect of High School Math on Labor Market Outcomes? *Journal of Human Resources, 44*(1), 171-198. doi:10.3368/jhr.44.1.171

Joensen, J. S., & Nielsen, H. S. (2018). Spillovers in education choice. *Journal of Public Economics, 157*, 158-183. doi:https://doi.org/10.1016/j.jpubeco.2017.10.006

Karbownik, K., & Ozek, U. (2019). *Setting a good example? Examining sibling spillovers in educational achievement using a regression discontinuity design*. NBER Working Paper 26411. Boston, MA.

Kofoed, M. S., & mcGovney, E. (2019). The Effect of Same-Gender and Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point. *Journal of Human Resources, 54*(2), 430-467. doi:10.3368/jhr.54.2.0416.7838R1

Kolesár, M., & Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review, 108*(8), 2277-2304. doi:10.1257/aer.20160945

Landersø, R. K., Nielsen, H. S., & Simonsen, M. (forthcoming). Effects of School Starting Age on the Family. *Journal of Human Resources*.

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *The Economic Journal, 122*(559), 208-237. doi:10.1111/j.1468-0297.2011.02463.x

Lim, J., & Meer, J. (2017). The Impact of Teacher–Student Gender Matches: Random Assignment Evidence from South Korea. *Journal of Human Resources, 52*(4), 979-997. doi:10.3368/jhr.52.4.1215-7585R1

Lindquist, M. J., Sol, J., & Praag, M. V. (2015). Why Do Entrepreneurial Parents Have Entrepreneurial Children? *Journal of Labor Economics, 33*(2), 269-296. doi:10.1086/678493

Nicoletti, C., & Rabe, B. (2018). Sibling spillover effects in school achievement. *Journal of Applied Econometrics, 0*(0). doi:10.1002/jae.2674

Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics, 122*(3), 1067-1101. doi:10.1162/qjec.122.3.1067

Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (1990). Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries. *American Economic Review, 80*(5), 1139-1156.

Porter, C., & Serra, D. (2019). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*.

Qureshi, J. A. (2018). Siblings, Teachers, and Spillovers on Academic Achievement. *Journal of Human Resources, 53*(1), 272-297. doi:10.3368/jhr.53.1.0815-7347R1

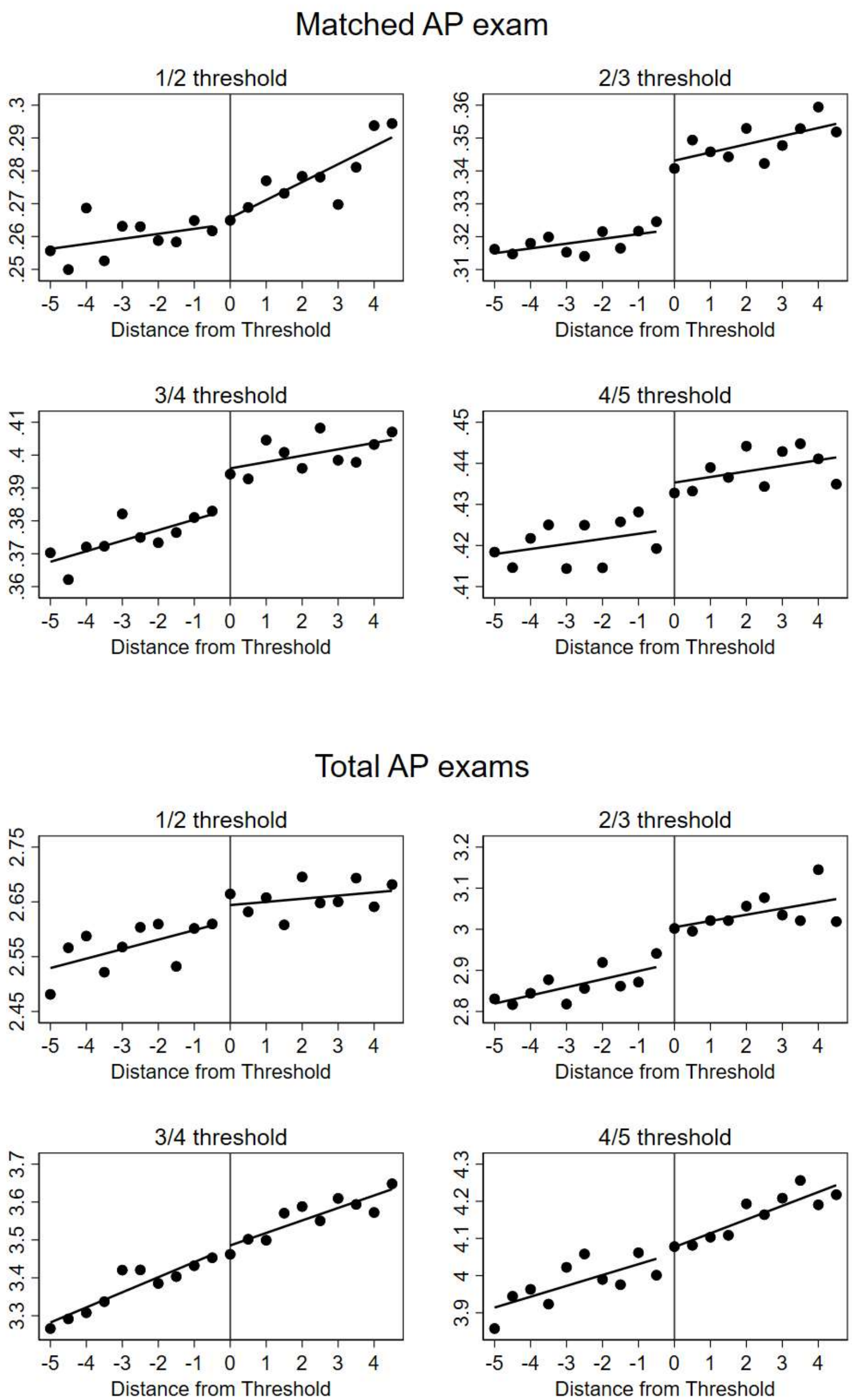
Rose, H., & Betts, J. R. (2004). The Effect of High School Courses on Earnings. *Review of Economics and Statistics, 86*(2), 497-513.

Sacerdote, B. (2007). How Large are the Effects from Changes in Family Environment? A Study of Korean American Adoptees. *The Quarterly Journal of Economics, 122*(1), 119-157. doi:10.1162/qjec.122.1.119

Smith, J., Hurwitz, M., & Avery, C. (2017). Giving College Credit Where It Is Due: Advanced Placement Exam Scores and College Outcomes. *Journal of Labor Economics, 35*(1), 67-147. Retrieved from http://www.journals.uchicago.edu/doi/abs/10.1086/687568

Yi, J., Heckman, J. J., Zhang, J., & Conti, G. (2015). Early Health Shocks, Intra-household Resource Allocation and Child Outcomes. *The Economic Journal, 125*(588), F347-F371. doi:10.1111/ecoj.12291

Figure 1: Impact of older sibling receiving a higher AP integer score on younger sibling outcomes



Notes. Figures identify the behaviors of younger siblings using a running variable that is the older siblings' continuous AP score. All pictures use five point bandwidths with bins of 0.5 points. Pictures correspond to regression results in Table 2, though use unweighted Stata lfits.

Figure 2: Ratio of female to male participation in AP exams, 2004-2009

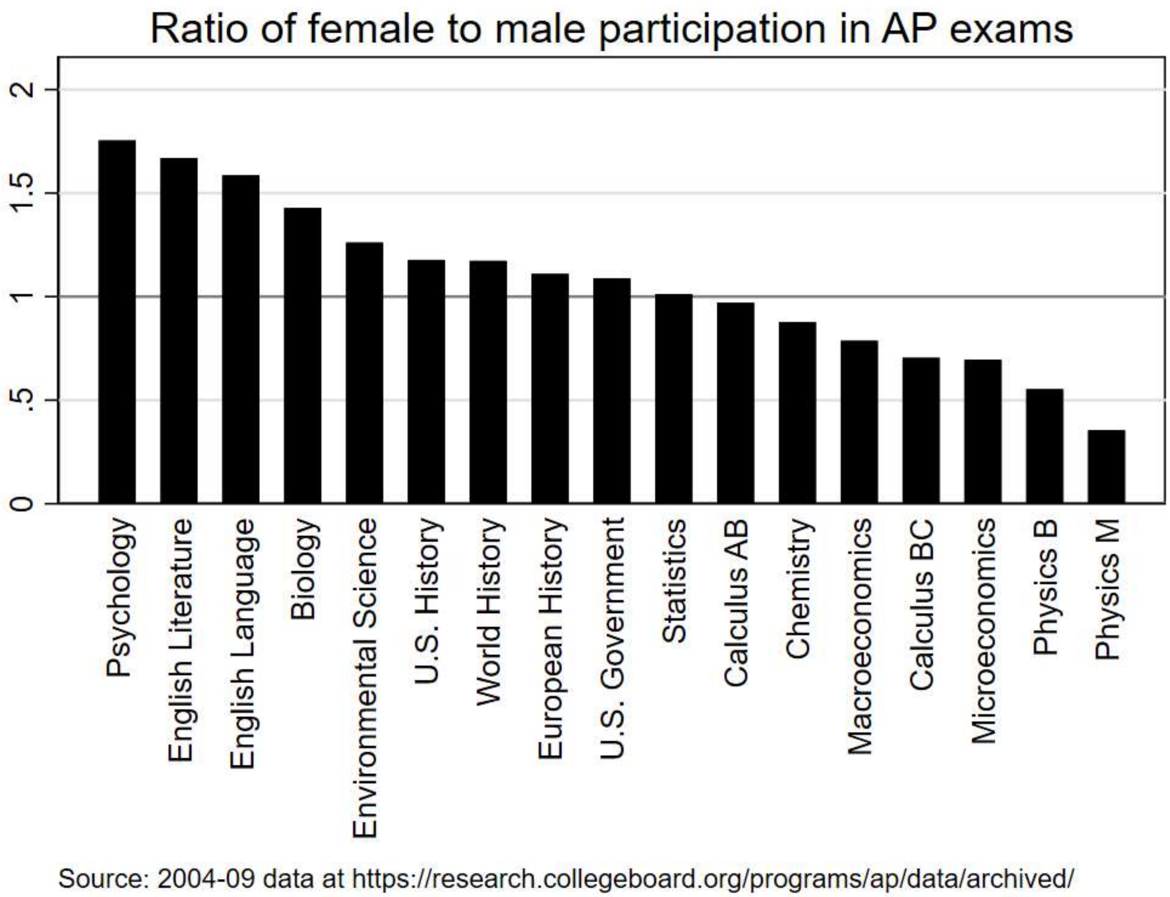
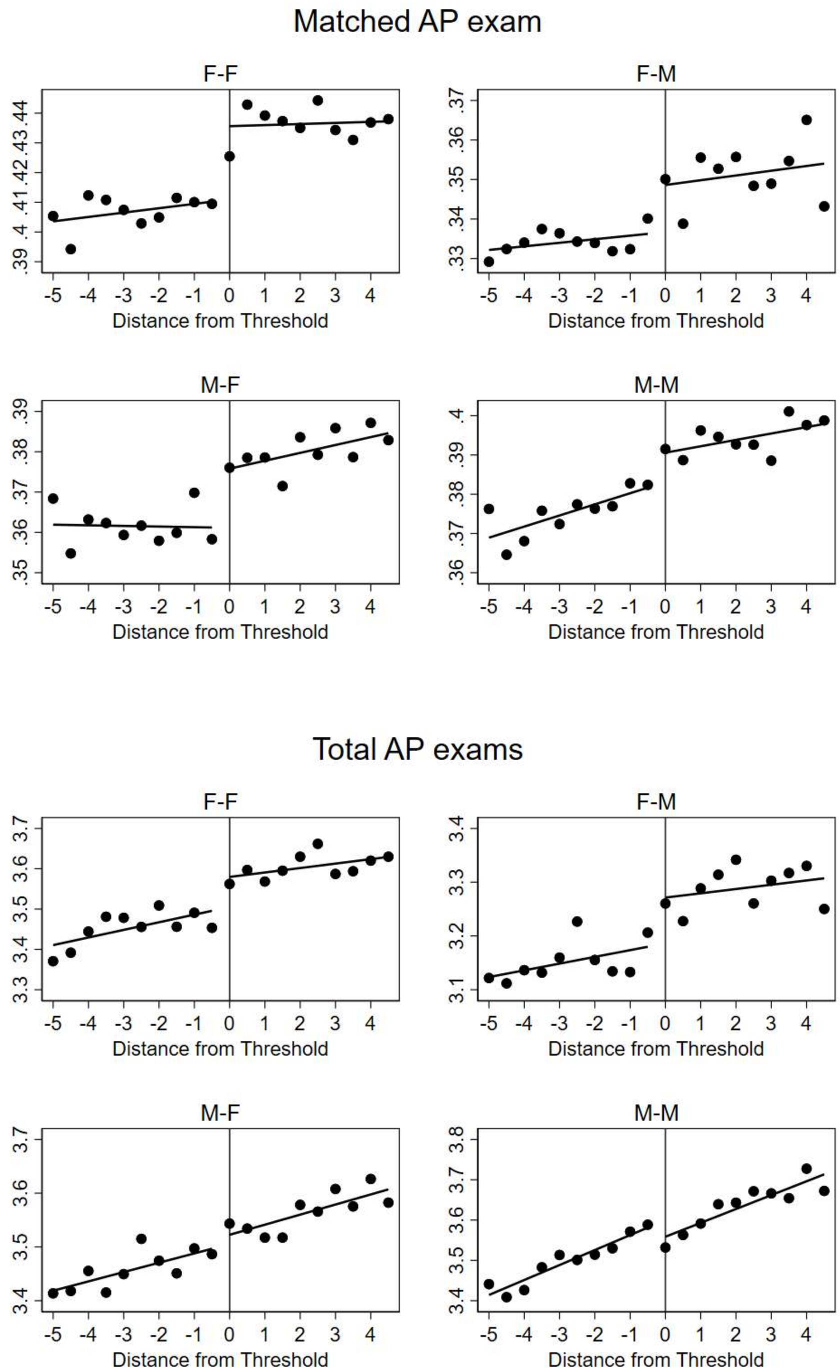


Figure 3: Impact of older sibling receiving a higher AP integer score on younger sibling outcomes, by sibling gender



Notes. Figures identify the behaviors of younger siblings using a running variable that is the older siblings' continuous AP score. With "M" for male and "F" for female, the first letter is the older sibling's gender and the second letter is the younger sibling's gender. All pictures use five point bandwidths with bins of 0.5 points. Pictures correspond to regression results in Table 4, though use unweighted Stata lfits.

Table 1. Descriptive statistics, younger siblings in analytic sample

	Full Sample	Within 5 points of threshold			
		1/2	2/3	3/4	4/5
Sample size	2347487	180187	321483	364037	258243
Male	49.1%	48.5%	49.0%	49.2%	49.4%
White	70.6%	63.4%	71.3%	74.8%	75.4%
Asian	13.4%	13.2%	12.6%	13.0%	14.6%
African-American	3.8%	6.5%	3.8%	2.4%	1.5%
Hispanic	7.8%	12.4%	7.9%	5.5%	4.2%
Parent education: HS or less	6.6%	11.1%	7.1%	4.7%	3.2%
Parent education: HS or more	12.7%	18.1%	15.0%	11.0%	7.8%
Parent education: Bachelor or moı	75.1%	64.4%	72.3%	79.0%	83.8%
Income: LT $50K	10.3%	15.5%	10.8%	8.0%	6.2%
Income: $50K - $100K	21.2%	23.0%	22.7%	21.0%	19.0%
Income: $100K+	24.0%	20.4%	23.1%	25.7%	27.6%
Took SAT	82.6%	81.3%	81.7%	82.9%	84.6%
Average SAT	1161	1079	1128	1181	1234
AP participation					
Took matched AP exam	36.5%	26.9%	33.4%	38.7%	42.9%
Took at least one AP exam	74.1%	66.6%	71.0%	76.3%	81.2%
Total AP exams	3.34	2.63	2.96	3.46	4.06

Notes. Includes all individuals observed within the College Board data who had an older sibling that took at least one AP exam.

Table 2. Impact of receiving a higher AP integer score on younger sibling outcomes

	N	(1) Took identical AP exam	(2) Total AP exams	(3) Took AP (extensive margin)	(4) Total AP (intensive margin)
Individual Thresholds	1119572	0.0137**	0.0290*	0.0027	0.0268*
		(0.0018)	(0.0115)	(0.0016)	(0.0122)
Baseline value		0.357	3.30	0.741	4.45
Individual Thresholds					
1/2	180187	0.0042	0.0332	0.0014	0.0432
		(0.0041)	(0.0268)	(0.0045)	(0.0306)
		0.263	2.61	0.664	3.92
2/3	321483	0.0219**	0.0828**	0.0111**	0.0524*
		(0.0033)	(0.0207)	(0.0032)	(0.0225)
		0.323	2.90	0.701	4.14
3/4	364037	0.0148**	0.0019	0.0009	-0.0003
		(0.0032)	(0.0206)	(0.0028)	(0.0212)
		0.382	3.44	0.763	4.51
4/5	258243	0.0090*	-0.0088	-0.0045	0.0163
		(0.0038)	(0.0258)	(0.0031)	(0.0256)
		0.424	4.03	0.809	4.99

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Regressions on individual thresholds sum to slightly more than the total sample using all thresholds as some individuals overlap the five point boundary; in these cases, individuals are assigned to the higher threshold. Baseline values below regression estimates include all individuals within one point below the threshold.

Table 3. Impact of receiving a higher AP integer score on younger sibling AP exam performance

	N	AP exams earned a 3, 4, or 5			AP exams earned a 1 or 2		
		Total	Identical exam	All other	Total	Identical exam	All other
Top three thresholds	943130	0.0260*	0.0123**	0.0137	0.0033	0.0035**	-0.0002
		(0.0118)	(0.0018)	(0.0108)	(0.0057)	(0.0012)	(0.0052)
Control Mean		2.495	0.273	2.222	0.931	0.101	0.830
Individual Thresholds							
1/2	180187	0.0171	0.0025	0.0146	0.0160	0.0015	0.0145
		(0.0212)	(0.0030)	(0.0197)	(0.0157)	(0.0034)	(0.0143)
Control Mean		1.398	0.113	1.284	1.208	0.150	1.058
2/3	321483	0.0658**	0.0142**	0.0516**	0.0173	0.0077**	0.0096
		(0.0179)	(0.0028)	(0.0165)	(0.0106)	(0.0024)	(0.0096)
Control Mean		1.832	0.192	1.641	1.073	0.132	0.941
3/4	364037	0.0011	0.0132**	-0.0121	0.0008	0.0016	-0.0008
		(0.0191)	(0.0030)	(0.0175)	(0.0090)	(0.0020)	(0.0083)
Control Mean		2.520	0.282	2.237	0.923	0.100	0.823
4/5	258243	0.0006	0.0079*	-0.0073	-0.0095	0.0011	-0.0105
		(0.0250)	(0.0037)	(0.0231)	(0.0095)	(0.0020)	(0.0089)
Control Mean		3.262	0.357	2.905	0.771	0.067	0.704

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. The top row combines estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Baseline values below regression estimates include all individuals within one point below the threshold.

Table 4. Impact of receiving a higher AP integer score on younger sibling outcomes, by gender

Oldest Sibling	Younger sibling	N	Took identical AP exam	Total AP exams	Took AP (extensive margin)	Total AP (intensive margin)
F		498058	0.0189**	0.0688**	0.0067**	0.0571**
			(0.0027)	(0.0174)	(0.0024)	(0.0182)
Baseline value			0.374	3.33	0.743	4.47
M		445072	0.0123**	-0.0143	-0.0008	-0.0123
			(0.0028)	(0.0187)	(0.0025)	(0.0193)
			0.374	3.54	0.768	4.60
	F	479012	0.0210**	0.0421*	0.0035	0.0355*
			(0.0027)	(0.0172)	(0.0024)	(0.0174)
			0.389	3.48	0.784	4.44
	M	464118	0.0103**	0.0156	0.0025	0.0111
			(0.0028)	(0.0189)	(0.0026)	(0.0202)
			0.359	3.37	0.725	4.64
PAIRS						
F	F	264877	0.0258**	0.0663**	0.0068*	0.0491*
			(0.0038)	(0.0233)	(0.0032)	(0.0237)
			0.410	3.47	0.783	4.43
F	M	248669	0.0120**	0.0722**	0.0066+	0.0670*
			(0.0038)	(0.0259)	(0.0037)	(0.0282)
			0.336	3.17	0.700	4.52
M	F	232949	0.0152**	0.0133	-0.0003	0.0180
			(0.0039)	(0.0254)	(0.0035)	(0.0257)
			0.364	3.49	0.785	4.45
M	M	233890	0.0091*	-0.0413	-0.0014	-0.0418
			(0.0040)	(0.0275)	(0.0036)	(0.0288)
			0.383	3.58	0.752	4.76

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions combine estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Baseline values below regression estimates include all individuals within one point below the threshold.

Table 5. Impact of receiving a higher AP integer score on younger sibling outcomes, by sibling gender and using interaction term for female-male ratio of AP exam participation

Oldest Sibling	Younger sibling	N		(1) Took identical AP exam	(2) Total AP exams
F	F	256833	Threshold effect	0.030**	0.032+
				(0.004)	(0.017)
			Interaction with female-male participation ratio	-0.010+	-0.016
				(0.005)	(0.022)
			Baseline mean	0.410	3.47
F	M	241225	Threshold effect	0.016**	0.051**
				(0.004)	(0.018)
			Interaction with female-male participation ratio	-0.008	-0.004
				(0.005)	(0.022)
			Baseline mean	0.336	3.17
M	F	222179	Threshold effect	0.015**	0.048*
				(0.004)	(0.019)
			Interaction with female-male participation ratio	0.004	-0.002
				(0.005)	(0.024)
			Baseline mean	0.364	3.49
M	M	222893	Threshold effect	0.010*	0.015
				(0.004)	(0.019)
			Interaction with female-male participation ratio	-0.007	-0.019
				(0.005)	(0.024)
			Baseline mean	0.383	3.58

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. All results are strictly on the 2/3 threshold and include an additional term that interacts the AP specific female-male ratio with the threshold to determine heterogeneous effects. Baseline values below regression estimates include all individuals within one point below the

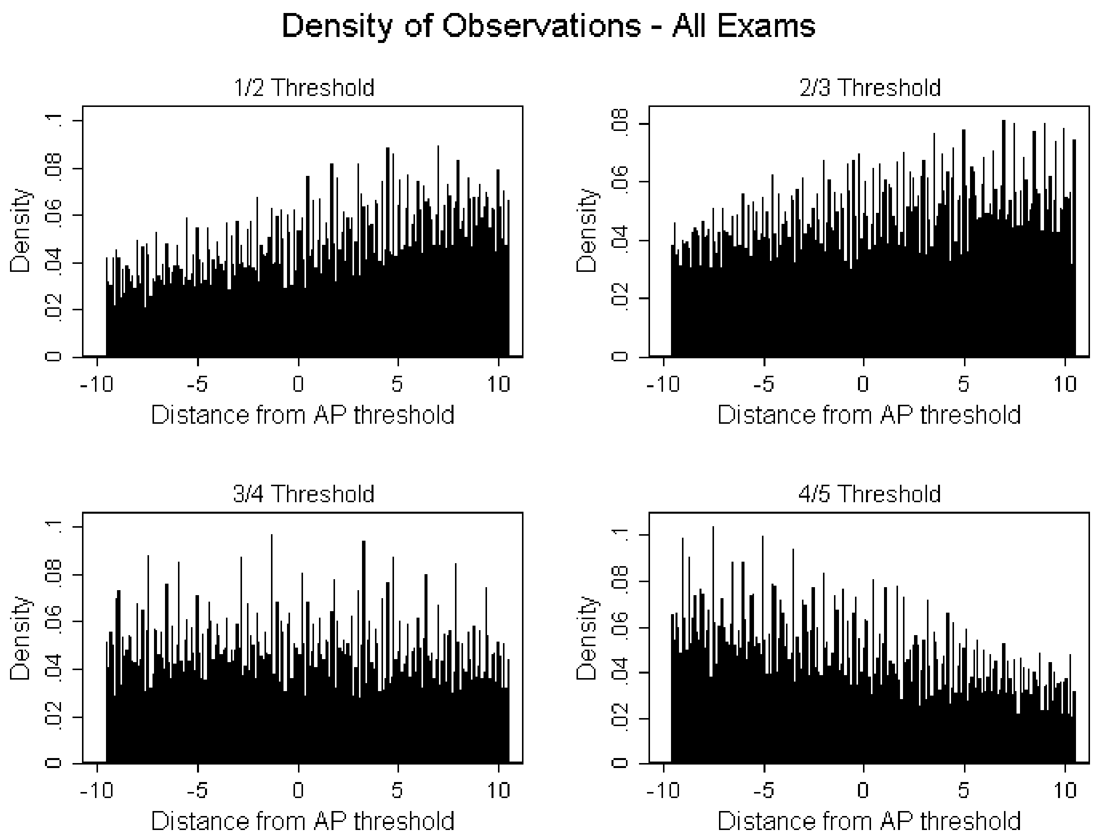
Table 6. Impact of receiving a higher AP integer score on younger sibling outcomes at top 3 thresholds, results for extended siblings in families with 3 siblings

	N	(1) Took identical AP exam	(2) Total AP exams	(3) Took AP (extensive margin)	(4) Total AP (intensive margin)
All siblings	103642	0.0136*	0.0298	-0.0007	0.0474
		(0.0065)	(0.0443)	(0.0063)	(0.0456)
		0.326	2.98	0.719	4.15
Impact on 2nd sibling only	51821	0.0201*	0.0727	0.0010	0.0930+
		(0.0082)	(0.0504)	(0.0078)	(0.0537)
		0.341	2.95	0.725	4.07
Impact on 3rd sibling only	51821	0.0071	-0.0131	-0.0024	0.0008
		(0.0081)	(0.0528)	(0.0080)	(0.0570)
		0.311	3.01	0.713	4.23
Impact of 2nd sibling on 3rd sibling	40676	0.0126	-0.0129	0.0066	-0.0510
		(0.0093)	(0.0596)	(0.0087)	(0.0630)
		0.346	3.21	0.734	4.37

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions combine estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Baseline values below regression estimates include all individuals within one point below the threshold.

APPENDICES – FOR ONLINE DISTRIBUTION

Appendix Figure 1: Density of AP scores for students with observed younger sibling, all students in 2004 through 2009 high school cohorts



Notes. Figure includes all student-exam observations of older siblings within 10 points of the integer AP score threshold for the seventeen AP exams listed in Table 2 for the years 2004 through 2009 graduating high school cohorts.

Appendix Table 1. Descriptive statistics of AP exams in older sibling sample

	Older sibling sample										National sample	
Exam	*Number of Exams*	*Average Score*	*Pass Rate*	*Average SAT score*	*African-American*	*Asian*	*Hispanic*	*White*	*Female*	*Ratio of female to males*	*Female*	*Ratio of female to males*
Non-STEM exams												
English Literature	359101	3.147	72.0%	1229	4.4%	10.5%	8.2%	72.4%	62.5%	1.67	63.3%	1.72
U.S. History	312369	3.037	63.6%	1232	3.9%	11.5%	8.0%	72.2%	54.1%	1.18	54.2%	1.18
English Language	252924	3.102	68.9%	1224	4.4%	11.4%	9.8%	69.9%	61.3%	1.59	62.5%	1.67
U.S. Government	193300	2.962	62.4%	1229	3.8%	12.1%	9.5%	70.0%	52.1%	1.09	52.6%	1.11
Psychology	97433	3.440	75.7%	1199	4.8%	12.5%	6.8%	71.3%	63.7%	1.75	64.1%	1.79
European History	93629	3.198	76.3%	1262	2.5%	11.1%	6.0%	75.8%	52.6%	1.11	53.0%	1.13
Macroeconomics	73964	3.002	60.6%	1262	3.4%	16.3%	10.1%	65.8%	44.0%	0.79	45.0%	0.82
World History	61729	3.054	65.7%	1229	4.9%	13.7%	10.0%	66.7%	54.0%	1.17	54.6%	1.21
Microeconomics	46104	3.136	67.4%	1276	3.2%	16.8%	7.3%	68.2%	41.0%	0.70	42.1%	0.73
STEM exams												
Calculus AB	259471	3.130	64.2%	1252	3.4%	13.1%	7.3%	72.2%	49.2%	0.97	48.5%	0.94
Biology	156550	3.217	66.7%	1243	4.0%	17.2%	6.8%	67.2%	58.8%	1.43	58.6%	1.41
Chemistry	104523	2.997	62.3%	1292	3.0%	18.8%	5.8%	67.9%	46.7%	0.88	46.8%	0.88
Statistics	97066	3.037	65.6%	1242	3.4%	14.7%	6.4%	71.3%	50.3%	1.01	50.8%	1.03
Calculus BC	88398	3.778	82.3%	1357	1.9%	21.5%	4.9%	67.2%	41.3%	0.70	40.9%	0.69
Physics B	67285	2.948	64.4%	1296	2.7%	16.5%	6.6%	69.8%	35.6%	0.55	34.8%	0.53
Environmental Science	54701	2.913	60.4%	1197	3.9%	10.8%	7.8%	73.0%	55.8%	1.26	55.6%	1.25
Physics M	38190	3.391	72.8%	1359	2.0%	18.2%	5.0%	70.3%	26.2%	0.36	26.4%	0.36

Notes. National sample ratios are calculated from 2004 through 2009 data available at https://research.collegeboard.org/programs/ap/data/archived/. Older siblings are from the 2004 to 2009 cohorts and are matched to at least one younger sibling.

Appendix Table 2. Impact of older sibling receiving a higher AP integer score on the likelihood of having a younger sibling

	N	(1) Has a sibling	(2) Total Siblings	(3) 1 sibling	(4) 2 siblings	(5) 3+ siblings
1/2	751068	0.0021	0.0020	0.0031	0.0003	-0.0013*
		(0.0023)	(0.0055)	(0.0022)	(0.0013)	(0.0006)
Control mean		0.433	0.994	0.328	0.087	0.018
2/3	1223351	-0.0016	-0.0032	-0.0012	-0.0007	0.0002
		(0.0018)	(0.0043)	(0.0017)	(0.0011)	(0.0005)
Control mean		0.462	1.060	0.347	0.097	0.018
3/4	1271393	0.0024	0.0058	0.0014	0.0012	-0.0001
		(0.0018)	(0.0043)	(0.0017)	(0.0011)	(0.0005)
Control mean		0.488	1.122	0.365	0.103	0.020
4/5	854024	-0.0015	-0.0011	-0.0034	0.0018	0.0001
		(0.0021)	(0.0052)	(0.0021)	(0.0013)	(0.0006)
Control mean		0.509	1.172	0.379	0.110	0.021
All stacked (5 point bw)	4084643	0.0007	0.0020	0.0002	0.0007	-0.0001
		(0.0010)	(0.0024)	(0.0009)	(0.0006)	(0.0003)
Control mean		0.475	1.092	0.356	0.100	0.019

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact of a student receiving a higher integer score on an AP exam on the likelihood of observing a younger sibling in the data. All regressions include fixed effects for the older sibling's high school graduation cohort and the exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Regressions on individual thresholds sum to slightly more than the total sample using all thresholds as some individuals overlap the five point boundary; in these cases, individuals are assigned to the higher threshold. Baseline values below regression estimates include all individuals within one point below the threshold.

Appendix Table 3. Covariate balancing tests of receiving a higher AP integer score on younger sibling characteristics

AP Integer Threshold	N	Male	White	Asian	Black	Hispanic	Parent Ed: Less than HS	Parent Ed: HS graduate	Parent Ed: BA or higher	Income < $50K	Income $50K - $100K	Income more than $100K	10th grade PSAT	11th grade PSAT	SAT math	SAT verbal	Joint F-test
1/2	180187	-0.0040	-0.0054	-0.0031	0.0008	0.0017	0.0009	-0.0028	-0.0072	-0.0009	-0.0044	0.0027	0.1281	0.3822	-0.2759	1.6299	0.060
		(0.0047)	(0.0045)	(0.0032)	(0.0023)	(0.0031)	(0.0033)	(0.0041)	(0.0050)	(0.0034)	(0.0040)	(0.0038)	(0.2821)	(0.2842)	(1.0461)	(1.0004)	
Baseline mean		0.488	0.635	0.137	0.063	0.124	0.114	0.182	0.642	0.157	0.233	0.233	142.0	153.1	551.5	525.7	
Baseline SD		0.500	0.481	0.344	0.244	0.329	0.317	0.385	0.480	0.363	0.423	0.423	25.2	26.8	102.1	98.5	
2/3	321483	-0.0044	-0.0017	0.0012	-0.0014	0.0022	-0.0009	-0.0020	0.0068+	0.0019	-0.0024	-0.0023	0.4063+	0.3307	0.9067	1.6790*	0.3341
		(0.0035)	(0.0032)	(0.0023)	(0.0014)	(0.0019)	(0.0020)	(0.0028)	(0.0035)	(0.0022)	(0.0030)	(0.0029)	(0.2122)	(0.2063)	(0.7681)	(0.7413)	
Baseline mean		0.493	0.708	0.126	0.041	0.081	0.073	0.154	0.715	0.108	0.232	0.232	148.6	159.8	573.9	550.1	
Baseline SD		0.500	0.455	0.332	0.198	0.273	0.261	0.361	0.452	0.310	0.422	0.422	24.5	25.9	99.2	94.8	
3/4	364037	-0.0000	-0.0020	0.0022	-0.0014	0.0003	-0.0007	-0.0002	-0.0019	0.0009	-0.0002	0.0010	-0.0812	-0.1159	-1.1959+	-1.1942+	0.6896
		(0.0033)	(0.0029)	(0.0022)	(0.0010)	(0.0015)	(0.0016)	(0.0023)	(0.0030)	(0.0018)	(0.0027)	(0.0029)	(0.1998)	(0.1894)	(0.7070)	(0.6938)	
Baseline mean		0.495	0.750	0.128	0.024	0.057	0.049	0.113	0.787	0.080	0.211	0.211	156.0	167.7	600.7	581.2	
Baseline SD		0.500	0.433	0.334	0.153	0.232	0.216	0.316	0.410	0.271	0.408	0.408	24.9	26.1	97.8	95.8	
4/5	258243	-0.0010	-0.0008	0.0012	0.0011	-0.0018	0.0008	0.0007	-0.0016	0.0028	0.0012	0.0014	-0.1257	-0.0676	-0.0645	-0.6874	0.8671
		(0.0039)	(0.0034)	(0.0028)	(0.0009)	(0.0016)	(0.0015)	(0.0023)	(0.0032)	(0.0019)	(0.0031)	(0.0035)	(0.2378)	(0.2217)	(0.8190)	(0.8143)	
Baseline mean		0.494	0.754	0.147	0.014	0.042	0.032	0.077	0.839	0.061	0.189	0.189	163.1	175.0	625.5	607.3	
Baseline SD		0.500	0.431	0.354	0.117	0.202	0.175	0.266	0.368	0.240	0.392	0.392	25.0	26.3	97.1	96.6	

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on the demographic and academic characteristics of the younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance.

Appendix Table 4. Impact of receiving a higher AP integer score on younger sibling outcomes, robustness tests (2/3 threshold)

	(1)	(2)	(3)	(4)	(5)	(6)
Functional Form	Linear	Quad	Linear	Quad	Linear	Quad
Kernel	Rect	Rect	Rect	Rect	Tri	Tri
Covariates	N	N	Y	Y	N	N
Took identical AP exam						
BW=2	0.0205**	0.0145+	0.0181**	0.0150*	0.0180**	0.0097
	(0.0052)	(0.0078)	(0.0049)	(0.0074)	(0.0058)	(0.0084)
BW=4	0.0236**	0.0184**	0.0206**	0.0180**	0.0219**	0.0167**
	(0.0037)	(0.0055)	(0.0035)	(0.0052)	(0.0041)	(0.0059)
BW=6	0.0219**	0.0213**	0.0190**	0.0195**	0.0225**	0.0203**
	(0.0030)	(0.0045)	(0.0029)	(0.0042)	(0.0033)	(0.0048)
BW=8	0.0214**	0.0213**	0.0193**	0.0190**	0.0222**	0.0214**
	(0.0026)	(0.0039)	(0.0025)	(0.0037)	(0.0029)	(0.0042)
Total AP exams						
BW=2	0.0581+	0.0153	0.0323	0.0278	0.0426	0.0144
	(0.0330)	(0.0492)	(0.0264)	(0.0392)	(0.0362)	(0.0525)
BW=4	0.0751**	0.0331	0.0422*	0.0287	0.0605*	0.0331
	(0.0233)	(0.0349)	(0.0187)	(0.0279)	(0.0256)	(0.0373)
BW=6	0.0739**	0.0559*	0.0414**	0.0363	0.0704**	0.0494
	(0.0191)	(0.0285)	(0.0153)	(0.0228)	(0.0209)	(0.0305)
BW=8	0.0748**	0.0594*	0.0509**	0.0340+	0.0729**	0.0589*
	(0.0167)	(0.0248)	(0.0133)	(0.0199)	(0.0182)	(0.0265)
Took AP (extensive margin)						
BW=2	0.0097+	-0.0066	0.0064	-0.0061	0.0037	-0.0080
	(0.0051)	(0.0076)	(0.0045)	(0.0067)	(0.0056)	(0.0081)
BW=4	0.0103**	0.0021	0.0063*	0.0011	0.0076+	0.0024
	(0.0036)	(0.0054)	(0.0032)	(0.0048)	(0.0039)	(0.0057)
BW=6	0.0102**	0.0071	0.0063*	0.0047	0.0095**	0.0056
	(0.0029)	(0.0044)	(0.0026)	(0.0039)	(0.0032)	(0.0047)
BW=8	0.0105**	0.0075*	0.0078**	0.0044	0.0098**	0.0075+
	(0.0026)	(0.0038)	(0.0023)	(0.0034)	(0.0028)	(0.0041)
Total AP (intensive margin)						
BW=2	0.0307	0.0641	0.0084	0.0493	0.0423	0.0695
	(0.0357)	(0.0531)	(0.0309)	(0.0459)	(0.0392)	(0.0567)
BW=4	0.0496*	0.0406	0.0264	0.0264	0.0445	0.0379
	(0.0252)	(0.0377)	(0.0218)	(0.0326)	(0.0277)	(0.0403)
BW=6	0.0480*	0.0426	0.0249	0.0242	0.0474*	0.0420
	(0.0207)	(0.0308)	(0.0179)	(0.0266)	(0.0227)	(0.0330)
BW=8	0.0470**	0.0445+	0.0328*	0.0219	0.0485*	0.0440
	(0.0181)	(0.0268)	(0.0156)	(0.0232)	(0.0197)	(0.0286)

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance.

Appendix Table 5. Impact of receiving a higher AP integer score on younger sibling outcomes, optimal bandwidth results

	(1)	(2)	(3)	(4)
	Took identical AP exam	Total AP exams	Took AP (extensive margin)	Total AP (intensive margin)
1/2 threshold				
Main estimate	0.0039	0.0246	0.0030	0.0234
	(0.0028)	(0.0192)	(0.0030)	(0.0234)
Bias-corrected estimate	0.0045	0.0292	0.0031	0.0295
	(0.0033)	(0.0227)	(0.0035)	(0.0275)
Optimal bandwidth	13.2	11.8	13.9	10.4
N	475646	424618	500820	249798
2/3 threshold				
Main estimate	0.0201**	0.0774**	0.0098**	0.0426**
	(0.0025)	(0.0159)	(0.0023)	(0.0145)
Bias-corrected estimate	0.0210**	0.0844**	0.0108**	0.0479**
	(0.0028)	(0.0176)	(0.0026)	(0.0166)
Optimal bandwidth	10.8	10.5	11.5	14.8
N	683467	661350	721100	658026
3/4 threshold				
Main estimate	0.0142**	0.0264*	0.0042*	0.0097
	(0.0021)	(0.0124)	(0.0017)	(0.0128)
Bias-corrected estimate	0.0151**	0.0219	0.0041*	0.0055
	(0.0023)	(0.0141)	(0.0020)	(0.0147)
Optimal bandwidth	13.9	17.0	17.6	17.1
N	964200	1151985	1183292	879043
4/5 threshold				
Main estimate	0.0095**	0.0006	-0.0007	0.0035
	(0.0028)	(0.0196)	(0.0023)	(0.0187)
Bias-corrected estimate	0.0102**	-0.0016	-0.0011	0.0021
	(0.0033)	(0.0234)	(0.0027)	(0.0223)
Optimal bandwidth	11.3	10.7	10.9	11.7
N	570772	540214	549473	477988

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Regressions on individual thresholds sum to slightly more than the total sample using all thresholds as some individuals overlap the five point boundary; in these cases, individuals are assigned to the higher threshold. Baseline values below regression estimates include all individuals within one point below the threshold.

Appendix Table 6. Impact of receiving a higher AP integer score on younger sibling outcomes, additional robustness tests (2/3 threshold)

	N	Took identical AP exam	Total AP exams	Took AP (extensive margin)	Total AP (intensive margin)
Sibling took only one AP	40004	0.0146+	0.0414	0.0180+	-0.0171
		(0.0087)	(0.0397)	(0.0100)	(0.0529)
		0.251	1.45	0.533	2.71
Inverse weighted by number of older sibling observations	321483	0.0219**	0.0931**	0.0127**	0.0615**
		(0.0034)	(0.0210)	(0.0035)	(0.0230)
		0.321	2.93	0.703	4.17
Only used one older sibling exam (random draw)	251798	0.0234**	0.1057**	0.0155**	0.0635*
		(0.0037)	(0.0227)	(0.0037)	(0.0250)
		0.313	2.74	0.687	3.99
Only 2007-2009	179349	0.0190**	0.0835**	0.0091*	0.0604*
		(0.0044)	(0.0284)	(0.0042)	(0.0305)
		0.334	3.06	0.717	4.27

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance.

Appendix Table 7. Impact of receiving a higher AP integer score on younger sibling outcomes, robustness tests by gender

	RD bandwidth			
Female-female	2	4	6	8
Took identical AP exam	0.022**	0.027**	0.027**	0.025**
	(0.006)	(0.004)	(0.004)	(0.003)
Total AP exams	0.097**	0.086**	0.065**	0.059**
	(0.037)	(0.026)	(0.021)	(0.019)
Took AP (extensive margin)	0.009+	0.009*	0.008**	0.008**
	(0.005)	(0.004)	(0.003)	(0.003)
Total AP (intensive margin)	0.079*	0.062*	0.039+	0.033+
	(0.038)	(0.027)	(0.022)	(0.019)
Female-male				
Took identical AP exam	0.006	0.012**	0.010**	0.009**
	(0.006)	(0.004)	(0.004)	(0.003)
Total AP exams	0.019	0.060*	0.059*	0.049*
	(0.041)	(0.029)	(0.024)	(0.021)
Took AP (extensive margin)	0.007	0.008*	0.005	0.008**
	(0.006)	(0.004)	(0.003)	(0.003)
Total AP (intensive margin)	-0.014	0.038	0.057*	0.020
	(0.045)	(0.032)	(0.026)	(0.023)
Male-female				
Took identical AP exam	0.016*	0.015**	0.015**	0.016**
	(0.006)	(0.004)	(0.004)	(0.003)
Total AP exams	0.028	0.009	0.002	0.010
	(0.040)	(0.029)	(0.024)	(0.021)
Took AP (extensive margin)	-0.002	-0.002	-0.000	0.001
	(0.006)	(0.004)	(0.003)	(0.003)
Total AP (intensive margin)	0.048	0.019	0.003	0.005
	(0.041)	(0.029)	(0.024)	(0.021)
Male-male				
Took identical AP exam	0.011	0.009*	0.010*	0.010**
	(0.006)	(0.005)	(0.004)	(0.003)
Total AP exams	-0.087*	-0.054+	-0.036	-0.019
	(0.044)	(0.031)	(0.026)	(0.022)
Took AP (extensive margin)	-0.000	-0.002	-0.001	-0.000
	(0.006)	(0.004)	(0.003)	(0.003)
Total AP (intensive margin)	-0.107*	-0.054+	-0.038	-0.023
	(0.046)	(0.033)	(0.027)	(0.023)

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Sample sizes for bandwidths of 4, 6, 8, and 10 include 287,833, 430,102, 570,040, and 7009,641 observations, respectively.

Appendix Table 8. Impact of receiving a higher AP integer score on younger sibling outcomes, by gender

Oldest Sibling	Younger sibling		Took identical AP exam	Total AP exams	Took AP (extensive margin)	Total AP (intensive margin)
F		Main estimate	-0.0010	0.0772*	0.0065	0.0565*
			(0.0056)	(0.0321)	(0.0049)	(0.0280)
		Bias-corrected estimate	-0.0036	0.0794	0.0059	0.0587
			(0.0064)	(0.0392)	(0.0060)	(0.0340)
		Optimal bandwidth	1.1	1.5	1.3	2.2
		N	135201	172780	153958	188485
M		Main estimate	0.0031	-0.0105	-0.0035	0.0018
			(0.0052)	(0.0291)	(0.0052)	(0.0358)
		Bias-corrected estimate	0.0015	-0.0128	-0.005	0.0092
			(0.0062)	(0.0351)	(0.0062)	(0.0427)
		Optimal bandwidth	1.5	2.1	1.2	1.5
		N	157675	219940	125838	118974
	F	Main estimate	-0.0022	0.0628*	-0.0020	0.0795*
			(0.0060)	(0.0259)	(0.0048)	(0.0324)
		Bias-corrected estimate	-0.0053	0.0658	-0.0037	0.0864
			(0.0068)	(0.0314)	(0.0056)	(0.0386)
		Optimal bandwidth	1.0	2.2	1.3	1.5
		N	118034	254977	145884	129763
	M	Main estimate	0.0020	0.0150	0.0063	-0.0284
			(0.0053)	(0.0353)	(0.0052)	(0.0320)
		Bias-corrected estimate	0.0008	0.0218	0.0065	-0.0372
			(0.0063)	(0.0428)	(0.0064)	(0.0378)
		Optimal bandwidth	1.4	1.4	1.3	2.0
		N	151671	155524	142176	159506
PAIRS						
F	F	Main estimate	-0.0032	0.0971*	0.0030	0.1219**
			(0.0079)	(0.0377)	(0.0063)	(0.0460)
		Bias-corrected estimate	-0.0071	0.1076	0.0012	0.1367
			(0.0089)	(0.0448)	(0.0075)	(0.0538)
		Optimal bandwidth	1.1	1.9	1.4	1.3
		N	69795	119445	84479	64630
F	M	Main estimate	0.0023	0.0554	0.0114	0.0074
			(0.0076)	(0.0488)	(0.0072)	(0.0500)
		Bias-corrected estimate	0.0005	0.0549	0.0114	0.0015
			(0.0090)	(0.0592)	(0.0088)	(0.0597)
		Optimal bandwidth	1.3	1.4	1.3	1.6
		N	72847	79663	78011	64872
M	F	Main estimate	0.0043	0.0200	-0.0054	0.0454
			(0.0074)	(0.0401)	(0.0060)	(0.0454)
		Bias-corrected estimate	0.0023	0.0137	-0.007	0.0411
			(0.0088)	(0.0478)	(0.0071)	(0.0542)
		Optimal bandwidth	1.4	2.1	1.8	1.7
		N	73721	106409	90743	66808
M	M	Main estimate	0.0036	-0.0424	0.0002	-0.0404
			(0.0062)	(0.0450)	(0.0072)	(0.0532)
		Bias-corrected estimate	0.0027	-0.0403	-0.0008	-0.0253
			(0.0075)	(0.0543)	(0.0086)	(0.0623)
		Optimal bandwidth	2.2	1.9	1.4	1.5
		N	112156	99891	71257	58344

observes their older sibling receive a higher integer score on an AP exam. All regressions combine estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance.

Appendix Table 9. Impact of receiving a higher AP integer score on younger sibling outcomes, by sibling gender at individual thresholds

		Individual Thresholds											
Oldest Sibling	Younger sibling	Took identical AP exam			Total AP exams			Took AP (extensive margin)			Total AP (intensive margin)		
		2/3	3/4	4/5	2/3	3/4	4/5	2/3	3/4	4/5	2/3	3/4	4/5
F	F	0.0286**	0.0276**	0.0180*	0.0762*	0.0502	0.0622	0.0140*	0.0055	-0.0035	0.0244	0.0324	0.0972*
		(0.0062)	(0.0062)	(0.0077)	(0.0371)	(0.0376)	(0.0488)	(0.0057)	(0.0051)	(0.0057)	(0.0390)	(0.0378)	(0.0476)
Baseline means		0.354	0.421	0.472	2.98	3.51	4.10	0.731	0.795	0.839	4.07	4.42	4.89
F	M	0.0197**	0.0081	0.0066	0.1638**	-0.0016	0.0505	0.0204**	0.0020	-0.0058	0.1250**	-0.0101	0.0959+
		(0.0061)	(0.0062)	(0.0078)	(0.0405)	(0.0418)	(0.0548)	(0.0065)	(0.0059)	(0.0067)	(0.0468)	(0.0452)	(0.0562)
Baseline means		0.287	0.345	0.391	2.62	3.24	3.83	0.637	0.712	0.770	4.10	4.55	4.98
M	F	0.0217**	0.0147*	0.0071	0.0656	0.0177	-0.0672	0.0062	0.0003	-0.0092	0.0520	0.0213	-0.0351
		(0.0068)	(0.0064)	(0.0073)	(0.0431)	(0.0413)	(0.0487)	(0.0065)	(0.0056)	(0.0059)	(0.0452)	(0.0418)	(0.0481)
Baseline means		0.324	0.375	0.392	3.03	3.47	3.99	0.741	0.789	0.824	4.09	4.40	4.85
M	M	0.0162*	0.0083	0.0025	0.0107	-0.0572	-0.0847	0.0002	-0.0034	-0.0007	0.0205	-0.0497	-0.0966+
		(0.0070)	(0.0065)	(0.0076)	(0.0463)	(0.0439)	(0.0536)	(0.0069)	(0.0058)	(0.0062)	(0.0509)	(0.0456)	(0.0535)
Baseline means		0.326	0.384	0.439	3.03	3.56	4.18	0.701	0.756	0.799	4.32	4.71	5.24

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance.

Appendix Table 10. Impact of receiving a higher AP integer score on younger sibling outcomes, by gender and exam type

			Higher integer score on STEM exam					*Higher integer score on non-STEM exam*			
				Total AP exams					Total AP exams		
Oldest Sibling	Younger sibling	N	Took identical AP exam	All	STEM	Non-STEM	N	Took identical AP exam	All	STEM	Non-STEM
--	--	334035	0.016**	0.010	0.002	0.007	174324	0.017**	0.050**	0.015*	0.032**
			(0.003)	(0.021)	(0.010)	(0.014)		(0.003)	(0.016)	(0.007)	(0.011)
			0.283	3.27	1.26	1.96		0.400	3.32	1.04	2.23
F	F	82509	0.034**	0.118**	0.047*	0.066*	163695	0.023**	0.057*	0.008	0.042*
			(0.006)	(0.042)	(0.019)	(0.029)		(0.005)	(0.029)	(0.012)	(0.020)
			0.290	3.34	1.16	2.12		0.439	3.34	0.93	2.35
F	M	77530	0.014*	0.057	0.017	0.042	135302	0.012**	0.098**	0.038**	0.057**
			(0.007)	(0.046)	(0.023)	(0.029)		(0.005)	(0.032)	(0.015)	(0.021)
			0.290	3.15	1.34	1.78		0.339	2.99	1.06	1.91
M	F	86877	0.008	-0.075+	-0.010	-0.063*	135774	0.020**	0.072*	0.015	0.056*
			(0.006)	(0.040)	(0.018)	(0.028)		(0.005)	(0.033)	(0.014)	(0.023)
			0.249	3.26	1.12	2.07		0.415	3.47	1.00	2.40
M	M	87119	0.010	-0.046	-0.035	-0.014	609095	0.009+	-0.032	-0.004	-0.030
			(0.006)	(0.043)	(0.022)	(0.028)		(0.005)	(0.036)	(0.017)	(0.024)
			0.304	3.31	1.42	1.87		0.409	3.54	1.22	2.28

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions combine estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Baseline values below regression estimates include all individuals within one point below the threshold.

Appendix Table 11. Impact of receiving a higher AP integer score on younger sibling outcomes, additional heterogeneous effects

	N	Took identical AP exam	Total AP exams	Took AP (extensive margin)	Total AP (intensive margin)
Ethnicity					
White	695902	0.0152**	0.0265+	0.0043*	0.0143
		(0.0023)	(0.0140)	(0.0021)	(0.0146)
		0.349	3.07	0.730	4.20
Asian	125138	0.0171**	0.0147	-0.0046	0.0497
		(0.0055)	(0.0396)	(0.0038)	(0.0382)
		0.449	4.86	0.862	5.64
URM	84596	0.0164**	0.0336	0.0041	0.0257
		(0.0064)	(0.0437)	(0.0062)	(0.0480)
		0.303	2.92	0.679	4.30
Parent education					
Parent BA or more	605782	0.0152**	0.0205	0.0015	0.0193
		(0.0025)	(0.0162)	(0.0020)	(0.0162)
		0.399	3.79	0.803	4.72
Parent Less than BA	170362	0.0155**	0.0345	0.0030	0.0344
		(0.0045)	(0.0291)	(0.0045)	(0.0321)
		0.305	2.78	0.677	4.11
Sibling age gap					
1 year	106922	0.0162**	0.0161	0.0000	0.0221
		(0.0058)	(0.0364)	(0.0053)	(0.0383)
		0.356	3.03	0.719	4.22
2 years	303710	0.0159**	0.0324	0.0040	0.0235
		(0.0035)	(0.0220)	(0.0031)	(0.0228)
		0.368	3.25	0.743	4.37
3 years	257567	0.0137**	0.0177	0.0036	0.0074
		(0.0038)	(0.0241)	(0.0033)	(0.0249)
		0.361	3.33	0.746	4.46
4 years	141236	0.0238**	0.0350	0.0010	0.0457
		(0.0051)	(0.0336)	(0.0045)	(0.0349)
		0.348	3.39	0.744	4.56
5 or more years	133695	0.0114*	0.0534	0.0057	0.0338
		(0.0051)	(0.0362)	(0.0047)	(0.0376)
		0.334	3.48	0.741	4.70

Notes. + p<0.1, * p<0.05, ** p<0.01. Each regression estimates the impact on AP participation for a younger sibling who observes their older sibling receive a higher integer score on an AP exam. All regressions combine estimates from the 2/3, 3/4, and 4/5 AP integer thresholds, and include fixed effects for the older sibling's high school graduation cohort and the AP exam type interacted with the year the exam was taken, using a bandwidth of five points on the underlying continuous measure of exam performance. Baseline values below regression estimates include all individuals within one point below the threshold.