

DISCUSSION PAPER SERIES

IZA DP No. 12314

**Girls, Boys, and High Achievers**

Angela Cools  
Raquel Fernández  
Eleonora Patacchini

APRIL 2019

## DISCUSSION PAPER SERIES

IZA DP No. 12314

# Girls, Boys, and High Achievers

**Angela Cools**

*Cornell University*

**Raquel Fernández**

*NYU, NBER, CEPR, ESOP and IZA*

**Eleonora Patacchini**

*Cornell University, EIEF, CEPR and IZA*

APRIL 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

### Girls, Boys, and High Achievers

This paper studies the effect of exposure to female and male “high-achievers” in high school on the long-run educational outcomes of their peers. Using data from a recent cohort of students in the United States, we identify a causal effect by exploiting quasi-random variation in the exposure of students to peers with highly-educated parents across cohorts within a school. We find that greater exposure to “high-achieving” boys, as proxied by their parents’ education, decreases the likelihood that girls go on to complete a bachelor’s degree, substituting the latter with junior college degrees. It also affects negatively their math and science grades and, in the long term, decreases labor force participation and increases fertility. We explore possible mechanisms and find that greater exposure leads to lower self-confidence and aspirations and to more risky behavior (including having a child before age 18). The girls most strongly affected are those in the bottom half of the ability distribution (as measured by the Peabody Picture Vocabulary Test), those with at least one college-educated parent, and those attending a school in the upper half of the socioeconomic distribution. The effects are quantitatively important: an increase of one standard deviation in the percent of “high-achieving” boys decreases the probability of obtaining a bachelor’s degree from 2.2-4.5 percentage points, depending on the group. Greater exposure to “high-achieving” girls, on the other hand, increases bachelor’s degree attainment for girls in the lower half of the ability distribution, those without a college-educated parent, and those attending a school in the upper half of the socio-economic distribution. The effect of “high-achievers” on male outcomes is markedly different: boys are unaffected by “high-achievers” of either gender.

**JEL Classification:** I21, J16

**Keywords:** gender, education, cohort study, high achievers, peers

**Corresponding author:**

Eleonora Patacchini  
Department of Economics  
Cornell University  
Ithaca, NY 14853  
USA

E-mail: [ep454@cornell.edu](mailto:ep454@cornell.edu)

# 1 Introduction

The gender composition of a class, or of a group of competitors, or of a team has been shown to affect both individual and group outcomes. These findings point to potentially important consequences for issues ranging from effective teaching, to the optimal structuring of teams, or to how best design evaluations in a variety of environments. In this paper we attempt to move beyond the question of gender per se and instead focus on investigating a particular characteristic: “high-achievers” of a given gender, which is the term we use to refer to students with very highly-educated parents. Does greater exposure to “high-achievers” of the same or different gender matter? To whom does it matter? And why?

We investigate these questions in the context of high-school education using the National Longitudinal Survey of Adolescent to Adult Health (Add Health) which was designed to be a nationally representative sample of students in grades 7-12 in the US. We make use of a predetermined student characteristic – whether at least one of their parents has some post-college education – to proxy for a bundle of student characteristics. Aggregating this number across students at the grade level by gender allows us to use plausibly exogenous variation across grades within the same school in the proportions of “high achievers” of each gender. Our main focus is on the longer-run education effects of this variation. We find a very strong asymmetric gender effect: the proportion of “high-achieving” boys has a statistically and economically significant negative effect on the probability that girls will end up with a bachelor’s degree some 14 years later. There is no similar asymmetric gender effect in the proportion of “high-achieving” girls: a greater proportion of these does not affect outcomes for either gender and, furthermore, boys are not affected by the proportion of male “high-achievers.” These results are robust to a wide variety of controls and alternative specifications, including the proportion of females in the class, the rank of the student, and the proportion of students of different races/ethnicities.

We investigate potential heterogeneity and non-linear effects. Performing various cuts of the data, we find that the negative effect of “high-achieving” boys on girls is concentrated in the lower half of the ability distribution (as measured by a student’s Peabody Picture Vocabulary Test score), among those with a (at least) college-educated parent, and in the upper half of the socio-economic distribution of schools (as measured by the fraction of students that are performing at or above grade level). We show that girls exposed to a higher proportion of boys with highly-educated parents tend to have a lower math and science grades in high school and to substitute away from a four-year college degree into a two-year college. Furthermore, they have lower labor force participation and higher fertility by the ages of 26-32.

We are especially interested in understanding the mechanisms that drive the asymmetric

gender results. Using questions in Add Health administered in the baseline year, we show that a larger proportion of “high-achieving” boys is associated with lower self-confidence/ambition in girls, an increase in their risky behavior, and a higher chance of becoming a mother before age 18. The data does not permit us to identify the exact mechanism by which this occurs. Although we call these students (and their parents) “high achievers” a complex mix of their characteristics, and responses these, could be responsible for the outcomes. For example, it could be that teachers pay less attention to girls when faced with boys who perform well in school. Alternatively, it could be that girls are more likely to feel discouraged in the face of competition from boys. Nor can we rule out that their parents interact differently with schools, although this must occur in such a way that is detrimental to girls. Thus, although we cannot isolate different potential mechanisms, we do provide suggestive evidence regarding the pathways at work as they affect girls’ propensity to engage in risky behavior and reduce their aspirations.

Our analysis relies on variation in the proportions of boys and girls that are “high-achievers” across grades within the same school. As the baseline grade-level data is a “snapshot” of a school (grades) in 1994, we also conduct our analysis with an alternative definition of the main variable – the number of high achievers – as the proportion may vary in a systematic fashion due to dropouts. We show the results are robust to this alternative specification. In addition to a school-specific time trend, the paper conducts a variety of checks to make sure that the variation obtained is “as good as random.” As in Lavy and Schlosser (2011), we conduct Monte Carlo simulations in which we randomly generate the post-college status of each student’s parents and compare the simulated vs empirical standard deviation in each school. We also show that variation in the key peer explanatory variables is not related, within a school, to a number of important individual characteristics. Lastly, as suggested by Athey and Imbens (2017), we examine the extent to which the results could have been obtained by chance by reassigning to each individual the proportion of “high-achieving” peers from another grade within the same school, keeping all other variables as in the data and then comparing the results of these placebo tests with the estimated treatment coefficients.

Our paper is related to a growing literature on asymmetric gender effects in a variety of contexts. Asymmetric gender effects have been identified, for example, in a series of experiments by Niederle, Segal, and Vesterlund (2013). They show that in addition to female subjects being less likely to enter competitive situations (tournaments) than male subjects, the gender composition of the other competitors matters. Women are markedly more likely to participate when the competition consists solely of other women. Bordalo, Coffman, Gennaioli, and Shleifer (2018) show that, controlling for ability, a female subject’s belief about the probability she answered a question correctly is more affected by the gender stereotype about the category in which a question is asked (e.g., cars and sports vs cooking and art) than a male subject’s. Furthermore, when subjects

play a cooperative game with a partner in which each needs to decide how willing they are to answer for the group, female subjects' beliefs become even more stereotyped if their partner is known to be male, leading them to decrease the probability with which they are answer and reducing the overall performance of their group. Our results suggest that the decline in self-confidence may be accentuated when faced with a signal of individual performance, a conjecture that would be of interest to confirm in the lab.

In the school setting there is also evidence that gender composition matters. The literature in this area, like ours, has mainly relied on quasi-random variation across cohorts or grades within a school to study topics ranging from the effect of immigrant peers in 5th grade on high-school outcomes in Israel (Gould, Lavy, and Paserman, 2009), to gender roles and their intergenerational perpetuation (Olivetti, Patacchini, and Zenou, 2018; Rodríguez-Planas, Sanz-de Galdeano, and Terskaya, 2018) to the long-run educational and labor-market consequences of disruptive peers in elementary school in Florida (Carrell, Hoekstra, and Kuka, 2018).<sup>1</sup> Hoxby (2000b) exploits idiosyncratic variation in gender and race composition of adjacent cohorts in Texas public schools and finds that a greater share of female peers improves reading and mathematics test scores for both genders. Using Israeli data and relying on variations in the proportion of female students across adjacent cohorts within the same school, Lavy and Schlosser (2011) find that a greater proportion of girls is associated with positive high-school outcomes for both sexes. On the other hand, Black, Devereux, and Salvanes (2013), using Norwegian administrative data, find that there are asymmetric gender effects: a larger proportion of females among ninth-grade peers reduces males' long-run educational attainment whereas it decreases women's rates of becoming a teenage mother and increases the likelihood that as adults they work full time and their earnings. As the studies are based in different countries and use different specifications, the difference in results could stem from a variety of sources including short vs longer-run effects.

The papers closest to ours are Mouganie and Wang (2017) and Feld and Zölitz (2018) as they too are concerned with both the gender of a peer and distinguishing its effect by gender as well.<sup>2</sup> Mouganie and Wang (2017) study high-school students in China and

---

<sup>1</sup>See the excellent review of the literature in Handbook of Education chapter by Sacerdote (2011).

<sup>2</sup>How high or low performing students affect various outcomes is the focus of a large literature. It does not, however, distinguish necessarily between male and female peers nor on their differential gender effects. The results of this literature are mixed. See, e.g., Imberman, Kugler, and Sacerdote (2012) that finds that high school students in the top of the student distribution benefit most from the arrival of high-performing peers (Katrina evacuees) and are the hurt by the arrival low-achieving peers whereas Angrist and Lang (2004) find no significant impact of low-performing students from the Metropolitan Council for Educational Opportunity desegregation program in Boston. Carrell, Fullerton, and West (2009) find that higher-ability peers at the US Air Force Academy provide greater positive peer effects for lower-ability students than for middle-ability students. Bifulco, Fletcher, Oh, and Ross (2014) study the effects of the greater exposure to school peers with a college-educated mother, also using Add Health data. Interestingly, they do not find any significant long-run effect on education but this may well be a result of not distinguishing between male and female peers with a college educated mother. Lastly, Fischer (2017) examines the impact of relatively high-achieving peers in an introductory chemistry classes at a large

find that high-performing male peers (defined by their performance on a national exam in mathematics prior to their entry in high school which occurs in 10th grade) reduce women’s likelihood of choosing a science track (relative to an arts track) for the remainder of high school whereas high-performing female peers have the opposite effect. Feld and Zölitz (2018) exploit the random assignment of first-year students within compulsory courses to teaching sections in a Dutch business school. They show that having male peers with higher pre-assignment GPA is associated with men taking more mathematical courses. Women, on the other hand, choose to take fewer mathematical courses and are less likely to choose a mathematically intensive major. Our analysis adds to these findings by showing that greater exposure to “high-achieving” males not only influences women’s fields of study but also their overall educational attainment. It has the advantage not only of being nationally representative but also of showing that these effects are already present in high school and have long-term consequences. Furthermore, and perhaps most importantly, the nature of our data set allows us to explore some of the pathways by which these effects occur (self-confidence and risk) and identify the characteristics of those girls who are most likely to be negatively affected.

Our paper proceeds as follows. In Section 2, we present the data and sample selection. In Sections 3 and 4, we detail the construction of the main variables and our identification strategy. Section 5 is devoted to the main regression analysis. We explore possible pathways for the effects in Section 6. Section 7 explores heterogeneity in the results and examines additional long-term consequences. In Section 8 we perform several robustness checks and conclude in Section 9.

## 2 Data and Sample Selection

This analysis uses data from the National Longitudinal Survey of Adolescent to Adult Health (Add Health).<sup>3</sup> Add Health is a school-based longitudinal survey designed to be nationally representative of students in grades 7-12. It examines students at a representative set of 132 schools in the United States, beginning in the 1994-1995 school year.<sup>4</sup> Add Health contains both in-school and in-home survey components. First, between

---

public university. She finds that being in a chemistry class with more high-ability peers decreases the likelihood that women complete a STEM degree whereas men are not affected. In this case as well, there is no differentiation in the gender of these peers.

<sup>3</sup>This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

<sup>4</sup>To select schools, Add Health used a stratified sampling design. High schools were chosen from strata based on the following characteristics: region, urbanicity, school type, ethnicity, and size. If a school

September 1994 and April 1995, an in-school survey was issued to students in each of the sample schools. Every student in attendance on the school's survey day was asked to complete an in-school questionnaire which included basic questions about the student's demographics (sex, age, race, nativity status) and information about the characteristics including educational attainment of a mother figure (biological mother, stepmother, foster mother, or adoptive mother) and father figure (biological father, stepfather, foster father, or adoptive father) living in the student's household (henceforth called "residential" parents). A total of about 90,000 students completed this in-school questionnaire.

Following the conclusion of the in-school surveys, Add Health randomly selected a subsample of 20,000 students from the roster of the sample schools for more detailed interviews in the in-home sample. Approximately 17 male and 17 female students from each grade level in each school were chosen for the core in-home sample.<sup>5</sup> The core was then supplemented with oversamples for particular populations of interest, defined by ethnicity, presence of siblings in the sample, adoption status, and disability. Interviews with the core and supplemental students took place between April and December 1995 in the students' homes (Wave I).<sup>6</sup> They included questions on parents' background for both residential parents (the mother and father figures living in the same household as the student) and biological parents.

Wave I also include questions on students' academic performance, attitudes, criminal behavior, and other sensitive topics. At the beginning of the Wave I survey, students were also asked to complete an abbreviated version of the Peabody Picture Vocabulary Test (PVT), a test widely used to measure verbal ability.<sup>7</sup> Interviewers also issued a survey to the student's residential parent, asking questions about attitudes and family income, among other topics.<sup>8</sup> Those individuals selected for the Wave I in-home sample were re-interviewed in 1996 (Wave II), 2001-2002 (Wave III), and 2008 (Wave IV). In these follow-up interviews, individuals were asked questions about their living situation, health behaviors, daily activities, and, importantly, level of educational attainment to date.

---

refused to participate in the survey, another school from the same stratum was selected. Participating high schools then assisted in the identification of feeder schools, typically a middle school, whose students tend to attend the sample high schools.

<sup>5</sup>At 2 large schools and 14 small schools, in-home interviews were administered to all students to achieve a saturated sample.

<sup>6</sup>The majority (92 percent) of Wave I interviews for individuals in our sample took place between April 1995 and August 1995. For these students, Wave I responses and information on peer characteristics are from the same school year. For those interviewed between September and December, Wave I responses are from the following school year. We include, therefore, an indicator for this in any regression with a Wave I outcomes.

<sup>7</sup>The Add Health PVT is a condensed version of the Peabody Picture Vocabulary Test-Revised (PPVT), a standard assessment of verbal ability used in the United States. Scores are standardized by age to a mean of 100 and standard deviation of 15 for each age group, and neither the student nor the interviewer is made aware of the results of the test. For more information, see <http://www.cpc.unc.edu/projects/addhealth/design/wave1>.

<sup>8</sup>The interviewers attempted to interview the student's resident mother. If unavailable, they interviewed another adult in the household. Overall, 93 percent of parent interviews took place with a female parent.



Our analysis uses both Wave I (in-school and in-home survey) and Wave IV information. All the information on school peers is obtained from the in-school survey. In addition to information on educational attainment, our data also provides rich information on behaviors and perceptions in adolescence, which enable us to inspect several mechanisms for our results. Throughout the analysis we use Wave I in-home information on residential rather than biological parents for consistency with the in-school sample (which only collects information on residential parents).

After dropping students who were not followed through Wave IV and those that cannot be matched with peer characteristics we also eliminated particular grades/schools (e.g., an all-male school, 7th and 8th graders from a school that doubles in size between 8th and 9th grade) the final sample consists of 10,853 students (5899 females and 4954 males) and 118 schools.<sup>9</sup> Summary statistics for the Wave IV-weighted sample are reported by sex in Table 1. On average, the students are almost 16 years old in July of 1995. About two-thirds of our sample is Non-Hispanic White, 15-17 percent is Black, 10 percent is Hispanic/Latino, and the remainder is Asian or other races. By the time of the Wave IV survey (2008), the vast majority of students in our sample (94 percent of females and 91 percent of males) have achieved a high school diploma or GED, and about one third (35 percent of females and 28 percent of males) have completed a bachelor’s degree. Data from the 2008 American Community Survey (ACS) that is re-weighted to match the age distribution of the Add Health sample is also presented in Table 1.<sup>10</sup> As shown, the Add Health population is broadly similar to the U.S. population as calculated from the ACS.

### 3 The Main Variables

The objective of this paper is to study the impact of “high-achieving” girls and boys on the long-run educational outcomes of their peers. In order to do so, we require a variable that is plausibly exogenous to a student’s experience. This prevents one from using grades as an outcomes as, for example, a high-ability student may encourage her peers to study harder and thus raise their grades. In this case of reverse causality, an association between peers’ achievement and an individual’s long-run outcomes may be due to the individual’s own ability. We therefore turn to a measure that is determined before individuals meet and interact with their fellow students at school: residential parents’ education.

---

<sup>9</sup>See the Appendix for the exact details. The robustness section examines whether attrition is a function of the main variables introduced in the next section and concludes that this is not a problem.

<sup>10</sup>We exclude those from the ACS sample who immigrated to the United States after 1994 as immigrants since 1994 would not have been in the Add Health sample.

### 3.1 Defining “High Achievers”

Parents’ educational attainment is usually determined before students enter school and the literature finds a strong, positive relationship between an individual’s academic achievement and the educational attainment of his or her parents.<sup>11</sup> We next document this link for individuals in our sample in two ways: via a student’s GPA and via a student’s ability as measured by their PVT score. It is important to reiterate here that parental education is not only a proxy for a student’s achievement but is also correlated with a bundle of other characteristics. Thus, we cannot disentangle which features associated with parental education are playing a critical role. Nonetheless, we think it is useful to document the correlation between parental education and student achievement.

To measure achievement using a student’s GPA, we average a student’s self-reported grades in the four subjects that are asked (English, History, Mathematics, and Science) obtained during the Wave I interview.<sup>12</sup> Table 2 shows the results of an OLS regression in which the dependent variable is GPA and the main independent variables are indicators of residential parental education.<sup>13</sup> A residential parent’s education is coded as less than high school, high school graduate (including a GED), some college if there is any education past high school that does not result in college graduation, college graduate, and post college if “any professional training beyond a four-year college or university.” Students who do not know or refuse to answer are included via a dummy variable for the presence of a missing value. If there is no residential mother (equivalently father), all education dummies are given a value of zero and we include a dummy variable for “mother (equivalently father) not in household.”

The first four columns of Table 2 are for girls; the last four are equivalent specifications for boys. Columns (1)-(3) (respectively, (5)-(7) for males) include increasing number of individual and family controls – first only residential parental education categories (including a category for missing), then age, race, and family income, and lastly the

---

<sup>11</sup>See, e.g., Davis-Kean (2005); Reardon (2011); and van Tetering, de Groot, and Jolles (2018).

<sup>12</sup>For each subject, the student was asked to report the following grade categories: A, B, C, D or lower, subject was not graded that way, or did not take this subject. We set A=4.0, B=3.0, C=2.0, D or lower=1.0 and averaged across the four subjects. If the student reported that the subject was not graded in this manner or did not take this subject (approximately 10 percent of the students), we do not include that subject in the calculation of the student’s GPA.

<sup>13</sup>Information on parental education is from the student’s responses in the Wave I survey. Students were asked, for each residential parent, to select how far the parent went in their education: never went to school; eighth grade or less; more than eighth grade, but did not graduate from high school; went to a business, trade, or vocational school instead of high school; high school graduate; completed a GED; went to a business, trade, or vocational school after high school; went to college, but did not graduate; graduated from a college or university; professional training beyond a four-year college or university; doesn’t know what level; doesn’t know if he/she went to school.

student's PVT score as a control for individual ability.<sup>14,15</sup> All specifications include grade and school fixed-effects to control for different grading philosophies across schools and also across grades. Lastly, specification (4) (respectively, (8) for males) includes a school-specific linear time trend to account for trends in grading practices or student composition at the school level.

As shown, having highly-educated residential parents is associated with a higher GPA for both girls and boys. For girls, a mother with a post-college education – defined as any education past college graduation – is associated with a GPA that is about 0.3-0.5 points higher relative to the baseline group of a mother without a high school degree. A father with a post-college education is associated with a GPA that is 0.3-0.4 points higher relative to the baseline group of a father without a high school degree. The results are similar for boys: a mother with a post-college education is associated with a GPA that is 0.2-0.4 points higher than the baseline group whereas a father with a post-college education is associated with a GPA that is about 0.4-0.5 points higher than the baseline group. These are sizable effects: *ceteris paribus*, they increase a girl's GPA from a mean of 2.9 to 3.2-3.4 and a boy's GPA from a mean of 2.7 to 2.9-3.2.

Although we have included several individual controls in the regressions of student achievement, note that if what one is interested is how other students react to high achievers, it is not clear that these controls are relevant. That is, other students will not care necessarily if a student is a high-achiever after controlling for, say, family income. Nonetheless, it is of interest to note that the correlation between parental education and achievement exists both with and without controls.

We next consider whether a parent with a post-college education predicts achievement more strongly than any other levels of parental education. For both girls and boys, a post-college mother is quantitatively more important than simply college. A post-college rather than only-college father is also highly predictive for boys but not for girls. For the latter, a college versus a post-college father is associated with roughly the same increase in GPA. To show this rigorously, we perform F-tests for equality between the coefficients on parents with post-college education and parents with only a college education. We focus on two sets of coefficients: mother with college versus mother with post-college, and father with college versus father with post-college. The equality p values are shown

---

<sup>14</sup>Using information on the student's month and year of birth, we create a variable for age in months that reflects each student's age in July 1995. Race controls include dummies for Black, Latino, Asian, and other races. Students are classified as Latino if they report any Hispanic/Latino ethnicity. They are classified as other races if they report more than one race (White, Black, Asian) or being of another race outside of these three categories.

<sup>15</sup>The Add Health parent survey asks the following question to collect family income information: "About how much total income, before taxes did your family receive in 1994? Include your own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources." The number is given in thousands and thus multiplied by 1000 before inclusion in the regression. If family income is missing, we impute it at the average of that reported by parents of other Wave I in-home survey students at their school and include a dummy for missing family income.

at the bottom of the table. The values are under 0.05 for mother post-college versus mother college, and are under 0.01 for father post college versus father college for boys. Consequently, we define a student as a “high-achiever” if at least one of their parents has a post-college education.<sup>16</sup>

We can repeat this exercise using a larger sample – the in-school sample – but in that case losing the information on household income and the student’s PVT score. In Table 3 we examine the predictive power of parental education on student achievement for this sample. As can be seen from this table, the results are similar to those obtained for the in-home sample: either parent post-college is the quantitatively most important education level for a student’s GPA.

As a final check, in Table 4, we return to our main sample and examine the relationship between a proxy for a student’s ability – their PVT score – and parental education. The results are similar to the ones obtained for GPA. In this case, a girl with a post-college mother is associated with a PVT score that is 7-9 points higher (more than half a standard deviation of PVT scores as these are normalized to have a standard deviation of 15); a boy with a post-college father has a 5-7 point higher PVT score.

### 3.2 Construction of the Main Variables

As discussed in the previous section, we call a student a “high achiever” if at least one parent has a post-college education. Next, we construct a composite measure for the extent to which a student faces or interacts with this type of student. A natural measure to use for this is the *fraction* of students in the same grade who are “high achievers,” i.e., who have residential post-college parents. We use the in-school survey which records the student’s response to the highest level of education attained by their residential father and residential mother and create a dummy variable  $PC_i$  for student  $i$  that takes the value one if either the residential mother or the residential father of student  $i$  has a post-college education, i.e., obtained any education beyond a four-year college degree, and takes the value of 0 otherwise. If a student either does not have a residential father/mother or the information is missing, we impute that parent’s level of education using the other parent’s education. For example, if the residential father’s education is missing, but the residential mother has a high-school education, we impute a value for father post-college by taking the average value of father post-college among students of the same gender and and in the same school who also have a residential mother with a high-school education.<sup>17</sup>

We measure the *fraction* of male and female high achievers separately. Specifically, we define the variables:

---

<sup>16</sup>Not only is this variable a stronger predictor of high achievement than college graduate but its interpretation is also clearer since the latter does not explicitly distinguish between a two and four-year college whereas a parent is only classified as post-college if they have gone beyond a four-year college.

<sup>17</sup>See the Appendix for further details.

$$MF_{igs} = \frac{1}{n} \sum_{j(i)=1}^n PC_{jgs}, \quad FF_{igs} = \frac{1}{q} \sum_{k(i)=1}^q PC_{kgs} \quad (1)$$

where  $j(i) = 1, \dots, n$  indexes student  $i$ 's male peers and  $k(i) = 1, \dots, q$  indexes student  $i$ 's female peers in grade  $g$  and school  $s$ . Thus,  $MF_{igs}$  (respectively,  $FF_{igs}$ ) is the fraction of male peers (respectively, female peers) in the same school and grade as  $i$  with at least one post-college parent. Both  $MF_{igs}$  and  $FF_{igs}$  are the sample moments of the *leave-one-out* distribution of students with a post-college parent belonging to a specific gender, grade, and school. That is, for each student  $i$ , these variables capture the proportion of students of each gender with a post-college parent computed from the school-grade distribution of students by gender after eliminating student  $i$  from the distribution.

## 4 Empirical Model and Identification Strategy

This section presents the benchmark regression model and conducts several tests. In particular, it reports the variation that exists once in the main variables net of fixed effects and school time trends. It performs Monte Carlo simulations and balance tests.

### 4.1 The Regression Model

In Wave IV we observe individuals at ages 26-32, by which point individuals are likely to have completed a substantial portion of their education.<sup>18</sup> Hence, the main long-run outcome that we are interested in studying and that is feasible given the nature of the data is completion of a bachelor's degree. We create a variable for "bachelor's degree" which equals 1 if the individual reports any of the following degrees: bachelor's degree, certificate from a 1- or 2-year post-baccalaureate academic program, master's degree, PhD degree or equivalent (EDD, DrPH, etc.), or professional doctorate (MD, JD, LLB, DDS, etc.). The variable "bachelor's degree" is set equal to 0 if the individual does not report any of these degrees.

We estimate the following regression model:

$$y_{igs,t+1} = \alpha_g + \beta_s + \delta_s g + \phi_1 MF_{igs} + \phi_2 FF_{igs} + \theta X_{i,t} + \gamma Z_{igs,t} + \varepsilon_{i,t+1} \quad (2)$$

where  $i$  denotes a student,  $g$  denotes grade or cohort,  $s$  denotes school, and  $t$  denotes time.  $y_{igs,t+1}$  is a dummy variable taking value 1 if, as of Wave IV ( $t+1$ ), the student has obtained a bachelor's degree.  $\alpha_g$  is a grade fixed effect,  $\beta_s$  is a school fixed effect, and  $\delta_s g$  is a school-specific linear time trend (which is equivalent to a linear trend in grade level).

---

<sup>18</sup>To the extent that younger individuals may still go on to finish college, this is controlled for by the grade fixed effect.

The linear time trend is implemented by creating dummy variables by school that are set equal to the student’s grade if they attend the given school, and 0 otherwise. We also include a vector of controls for individual characteristics,  $X_i$ , and a vector of other peer characteristics  $Z_{igs,t}$ , as measured in Wave I. Finally,  $\varepsilon_{i,t+1}$  are i.i.d., mean 0 innovations.

Our empirical strategy exploits idiosyncratic variation in exposure to high achievers across different cohorts of high-school students within a given school, a common approach in the literature.<sup>19</sup> The grade fixed-effects control for initial differences across cohorts whereas the school fixed-effects control for unobserved differences in average student characteristics across schools as well as for aspects of school quality that are constant across cohorts within a school. The main assumption we make is the usual one: we assume that while parents may make decisions based on overall characteristics of a school, they do not do so based on the specific characteristics of their child’s cohort within the school. Thus, the variation due to differences in cohorts across schools can be treated as quasi-random. To deal with the possibility that the average characteristics of a school may be changing over time/grade, and thus that there may also be changes in selection over time, we include a school linear time trend in all specifications. Thus, the quasi-random variation is obtained from the deviation from this trend, rather than simply from the deviation around the average cohort in the school.

## 4.2 Evidence in Support of the Identification Strategy

Our ability to exploit this identification strategy relies on there being sufficient residual variation in the main variables. Table 5, panel (a), reports variation in the fraction of male (female) “high achievers” (always using the leave-one-out distribution as described earlier) by sex. As shown in the first row, the average of  $MF$  is 0.145 for females and 0.143 for males (that is, about 14.3-14.5 percent of male peers have a post-college parent), and the standard deviation is about 0.10. After removing grade fixed effects and the school time trend, the residual variation is about 0.02, accounting for just under one-fifth of the overall raw variation for both genders. The average of  $FF$  is slightly lower, about 0.12, and the standard deviation is approximately 0.10.<sup>20</sup> The residual variation in  $FF$  of 0.02 accounts for about one-fifth of the overall variation as well.

Next, to test whether the variation we observe in our main variables is “as good as random” we perform Monte Carlo simulations.<sup>21</sup> Specifically, for each student  $i$  in our sample we randomly generate the post-college status of  $i$ ’s parents using a binomial

<sup>19</sup>See, e.g., Hoxby (2000a, 2000b); Angrist and Lang (2004); Gould et al. (2009); Lavy and Schlosser (2011); Lavy, Paserman, and Schlosser (2011); Olivetti et al. (2018); and Merlino, Steinhardt, and Wren-Lewis (2019).

<sup>20</sup>This could be due to girls with more-educated parents having a greater tendency to attend all-girls schools.

<sup>21</sup>See, e.g., Lavy and Schlosser (2011) and Rodríguez-Planas et al. (2018).

distribution function with  $p$  equal to the fraction of students of the same gender as  $i$  who have post-college parents in that student’s school. We then compute the within-school standard deviation, by gender, using the residuals from regressions of  $MF$  and  $FF$  on school fixed effects, grade fixed effects, and school-specific time trends. We repeat this process 1,000 times to obtain an empirical 90 percent confidence interval for the standard deviation in  $MF$  and  $FF$  by gender, for each school.<sup>22</sup>

Figure 1 shows the simulated standard deviation of  $MF$  and  $FF$  for females (panels a and c) and males (panels b and d). The upper and lower edges of the bars represent the 5th and 95th percentiles of the simulated standard deviation. The dot represents the empirical standard deviation. As shown in the figure, about 90 percent of our schools have a standard deviation of  $MF$  and  $FF$  within the 90 percent confidence interval obtained from our simulations for both males and females.<sup>23</sup>

Following Lavy and Schlosser (2011), we next investigate the validity of the identification strategy by examining whether the variation in the main variables is related to the variation in a number of predetermined student characteristics. We consider family income, family social structure (as captured by whether the mother/father lives with the student), the student’s ability (as measured by the Picture Vocabulary Test), race (whether or not the student is Black), and age in months. We run separate regressions with each of these as alternate dependent variables and the main variables as the independent variables, always including grade and school fixed effects and time trends. Note that there is a mechanical negative correlation between girls’  $FF$  (boys’  $MF$ ) and own parent’s post-college status as a result of the leave-one-out strategy. For example, a girl with a post-college parent will have a lower  $FF$  than her peers without a post-college parent since the former’s parents are not included in the grade average of  $FF$ . To eliminate this source of bias, we control for own parent’s post-college status in the regression with girls and  $FF$  (respectively, boys and  $MF$ ). As shown in Table 6, only one of the estimated correlations is significantly different from zero at the five percent level, which is slightly less than what would be expected by chance. This evidence mitigates concerns regarding systematic differences due to sorting across grades within the same school along observable

---

<sup>22</sup> We perform this exercise only for schools with at least 3 grades (71 schools); for those with fewer grades, the variation in our main variables is absorbed by the school fixed effect and its time trend. Specifically,  $MF$  necessarily takes the same value for all girls in the same grade and school; likewise for boys and  $FF$ . As a result, variation in  $MF$  for females and  $FF$  for males will be completely absorbed by the school fixed effect and the time trend in schools with fewer than three grades.  $MF$  for males and  $FF$  for females take on separate values for students in the same grade and school due to the leave-one-out nature of the variable construction.

<sup>23</sup> Specifically, 96 percent (females) and 90 percent (males) of schools have a standard deviation of  $MF$  falling within the estimated 90 percent confidence interval, and 90 percent (females) and 92 percent (males) have a standard deviation of  $FF$  falling within the estimated 90 percent confidence interval.

students' characteristics.<sup>24,25</sup> Taken together, the results in Tables 5 and 6, along with Figure 1 lend support to the identification strategy.

## 5 Results

This section presents the benchmark results. In response to the potential concern that the main results could be driven by dropouts, it then explores an alternative measure of exposure that relies on numbers rather than proportions. Lastly, to assess the likelihood that the results could have occurred by chance, it conducts a simulation exercise.

### 5.1 The Benchmark Results

Table 7 reports the estimation results of the model in equation (2), with the completion of a bachelor's (four-year college) degree as the dependent variable. The results are shown with increasing controls and all specifications include grade fixed-effects, school fixed-effects, and a school time trend.<sup>26</sup> Standard errors are clustered at the school level. The first six columns are for girls; the last six are for boys.

In column (1), we include only grade and school fixed effects and linear time trends. In column (2), we add controls for individual characteristics and parental background. We include age in months, an indicator for foreign born, race, the student's PVT score, parental education (coded as discussed previously), an indicator for mother or father not in the household, and the log of household income. Looking across columns, note that, conditional on student grade and school, age in months is negatively associated with attainment of a bachelor's degree. This may in part reflect students who have been held back as these tend to complete college at lower rates (or students who have skipped grades could be completing at higher rates). An individual's foreign-born status has a positive association with college completion.

As expected, both higher parental education and higher parental income increase the likelihood of completing college, as does a higher PVT score. In column (3) (respectively, column (9) for males) we include the PVT percentile rank of the student in addition to

---

<sup>24</sup>Altonji, Elder, and Taber (2005) suggest that the degree of selection on observables can provide a good indicator of the degree of selection on unobservables. In light of this argument, the evidence of no correlation from Table 6 support the notion that our model specification identifies an exogenous source of variation.

<sup>25</sup>We can also examine whether the residual variations in  $MF$  and  $FF$  are related. To test this, we regress the fraction of females with a post college parent on the fraction of males with a post college parent, as well as grade and school fixed effects, and linear time trends. The correlation between the two variables is 0.12 and statistically insignificant.

<sup>26</sup>All specifications include a dummy for whether the Wave I interview was completed between April and August 1995 (the 1994-1995 school year) or between September and December 1995 (the 1995-1996 school year). Since the in-school survey was completed during the 1994-1995 school year, this dummy accounts for whether students answer the in-home survey with respect to contemporaneous peers or the peer composition in the prior school year.



their PVT score, where the rank is calculated relative to the other students in the in-home sample from the same grade and school as only these students have PVT reported. The PVT rank has been shown in previous literature to affect educational attainment and career aspirations [see Elsner and Isphording (2017, 2018)].<sup>27</sup> As can be seen, the PVT rank is positively and significantly associated with college graduation for women (but not for men which may be a function of the sample size over which rank is calculated). The coefficients on the main variables are unaffected by its inclusion. In column (4) (respectively, column (10) for men), we include the fraction of peers who are female as the literature has tended to find that students perform better when the fraction of boys is smaller (see, e.g., Lavy and Schlosser, 2011) and that long-run educational attainment increases with the fraction of same-gender peers (Black et al., 2013). The results here are in agreement with (Black et al., 2013): a higher fraction of girls increases long-run education outcomes for girls and decreases them for boys. In column (5) (respectively, column (11) for men), we include other peer variables: the fractions foreign born, Black, Asian and Latino.<sup>28</sup> These fractions are calculated in the same “leave-one-out” fashion described earlier for the main variables.

Finally, in column (6) (respectively, column (12) for males), we remove the individual’s PVT score but keep all other control variables. The PVT score is generally viewed as a measure of innate ability and the convention in the literature is to include it in the set of baseline individual controls.<sup>29</sup> However, to the extent that PVT measures content or skills learned in the classroom, it could be an endogenous outcome and should thus be excluded. We thus show the results without the PVT score to allay possible concerns about its endogeneity. Note that once it is omitted, the coefficient on foreign born is insignificant, as would be expected if the test requires proficiency in English.

As can be seen clearly in this table, across specifications there is a great deal of stability in the magnitude of the estimated coefficients on all variables. Across all specifications, a greater fraction of “high-achieving” male peers decreases the likelihood of a girl completing a bachelor’s degree. The effect is sizable: using the most complete specification (column

---

<sup>27</sup>We follow Elsner and Isphording (2017, 2018) and first calculate the absolute rank of each student relative to others in her grade and school (with the worst-performing student having a value of 1) before converting into a percentile to render the measure comparable across samples of different sizes. To convert absolute rank to a percentile rank, the following formula is used:  $\text{Percentile Rank} = (\text{Absolute Rank} - 1) / (\text{Number of Students in the Same Grade and School} - 1)$ . Thus, the percentile rank ranges from 0 (the worst student) to 1 (the best student).

<sup>28</sup>Students report their race and ethnicity in the in-school survey. Individuals are classified as Hispanic if they report their ethnicity as Hispanic/Latino. Individuals are given a race/ethnicity of other if they report multiple races, do not report a race, or report an “other” race (not Black, White, or Asian). There are a number of students who do not report their foreign born status. To construct the fraction of peers who are foreign born, we impute the missing values by using the average fraction foreign born among students of the same race/ethnicity in the same grade and school. Bifulco, Fletcher, and Ross (2011) find that a higher fraction of Black and Hispanic peers is associated with worse reported student-teacher relationships and an increase in disruptive behaviors.

<sup>29</sup>See, e.g., Bifulco et al. (2011); Elsner and Isphording (2017); Olivetti et al. (2018).

5), an increase in  $MF$  by a standard deviation (2.0 percentage points) is associated with a 2.2 percentage point decrease in the probability that these girls go on to complete a four-year college degree by Wave IV.<sup>30</sup> The effect of female high-achievers, on the other hand, while positive is statistically insignificant. The results are radically different for males. The effect of the fraction of “high-achieving” males on males’ college completion, while positive for the most part, is not statistically different from zero. The effect of “high-achieving” female peers again is positive and insignificant.

As shown in columns (6) and (12), excluding the PVT score as an individual control leaves the results unchanged.<sup>31</sup> Since the results do not differ substantially with or without the PVT score, we follow the convention in the literature and focus on specification (5). Furthermore, as shown previously in Table 6, there is no evidence that the PVT score covaries with  $MF$  or  $FF$ .

## 5.2 Alternative Measure of Exposure to “High Achievers”

The results above show that “high-achieving” boys have a negative influence on girls’ long-run education outcomes. This is a strong and intriguing result since, if anything, one might expect their effect to be positive. A potential concern is that the variation in the fraction of high achievers could be driven by variation in the number of dropouts across grades in a school. Dropouts tend to have parents with relatively lower levels of education, and therefore a higher-than-average number of dropouts in a grade would mechanically generate a disproportionately high fraction of “high achievers.”<sup>32</sup> A high value of  $MF$  (or  $FF$ ) could thus reflect a high number of “bad boys” (or “bad girls”) in previous years rather than a high number of “good boys” (or “good girls”). In that case, the correct interpretation of the result would be the *opposite* of the conclusion we offered: it would be that having been previously in a grade with a disproportionately large number of future male dropouts is bad for girls. We deal with this important concern in a variety of ways.

First, to the extent that the dropout rate results in linear trend in the main variables, this issue is taken care of in the school time trend. To the extent that it does not, we can eliminate the ambiguity introduced by relying on proportions and use the *number* of “high achievers” rather than the *fraction* of these. The number of “high achievers” is not affected by variation in the number of dropouts. Thus, we define, for each student  $i$  in grade  $g$  and school  $s$ , the *number* of “high achievers” of a given sex (i.e., the number of

---

<sup>30</sup>Whenever the effect of a standard deviation change in a variable is considered, it is *the standard deviation net of fixed effects and of time trend*.

<sup>31</sup>The sample size increases as there were around 250 observations, by gender, dropped because of missing PVT scores.

<sup>32</sup>For example, among ninth graders in our sample interviewed before September 1995, all those with a post-college parent are in school the following year. In contrast, 96.7% of those without a post-college parent are in school the following year.

students of a given gender with a post-college parent):

$$MN_{igs} = \sum_{j(i)=1}^n PC_{jgs}, \quad FN_{igs} = \sum_{k(i)=1}^q PC_{kgs} \quad (3)$$

where  $j(i) = 1, \dots, j$  indexes student  $i$ 's male peers and  $k(i) = 1, \dots, q$  indexes  $i$ 's female peers. Therefore,  $MN_{igs}$  (equivalently,  $FN_{igs}$ ) is the number of male peers (equivalently, female peers) in the same school and grade as  $i$  who are high achievers. As before, both  $MN_{igs}$  and  $FN_{igs}$  are the sample moments of the leave-one-out distribution of the students with a post-college parent belonging to a specific grade and school. Table table:variation, panel(b) reports variation in the number of male (female) “high achievers” (always using the leave-one-out distribution as described earlier).

To facilitate the interpretation of the coefficients on the main variables, we perform an inverse hyperbolic sine transformation on  $MN$  and  $FN$ . This allows the coefficients  $\phi_1$  and  $\phi_2$  to be interpreted as measuring the effect of a percentage change in  $MN$  or  $FN$ , respectively, and, unlike logs, be defined at zero.<sup>33</sup> We use:

$$y_{igs,t+1} = \alpha_g + \beta_s + \delta_s g + \phi_1 IHS(MN_{igs}) + \phi_2 IHS(FN_{igs}) + \theta X_{i,t} + \gamma Z_{igs,t} + \varepsilon_{i,t+1} \quad (4)$$

where all variables are as defined previously. The identification tests (the Monte Carlo simulation and the check on the balance in other peer variables) for  $MN$  and  $FN$  are in the Appendix and provide evidence that the variation in these variables is random after controlling for fixed effects and time trends.

Table 8 reports the estimation results of regression model (4) using  $MN$  and  $FN$  as the main explanatory variables and the completion of a bachelor's degree as the dependent variable. As noted, we use the inverse hyperbolic sine of  $MN$  and  $FN$  – our main explanatory variables – and consequently of all peer variables such as the number of Asians, Blacks, and Hispanics.<sup>34</sup> The table follows the same format as Table 7: the first six columns report the results for females; the last six columns do the same for males. As before, all columns include grade fixed effects, school fixed effects, and school linear time trends. Standard errors are clustered at the school level.

Analogously to the results in Table 7, we find that “high-achieving” males reduce females' college completion. Across the columns, the effect of  $MN$  on the proportion of girls who graduate from college stays remarkably constant. Different from Table 7, we now find a statistically significant positive effect of the number of “high-achieving” girls

<sup>33</sup>The inverse hyperbolic sine of  $x$  is defined as:  $\log(x + (1 + x^2)^{0.5})$  [see, e.g., Burbidge, Magee, and Robb (1988)].

<sup>34</sup>We use number rather than fraction for the same reason that we turned to numbers to measure “high-achieving” peers, i.e., the concern that the results may be driven by selection into dropping out/being absent. The only exception is for percent female. We kept this as a percent as including the number of girls would transform our main variable into a percentage again.

on the outcomes of other girls. Interestingly, the coefficient on the fraction of girls in the grade becomes insignificant. As before, the effect of own-parent education, household income, and PVT score are positive and significant, as expected, as is (for girls) the PVT rank. For boys, the impacts of both  $MN$  and  $FN$  are positive but statistically insignificant.

The magnitudes implied by the estimated coefficients on  $MN$  for girls are again substantial. On average, a one standard-deviation increase in the number of “high-achieving” boys (2.7) from a mean of 16.4 is associated with about a 1.4 percentage point decrease in women’s college graduation. Since roughly 35 percent of women in our sample complete a bachelor’s degree, a one standard-deviation increase in  $MN$  is associated with about a 4 percent decrease in the proportion of women who attain bachelor’s degrees.

Lastly, in addition to demonstrating that the results are robust to the use of numbers, rather than fractions, of “high achievers” one can also test directly whether the variation in  $MF$  and  $FF$  is driven by dropouts. One way to do this is to examine the correlation between the fraction of boys with post-college parents, net of fixed effects and school time trend, and the number of boys in the grade (again net of fixed effects and time trend). If higher levels of  $MF$  were driven by higher dropout-rates of boys with less-educated parents, we would expect a negative and significant relationship between these two variables. Instead, the correlation is 0.13. The same exercise but for fraction of girls with post-college parents and the number of girls yields a correlation of 0.05. This exercise provides further evidence that the variation in the main variables is not driven by variation in dropouts.

Since the results are robust to using the main variables expressed both as fractions and as numbers and given that there is not a negative correlation between the proportion of “high achievers” of a given gender and the total number of individuals of the same gender, we henceforth restrict the remainder of the analysis to  $MF$  and  $FF$ .

### 5.3 Placebo Tests

We next use a simulation exercise, as described in Athey and Imbens (2017), to study the likelihood that the results obtained could have occurred by chance. For both the fraction and number specifications of the main variables, we calculate the likelihood of obtaining the observed treatment effects by chance by generating randomness in the exposure of individuals to “high-achieving” students. We do this by re-assigning to each individual in the sample the  $MF$  of a random grade within the same school, keeping all other variables (both own and those of one’s peers) at their true levels. We repeat this procedure 1,000 times, and run the fullest specification of Table 7 using  $MF$  and  $FF$  as the key explanatory variables.

The distributions of the estimated coefficients on  $MF$  and  $FF$ , by gender, are shown

in Figure 2. The vertical line in each graph indicates the estimated treatment effect we obtained in Table 7. The share of estimates that is larger in absolute value than the dashed line (actual treatment) represents the randomization-based p-value. As can be seen in the figure, the estimated coefficient of  $MF$  for girls is larger in absolute value than any of the randomization-based estimates, providing evidence that this is unlikely to have occurred by chance. The share of estimates of  $FF$  that is larger in absolute value than the line (actual treatment) is 0.27. For boys, the share of estimates that is larger in absolute value than the estimated coefficient for  $MF$  is 0.66 and of  $FF$  is 0.46.<sup>35</sup>

## 6 Exploring Possible Mechanisms

We next explore mechanisms that may be responsible for our most striking result: the negative effect of “high-achieving” boys on their female peers’ probability of graduating from college. Although we cannot determine the exact mechanism (e.g., do teachers pay more attention to “high-achieving” boys to the detriment of girls? Are girls more discouraged when they face greater competition from “high-achieving” boys? etc.), we are able to explore various pathways that lead to girls achieving a worse long-run education outcome.

### 6.1 Two vs Four-Year College

Having established that girls exposed to greater levels of “high-achieving” boys are less likely to graduate with a bachelor’s degree, we next ask what level of education they attain. A natural possibility is that they obtain a different type of post-secondary degree such as a vocational or associate’s degree. The latter is a two-year post-high-school degree, usually obtained from a community college. These degrees represent a level of education greater than a high-school degree but less than a bachelor’s degree.

In Table 9, column (1) (respectively, column 3 for men), we use the most complete specification of Table 7 but with the dependent variable equal to 1 if the individual has a vocational or associate’s degree as her highest level of education. This implies that the individual reported that they had a “certificate or degree from a 1-, 2-, or 3-year vocational/technical program (after high school)” or an associate’s degree.<sup>36</sup> As shown, girls exposed to “high achievers” are more likely to compete a vocational or associate’s degree, substituting this for a bachelor’s degree basically one for one as the absolute

---

<sup>35</sup>Repeating this exercise for  $MN$  and  $FN$ , the results by gender are shown in Figure 3. The results for  $MN$  for females are similar to those obtained for  $MF$ : the estimated treatment coefficient is larger in absolute value than any of the placebo estimates. For  $FN$ , the estimated treatment coefficient is larger in absolute value than about 96% of the placebo estimates. As expected, the randomization-based estimates of the treatment effects for both  $MN$  and  $FN$  for boys are likely to be obtained by chance.

<sup>36</sup>The variable is coded as 0 if the individual reports a degree from a vocational program lasting less than 1 year.

magnitude is similar and the sign is opposite to that obtained for the completion of a bachelor’s degree. This indicates that “high-achieving” males influence the decision about the type of degree to pursue rather than the decision to pursue any education after high school.

Finally, we can consider the impact of “high achievers” on the probability of high-school completion. In column (2) (respectively, column (4) for males), the dependent variable is set equal to 1 if the individual reports their highest level of education as “high-school graduate” or greater. As shown, there is no impact of either the proportion of “high achievers” on high-school completion for either girls or boys.

## 6.2 GPA

Why should greater exposure to “high-achieving” boys lead girls to switch from a bachelor’s to a junior college degree? One possibility is that greater exposure to these boys may reduce girls’ grades. This may happen mechanically if teachers conform to a fixed grade distribution and, for example, only give out grades of “A” to some fixed percent of the class. In that case, however, we would expect a symmetric effect on boys from greater exposure to “high-achieving” girls. Alternatively, the presence of these boys may direct teachers’ attention away from female students or simply discourage the latter’s efforts, resulting in lower grades.

In column (1) of Table 10 (respectively, column (5) for boys), we repeat the main specification with the student’s average GPA in Wave I as the dependent variable. As shown, “high-achieving” males are associated with a negative but statistically insignificant impact on GPA for girls whereas “high-achieving” females have a positive and statistically insignificant effect. For boys, the signs of these two variables are positive, but they are both insignificant at conventional levels.

Turning to subject-specific grades, we find that “high-achieving” boys have a significant negative effect on girls’ math and science grades, with a one standard-deviation increase in  $MF$  decreasing female math grades by 0.05 points, or 5 percent of a standard deviation. These boys also negatively impact females’ science grades, with a similar magnitude associated with a one-standard deviation increase in their proportion as for math grades.<sup>37</sup> There is no statistically significant effect of either peer variable on boys’ grades in math or science.

The results above are suggestive of findings in Bordalo et al. (2018). Their set of experiments show that, for a given level of difficulty, the greater the average performance gender gap in a domain, the less confident girls are that their answers are correct. Boys, on the other hand, are not affected. In our setting, a similar phenomenon may be taking

---

<sup>37</sup>Students may not be taking all subjects, but we do not find a significant effect of  $MF$  on girls’ probability of taking any of the four subjects.

place. Faced with a greater proportion of “high-performing” boys, girls may become less self-confident about their own ability in traditionally male-dominated fields such as math and science. More generally, these high-school girls may become more discouraged or think themselves less competent which could then affect their actual performance.

Our results also complement those in Feld and Zölitz (2018) and Mouganie and Wang (2017) that find that high-performing male peers reduce females’ completion of mathematics and/or science courses. Feld and Zölitz (2018) study outcomes of random assignment of students to sections in the first year of study in a Dutch business school, finding that women in sections with men with a higher average GPA (as measured prior to the start of the course) tend to choose a mathematical type major less frequently. Being in a section in which women have a high average GPA does not affect either men or women’s choice of major, on the other hand. Mouganie and Wang (2017) study high-school students in China and find that high-performing male peers in tenth grade (defined as those scoring in the top 20 percent in the national high-school-entrance mathematics exam) reduce girls’ likelihood of choosing a science track relative to an arts track for the remainder of high school.

### 6.3 Self Confidence and Aspirations

We next turn to psychological mechanisms such as self-confidence and aspirations/ambition more directly related to college attendance. For this, we can use three questions in Add Health. First, from the question “On a scale of 1 to 5, where 1 is low and 5 is high, how much do you want to go to college?”, we create the variable “Want College”, which equals 1 if the student reports that they want to go to college as a 5 and equals 0 otherwise.<sup>38</sup> Second, from the question “On a scale of 1 to 5, where 1 is low and 5 is high, how likely is it that you will go to college?” we create the variable “College likely” which equals 1 if the student says the likelihood that they go to college is a 5 and equals 0 otherwise.<sup>39</sup> Finally, from the question “Compared with other people your age, how intelligent are you?” (answers are: moderately below average, slightly below average, about average, slightly above average, moderately above average, extremely above average), we create the variable “Very intelligent” which equals 1 if the student reports that their intelligence level is “moderately above average” or “extremely above average” relative to others their own age and equals 0 otherwise.<sup>40</sup>

Since the above indicators of confidence and motivation are highly correlated, we use

---

<sup>38</sup>In our sample, 75 percent of girls and 68 percent of males rate the amount that they want to go to college as a 5.

<sup>39</sup>In our sample, 61 percent of girls and 50 percent of boys answer 5. We also tried classifying the dependent variable as 1 if the individual answers either 4 or 5 for this question, with similar results.

<sup>40</sup>The results are similar with a dependent variable equal to 1 only if the individual answers that she is “extremely above average”. Overall, 33 percent of females and 35 percent of males in our sample rate their intelligence as “moderately above average” or “extremely above average”.

factor analysis to reduce the dimensionality of the dependent variables.<sup>41</sup> We perform factor analysis separately for males and females on the set of variables described above and use the first factor (the only one with an eigenvalue greater than 1) as an index of confidence and motivation. Table 11 shows the variance explained by the first two factors, their eigenvalues, and the factor loadings. As shown, the measures of desire to attend college and likeliness of attending college load most strongly. By construction, the index has a mean of 0 and a standard deviation of 1.

Column (1) in Table 12 reports the results of using the main specification with the confidence index as the dependent variable. As shown in the table, a one standard deviation increase in the proportion of “high-achieving” boys decreases girls’ confidence and motivation by about 3 percent of a standard deviation; it has no effect on boys.<sup>42</sup> The effect of “high-achieving” girls is positive but statistically insignificant for both boys and girls irrespective of the specification of the main variable.

## 6.4 Risky Behavior

Another channel that can affect the ability and desire to attend and graduate from college is the extent to which students engage in risky behavior. As shown by Elsner and Isphording (2018), students with lower-ranked PVT scores relative to their peers are more likely to engage in risky behavior.<sup>43</sup> We can use behavioral questions asked in Wave I to examine whether “high achievers” influence the extent to which individuals engage in risky behavior such as drug and alcohol use, unprotected sex, and smoking among others.

Since many of the indicators of risky behavior are highly correlated, we use factor analysis to reduce the dimensionality of the dependent variables. Using behavioral questions from Wave I, we consider the following outcomes: “any alcohol”, which equals 1 if the individual has ever had more than a “couple of sips” of alcohol and equals 0 otherwise; “any cigarettes,” which equals 1 if the individual has ever smoked cigarettes and equals 0 otherwise; “any marijuana,” which equals 1 if the individual has smoked any marijuana in the past 30 days and equals 0 otherwise; “binge drinking,” which equals 1 if the individual has had 5 or more drinks “in a row” in the past year and equals 0 otherwise; “drunk,” which equals 1 if the individual reports being drunk in the past year and equals 0 otherwise; “fight,” which equals 1 if the individual reports getting in a “serious physical fight” in the past year and 0 otherwise; “unprotected sex,” which equals 1 if the individual did not use any form of birth control the most recent time she had sex and 0 otherwise;

---

<sup>41</sup>Factor analysis is an orthogonal transformation that converts a set of correlated variables into a fewer number of orthogonal variables. Each of the confidence and motivation outcomes is then viewed as a function of the “latent” variables reflected in the factors.

<sup>42</sup>The results for each of the variables in the index separately are shown in Appendix Table 2.

<sup>43</sup>The authors did not distinguish effects by gender either of the individual or their peers. As can be seen, both the gender of the individual and her peers seems to matter – only “high-achieving” male peers affect girls’ risky behavior.



and arrest before 18 which equals 1 if the individual was arrested before age 18 and 0 otherwise.<sup>44,45</sup>

We perform factor analysis separately for males and females on the set of risky behaviors described above and use the first two factors as indices of risky behavior (both have eigenvalues greater than 1).<sup>46</sup> Table 13 shows the variance explained by the first two factors, their eigenvalues, and the factor loadings. As shown, the various measures of alcohol use load most strongly for both girls and boys for the first factor (the first index of risky behavior). Physical fights and arrests load most strongly for the second factor.

Lastly, we explore an additional measure of behavior that may make college less likely: having a child before age 18. Wave IV asks individuals about the date (month and year) of each birth. We create a dummy variable for whether the individual first had a child before 18 years of age and set it equal to one if yes and to zero otherwise. In the sample, 7 percent of girls and 2 percent of boys reported having a child before age 18.

Table 12 reports the results using the full specification as usual. As shown in columns (2) and (3), exposure to a greater proportion of “high-achieving” boys increases girls’ risky behavior as measured by index 1 (with no significance for the index 2). Furthermore, as shown in column (4), greater exposure also increases the probability of having a child before age 18. A one standard-deviation increase in  $MF$  is associated with an increase of almost 1 percentage point in this probability.  $FF$ , on the other hand, decreases the first index of risky behavior (associated with drinking) but increases the second index (associated with fighting and arrests) and the probability of having a child before age 18. This effect could be driven by “high-achieving” boys lowering girls’ ability rank, as Elsner and Isphording (2018) find that lower PVT rank leads to greater risky behavior. To see whether this is the case, Table 12 repeats the analysis including a control for PVT rank calculated in the manner described in Section 5. The results given in Appendix Table 4 are very similar to those without rank control.

For boys, there is no significant impact of “high-achieving” boys as shown in columns (5)-(8). The effect of greater exposure to “high-achieving” girls is positive and significant: it reduces risky behavior as measured by both indices and also the likelihood of having a child before age 18.

Altogether, we take this as evidence that a possible pathway by which “high-achieving” boys affect girls is via increasing the latter’s propensity to engage in risky behavior including having a child before the age of 18. Exactly why this happens, as discussed previously, we cannot determine as it may be a reaction to either teacher or student

---

<sup>44</sup>Since there is no question on arrests in Wave I, we use the Wave IV question: “How many times were you arrested before your 18th birthday?”, and create a dummy equal to 1 if the individual was arrested before age 18.

<sup>45</sup>The means of these variables are reported in the base of Appendix Table 3.

<sup>46</sup>The loadings are very similar for girls and boys.

behavior.<sup>47</sup>

## 7 Heterogeneous Effects and Further Outcomes

The previous section showed that decreased confidence/aspirations and increased risky behavior were possible pathways leading girls to have worse long-run educational outcomes when exposed to a greater fraction of “high-achieving” boys in high school. We next examine heterogeneity in the main results, both with respect to individual/family characteristics and those of a school, as this will allow us to increase our understanding of the mechanisms at play. We also consider other long-run outcomes.

### 7.1 Heterogeneity

Individual ability and family background are strongly associated with the probability of graduating from college. To examine how these attributes matter we can split the sample according to (i) individual PVT score and (ii) parents’ education levels. For PVT scores, we divide the sample into at-or-below the median and above the median PVT.<sup>48</sup> For parental education, we divide the sample into students who have at least one residential parent with any kind of college degree or higher levels of education than this, and those students whose parents have lower levels of education than a college degree.<sup>49</sup>

Table 14 displays the results. As can be seen in columns (1) and (2), higher levels of  $MF$  reduce the likelihood that girls will graduate with a bachelor’s degree if their PVT is below the median. Specifically, a one standard deviation increase in  $MF$  (0.016 for girls with below-median PVT, always net of fixed effects and time trend) decreases bachelor’s degree attainment by 2.3 percentage points for this group. This is a very large effect: 21 percent of girls in this group on average graduate with a bachelor’s degree so this is over a 10 percent decrease. The effect on girls with an above-median PVT score is negative but not statistically significant.<sup>50</sup> A higher proportion of “high-achieving” girls, on the other hand increases college completion for below-median PVT females. From the magnitude of the coefficients, it is clear that an equal proportion of male and female “high-achievers” would have essentially a zero net effect on these girls. For girls with an above-median

---

<sup>47</sup>If we include the measures of risky behavior or early motherhood as additional controls in our main regression with bachelor’s degree as the dependent variable, the point estimate of  $MF$  is reduced in magnitude but remains significant. These regressions do not have a straightforward interpretation, however, these controls are themselves endogenous variables and hence we omit them.

<sup>48</sup>The median PVT score for females is 100 and we use this cutoff for both males and females to construct the samples.

<sup>49</sup>That is, if the student answers that either residential parent “graduated from a college or university” or has “professional training beyond a four-year college or university”, this is coded as college degree for the purpose of this question.

<sup>50</sup>A Wald test, however, is unable to reject equality of the coefficients on  $MF$  across columns (1) and (2) (p value=0.30).

PVT score, there is no statistically significant effect from female “high-achievers.”<sup>51</sup> This result is consistent with research suggesting that lower-ability females may be particularly positively influenced by higher-performing friends.<sup>52</sup> As can be seen in columns (7) and (8), there is no statistically significant effect of “high achievers” of either gender on boys.

Turning to parental education, we use the information on the highest level of education obtained by the residential mother and residential father as explained previously.<sup>53</sup> Contrasting column (3) with column (4), it is clear that the negative impact of “high-achieving” boys is concentrated among girls with a college-educated parent; there is no effect on the other group of girls. The magnitude of the estimate for girls with a college-educated parent is twice the one we obtained previously for the entire sample. A one standard-deviation increase in  $MF$  (2.1 percentage points for this sample) leads to a 4.5 percentage point decrease in college completion. This is roughly a 7 percent decrease on a mean of 61 percent. This suggests that the negative impact of boys is precisely on those girls who, from a family-background perspective, would be most likely to attend and graduate from college. The impact of  $FF$  is restricted to girls whose parents do not have college degree. For this group, a one standard deviation increase in  $FF$  is associated with a 1.3 percentage point increase in the probability of obtaining a college degree. Lastly, note that the equivalent sample split for boys in columns (9) and (10) once again shows no significant effect of either male or female “high achievers.”

Finally, we split the sample by an indicator of the socio-economic characteristics of the school rather than by individual characteristics. We rely on a question from the Add Health school administrator survey given to an administrator (e.g., the principal) in each sample school. Add Health asks the following question: “According to standardized achievement tests, approximately what percentage of all students at this school are testing: below grade level, at grade level, above grade level?” The fraction reported as testing at or above grade level is positively correlated with the average family income and average parental education level of students in the school.<sup>54</sup> The median fraction testing at or above grade level is 80 percent. We split schools according to whether they are strictly above this median or not.

Consistent with the results obtained when we split the sample by parental education, as can be seen comparing columns (5) and (6), the negative impact of “high-achieving” boys on girls appears in higher socio-economic/better performing schools. The effect of “high-achieving” girls is positive in this group as well. A one standard deviation increase in  $MF$  (2.0) is associated with a 2.7 percentage point decrease in bachelor’s

---

<sup>51</sup>In this case a Wald test can reject equality of the coefficient across the two columns at the 1 percent level (p value=0.003).

<sup>52</sup>See Hahn, Islam, Patacchini, and Zenou (2017).

<sup>53</sup>We exclude students for whom both parents’ educational information is missing. If there is only one parent in the household, we use that parent’s educational attainment.

<sup>54</sup>The correlation between the fraction in the school testing at or above grade level and the median family income of the Wave I students in the school is 0.46.

degree attainment – a 7 percent decrease on its mean of 36 percent. A one standard deviation increase in  $FF$  (1.8) increases bachelor’s degree attainment by 2.1 percentage points (or 6 percent of the mean). For boys, we again find no impact of “high achievers” in either group of schools.

## 7.2 Further Outcomes

Below we explore further the effect of “high achievers” on other long-term outcomes such as choice of major, labor force participation, fertility, and marriage.

In Wave III (when respondents are approximately 19-26), Add Health asks individuals who have some sort of college degree (including those with an associate or junior college degree (AA)) to list up to two major or minor fields of study.<sup>55</sup> We use this question to examine the impact of “high-achieving” peers on major/minor choice, creating a dummy variable equal to 1 if the individual reported a STEM field, where the classification of the latter is based on the National Science Foundation (NSF) Classification of STEM fields.<sup>56</sup> We restrict the sample to those individuals who were in at least grade 10 in Wave I (1995) and therefore would have been able to graduate with a BA by the summer of 2001 (Wave III). We focus on STEM fields because recent work has indicated that “high-achieving” males may affect women’s likelihood of choosing or completing mathematics-intensive and/or STEM majors.<sup>57</sup>

Table 15, column (1) (respectively column (5) for males) examines the effect on choice of major using STEM major/minor as a dependent variable. We include the same controls as in our main specification and in addition control for whether the individual has completed a bachelor’s degree (relative to a junior college degree) by Wave III. Although the coefficient on  $MF$  is negative, it is not statistically significant. There is, on the other hand, evidence that  $FF$  is associated with a higher proportion of STEM majors. The effect is large: a one standard deviation increase in  $FF$  (.019) is associated with an increase of 1.9 percentage points, a 13.2 percent increase over the mean of 14.5 percent.

Turning next to labor force participation, marriage, and fertility, we start by constructing indicators or measures of each. For labor force participation, we create a dummy equal to one if an individual states that they are currently employed, are on sick leave or temporarily disabled, are on maternity/paternity leave, or are unemployed and looking for work; the dummy is set equal to zero otherwise.<sup>58</sup> For the purposes of this question, we exclude those in the military or prison in Wave IV and restrict the sample to those individuals who were in 9th-12th grades in Wave I. This ensures that they are 28-32

<sup>55</sup>Unfortunately, field of study was not asked in Wave IV.

<sup>56</sup>See [https://www.lsamp.org/help/help\\_stem\\_ip2015.cfm](https://www.lsamp.org/help/help_stem_ip2015.cfm).

<sup>57</sup>E.g., Mouganie and Wang (2017) and Feld and Zölitz (2018).

<sup>58</sup>Employment information is based on the question: “are you currently working for pay at least 10 hours per week?”, where yes=1 and no=0. Individuals who report still working at their first full-time job are not asked this question so we code them as 1.

years old by Wave IV, and thus likely to have completed all schooling.<sup>59</sup> Marriage and fertility are based on questions to all sample respondents in Wave IV about whether they have ever been married (1=yes, 0=no) and the total number of (non-deceased) biological children they have.

The results are shown in columns (2)-(4) (respectively, columns (6)-(8) for males) of Table 15. As shown, a one standard deviation increase in  $MF$  (0.023) is associated with a 1.4 percentage point decline in labor force participation of women.<sup>60</sup> In contrast, a one standard deviation increase in  $FF$  (0.021) is associated with higher labor force participation of a slightly smaller magnitude. There is no impact of “high achievers” on a woman’s probability of having ever being married, but a one standard deviation increase in  $MF$  (0.020) increases a woman’s total number of biological children by 0.03, a 3 percent increase over its mean of 1.07. There is no effect of  $FF$  on males’ labor force participation or marriage, but a one standard deviation increase in  $FF$  decreases a man’s total number of biological children by 0.04, a 5 percent decrease on a mean of 0.72.

## 8 Robustness Checks

In this section we investigate the sensitivity of the results to alternative definitions of the main variable, the exclusion of outliers, alternative specifications, and attrition.

A natural alternative formulation of the explanatory variable would be the fraction of “high achievers” of a given gender among *all* students in the grade. It would be interesting and potentially informative if this formulation gave different answers, but as shown the answers are on the whole similar.<sup>61</sup> Table 16 repeats the most complete specification from Table 7 with alternative explanatory variables. In column (1) (respectively, column (3) for males)  $MF$  and  $FF$  replaced with  $MFA$  and  $FFA$ , where the  $A$  indicates that the high-achievers are measured as a fraction of the entire grade (*All*). As is clear, the negative impact of “high-achieving” males on females’ bachelor degree completion is a robust result. Furthermore, there is no symmetric negative effect of “high-achieving” girls on boys’ long-run education. Quantitatively, the coefficient on  $MFA$  indicates that a one standard deviation increase (0.011) in this fraction decreases females’ bachelor’s degree attainment by 1.7 percentage points.

An alternative definition would be to require both parents, rather than solely one, to have a post-college education. We denote these main variables  $MFB$  and  $FFB$ , for which each student receives a value of 1 if both parents have post-college education and

---

<sup>59</sup>In total, 16 females and 75 males are dropped due to being in the military or prison in Wave IV.

<sup>60</sup>We also explore employment as an outcome, and find a significant negative impact of  $MF$  on employment. This appears to be driven primarily, however, by differences in the work behavior of those who say they are students in Wave IV.

<sup>61</sup>For example, if we found that the results only hold when using  $MF$  and  $FF$  then we would be inclined to think that the pathway would be related to the formation of a perception about males.

0 otherwise. This more restrictive version of high-achievers necessarily results in lower means (e.g., for females the mean of  $MFB$  is 0.040 versus 0.145 for  $MF$ ; the mean of  $FFB$  is 0.035 versus 0.122 for  $FF$ ). As indicated by the coefficients in column (2), a one standard deviation increase in  $MFB$  (1.1 percentage points) decreases females' college completion by 1.7 percentage points. There is no significant impact of  $FFB$  for girls nor of any of the main peer variables for boys. There is no statistically significant effect of high-achievers on boys' outcomes.

We next consider our results without school-specific linear time-trends. While these allow us to control for time-varying attributes of a school in a linear fashion, they also decrease the amount of variation in the main variables. Columns (1)-(2) (respectively, columns (5)-(6) for boys) of Table 17 display our main results without time trends. As is clear from the table, the results remain strong and significant. A one standard deviation increase in  $MF$  for females (0.027) decreases the likelihood of graduating with a bachelor's degree by 2.3 percentage points, a similar magnitude to that obtained for our main specification (Table 7).

We can also examine the robustness of the results by eliminating outliers. As discussed in Section 4, the great majority of schools in our sample have variation in  $MF$  and  $FF$  similar to what would be expected by chance. For about 10 percent of schools, however, the variation lies outside the 90 percent confidence interval obtained through the Monte Carlo simulation. To test whether these outliers are driving the results, we re-run our main specifications [columns (5)-(6) of Table 7] excluding the schools whose within-school standard deviations of  $MF$  and  $FF$  lie outside of the 90 percent confidence interval. The results are shown in columns (3)-(4) (respectively, columns (7)-(8) for boys) of Table 17. As shown, the magnitude and significance of the coefficients are similar to what we obtained previously.

Lastly, we can also examine whether the results are due to differential attrition between Wave I and Wave IV as this could generate bias in the results. To do so, we regress a dummy for whether the individual remained in the sample between Wave I and Wave IV on the main variables. As shown in Appendix Table 5, there is no significant association between these and the dummy, suggesting that the results are unlikely to be driven by selective attrition of specific types of girls and boys from the sample.

## 9 Conclusion

This paper investigated the long-run effects of exposure to "high-achieving" boys versus girls in high school. Using a predetermined student characteristic – whether at least one of their parents has some post-college education – to proxy for a bundle of student characteristics, we investigated the consequences of quasi-random variation in the proportion of "high-achieving" girls and boys, separately, across grades within the same

school. As shown, we found a very strong asymmetric gender effect: the proportion of “high-achieving” boys has a statistically and economically significant negative effect on the probability that girls will end up with a bachelor’s degree some 14 years later. Boys, on the other hand, are not affected by their exposure to “high-achievers” of either gender.

Our paper suggests that in the future it would be useful to conduct experiments that examine how female subjects vs male subjects are affected by competing with varying proportions of high-performing males vs high-performing females. Are asymmetric gender effects also present there? Are any particular interventions helpful? It would also be of interest to examine whether the results the paper obtained are present in other data sets and in contexts other than high schools.

The data does not allow us to distinguish among various potential mechanisms. It could be that interacting with “high-achieving” boys has a direct negative effect on fellow female students. Or, the effect could be more indirect, e.g., arising from how teachers react to these students or even from how the parents of these boys affect teachers or the allocation of resources at the grade level. Nonetheless, we can identify some of the pathways. In particular, we show that greater exposure to “high-achieving” boys decreases girls’ self confidence and aspirations and increases their risky behavior including increasing teen-age motherhood. The girls especially affected are those in the lower half of the ability distribution as measured by their PVT score. Policies that target these more marginal girls – that increase their ambition and self-confidence or that decrease their exposure to these boys – are likely to have beneficial effects. Furthermore, our findings suggest that exposure to “high-achieving” girls has a positive effect of essentially the same absolute magnitude indicating that the negative effects of “high-achieving” boys can be counterbalanced through exposure to their female counterparts.

## 10 References

- Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), 151–184.
- Angrist, J. D., and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston’s Metco Program. *American Economic Review*, 94(5), 1613–1634.
- Athey, S., and Imbens, G. W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments* (Vol. 1, pp. 73–140). Elsevier.
- Bifulco, R., Fletcher, J. M., Oh, S. J., and Ross, S. L. (2014). Do high school peers have persistent effects on college attainment and other life outcomes? *Labour Economics*, 29, 83–90.
- Bifulco, R., Fletcher, J. M., and Ross, S. L. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1), 25–53.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1), 119–153.
- Bordalo, P., Coffman, K. B., Gennaioli, N., and Shleifer, A. (2018). Beliefs about Gender. Working Paper.
- Burbidge, J. B., Magee, L., and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401), 123–127.
- Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3), 439–464.
- Carrell, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11), 3377–3415.
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294.
- Elsner, B., and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3), 787–828.
- Elsner, B., and Isphording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, 53(2), 356–381.
- Feld, J., and Zölitz, U. (2018). Peers from Venus and Mars: Higher-Achieving Men Foster Gender Gaps in Major Choice and Labor Market Outcomes. Working Paper.
- Fischer, S. (2017). The downside of good peers: How classroom composition differentially



- affects men's and women's stem persistence. *Labour Economics*, 46, 211–226.
- Gould, E. D., Lavy, V., and Paserman, M. D. (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, 119(540), 1243–1269.
- Hahn, Y., Islam, A., Patacchini, E., and Zenou, Y. (2017). Do Friendship Networks Improve Female Education? Working Paper.
- Hoxby, C. M. (2000a). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4), 1239–1285.
- Hoxby, C. M. (2000b). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867.
- Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5), 2048–82.
- Lavy, V., Paserman, M. D., and Schlosser, A. (2011). 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.
- Lavy, V., and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2), 1–33.
- Merlino, L., Steinhardt, M. F., and Wren-Lewis, L. (2019). More than just friends? School peers and adult interracial relationships. *Forthcoming in Journal of Labor Economics*.
- Mouganie, P., and Wang, Y. (2017). High Performing Peers and Female STEM Choices in School. Working Paper.
- Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1), 1–16.
- Olivetti, C., Patacchini, E., and Zenou, Y. (2018). Mothers, Peers, and Gender-Role Identity. *Journal of the European Economic Association*.
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity*, 91–116.
- Rodríguez-Planas, N., Sanz-de Galdeano, A., and Terskaya, A. (2018). Independent Thinking and Hard Working, or Caring and Well Behaved? Short-and Long-Term Impacts of Gender Identity Norms. IZA Discussion Paper No. 11694.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the economics of education* (Vol. 3, pp. 249–277). Elsevier.
- van Tetering, M. A., de Groot, R. H., and Jolles, J. (2018). Boy–Girl Differences in Pictorial Verbal Learning in Students Aged 8–12 Years and the Influence of

Parental Education. *Frontiers in Psychology*, 9.

# 11 Tables

Table 1: Summary Statistics

	Add Health				2008 ACS			
	Females		Males		Females		Males	
	mean	sd	mean	sd	mean	sd	mean	sd
Age in Years (July 1995)	15.77	1.75	15.88	1.80				
Age in Years (July 2008)	28.77	1.75	28.88	1.80	28.77	1.75	28.88	1.80
White	0.65	0.48	0.67	0.47	0.67	0.47	0.67	0.47
Black	0.17	0.38	0.15	0.36	0.14	0.35	0.13	0.34
Latino	0.10	0.30	0.11	0.31	0.13	0.34	0.15	0.35
Asian	0.03	0.17	0.04	0.19	0.03	0.17	0.03	0.17
Other Race	0.05	0.21	0.04	0.20	0.03	0.16	0.03	0.16
Foreign Born	0.05	0.22	0.05	0.22	0.07	0.25	0.07	0.26
HS Graduate	0.94	0.24	0.91	0.29	0.91	0.28	0.88	0.33
Bachelor's Degree	0.35	0.48	0.28	0.45	0.33	0.47	0.26	0.44
Observations	5899		4954		153,269		148,470	

Note: This table reports summary statistics for the Add Health data sample used in the paper and the 2008 American Community Survey. The ACS sample excludes those who immigrated to the United States after 1994. ACS age in years is the average age in years of the sample with responses pooled over all survey months (ranging from January 2008 - December 2008) as birth dates are not reported. The ACS female and male samples are restricted to those aged 25-34 and re-weighted to match the age distribution of the Add Health female and male final samples, respectively. Wave IV weights are used in Add Health data.

Table 2: Parental Education and Child GPA: In-Home Sample

	Dependent Variable: Grade Point Average							
	Females (1)	(2)	(3)	(4)	Males (5)	(6)	(7)	(8)
Mother HS Grad	0.136*** (0.044)	0.097** (0.045)	0.074 (0.045)	0.069 (0.046)	0.138** (0.055)	0.116** (0.055)	0.091* (0.054)	0.074 (0.054)
Mother Some College	0.221*** (0.048)	0.170*** (0.048)	0.118** (0.046)	0.108** (0.044)	0.219*** (0.059)	0.181*** (0.058)	0.139** (0.057)	0.133** (0.057)
Mother College Grad	0.319*** (0.052)	0.245*** (0.053)	0.187*** (0.050)	0.187*** (0.049)	0.247*** (0.066)	0.194*** (0.066)	0.149** (0.065)	0.144** (0.065)
Mother Post College	0.488*** (0.065)	0.402*** (0.066)	0.291*** (0.059)	0.294*** (0.060)	0.381*** (0.077)	0.327*** (0.077)	0.265*** (0.072)	0.246*** (0.071)
Mother Not in HH	0.053 (0.060)	0.039 (0.061)	0.049 (0.062)	0.059 (0.064)	-0.155* (0.082)	-0.147* (0.078)	-0.154* (0.083)	-0.158* (0.086)
Father HS Grad	0.106** (0.047)	0.073 (0.047)	0.055 (0.046)	0.074* (0.043)	0.065 (0.064)	0.025 (0.057)	-0.001 (0.060)	0.009 (0.059)
Father Some College	0.212*** (0.062)	0.166*** (0.061)	0.126** (0.058)	0.152** (0.059)	0.228*** (0.066)	0.175*** (0.059)	0.134** (0.061)	0.130** (0.059)
Father College Grad	0.359*** (0.056)	0.302*** (0.057)	0.258*** (0.054)	0.266*** (0.051)	0.256*** (0.073)	0.209*** (0.066)	0.162** (0.069)	0.158** (0.066)
Father Post College	0.370*** (0.066)	0.306*** (0.064)	0.260*** (0.064)	0.275*** (0.063)	0.486*** (0.077)	0.418*** (0.075)	0.382*** (0.080)	0.409*** (0.078)
Father Not in HH	0.016 (0.043)	0.028 (0.043)	0.004 (0.044)	0.011 (0.042)	-0.042 (0.067)	-0.028 (0.061)	-0.057 (0.066)	-0.067 (0.064)
Age in Months		-0.017*** (0.002)	-0.013*** (0.002)	-0.013*** (0.002)		-0.017*** (0.002)	-0.013*** (0.003)	-0.012*** (0.003)
Foreign Born		0.233*** (0.071)	0.325*** (0.065)	0.308*** (0.068)		0.141* (0.080)	0.202** (0.091)	0.219** (0.088)
Log Family Income		0.038** (0.015)	0.029* (0.016)	0.026 (0.016)		0.039** (0.019)	0.028 (0.019)	0.023 (0.020)
PVT Score			1.204*** (0.132)	1.195*** (0.137)			0.945*** (0.135)	0.999*** (0.133)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Linear TT	No	No	No	Yes	No	No	No	Yes
Observations	5788	5788	5548	5548	4851	4848	4611	4611
R <sup>2</sup>	0.193	0.224	0.257	0.293	0.183	0.225	0.246	0.290
Equality P Value Moth Coll vs Moth Postcoll	0.001	0.001	0.017	0.013	0.013	0.016	0.021	0.049
Equality P Value Fath Coll vs Fath Postcoll	0.825	0.926	0.968	0.857	0.000	0.000	0.000	0.000

Note: This table reports parameter estimates and standard errors (in parentheses) from a regression of the student's grade point average on individual characteristics for the in-home sample. Grade point average is calculated based on self-reported student grades in math, science, english, and history from the Wave I in-home survey where A=4, B=3, C=2, and D or lower=1. All columns include dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Race fixed effects include dummies for Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01

Table 3: Parental Education and Child GPA: In-School Sample

	Dependent Variable: Grade Point Average					
	Females (1)	(2)	(3)	Males (4)	(5)	(6)
Mother HS Grad	0.125*** (0.016)	0.117*** (0.015)	0.115*** (0.015)	0.100*** (0.019)	0.094*** (0.019)	0.092*** (0.019)
Mother Some College	0.232*** (0.019)	0.227*** (0.018)	0.225*** (0.018)	0.204*** (0.023)	0.201*** (0.023)	0.200*** (0.022)
Mother College Grad	0.306*** (0.018)	0.284*** (0.019)	0.283*** (0.018)	0.270*** (0.024)	0.256*** (0.024)	0.253*** (0.024)
Mother Post College	0.339*** (0.024)	0.318*** (0.024)	0.317*** (0.024)	0.324*** (0.024)	0.313*** (0.024)	0.309*** (0.024)
Mother Not in HH	-0.008 (0.020)	-0.016 (0.018)	-0.017 (0.018)	0.000 (0.021)	-0.002 (0.020)	-0.005 (0.020)
Father HS Grad	0.066*** (0.018)	0.051*** (0.018)	0.049*** (0.018)	0.073*** (0.020)	0.066*** (0.020)	0.066*** (0.020)
Father Some College	0.227*** (0.022)	0.208*** (0.021)	0.207*** (0.021)	0.205*** (0.023)	0.192*** (0.022)	0.195*** (0.022)
Father College Grad	0.272*** (0.021)	0.240*** (0.021)	0.238*** (0.021)	0.299*** (0.024)	0.275*** (0.024)	0.279*** (0.024)
Father Post College	0.381*** (0.028)	0.347*** (0.027)	0.344*** (0.027)	0.380*** (0.029)	0.354*** (0.028)	0.357*** (0.028)
Father Not in HH	-0.079*** (0.017)	-0.063*** (0.016)	-0.063*** (0.016)	-0.058*** (0.018)	-0.041** (0.017)	-0.039** (0.017)
Foreign Born		0.136*** (0.023)	0.135*** (0.022)		0.119*** (0.023)	0.119*** (0.023)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	No	Yes	Yes
School Linear TT	No	No	Yes	No	No	Yes
Observations	39181	39181	39181	38021	38021	38021
$R^2$	0.171	0.190	0.201	0.160	0.175	0.182
Equality P Value Moth Coll vs Moth Postcoll	0.049	0.033	0.038	0.001	0.001	0.001
Equality P Value Fath Coll vs Fath Postcoll	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table reports parameter estimates and standard errors (in parentheses) from a regression of the student's grade point average on individual characteristics for the in-school sample. Grade point average is calculated based on self-reported student grades in math, science, english, and history from the in-school survey where A=4, B=3, C=2, and D or lower=1. Race fixed effects include dummies for Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. Standard errors clustered at the school level. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01

Table 4: Parental Education and Child PVT: In-Home Sample

	Dependent Variable: PVT Score					
	Females (1)	(2)	(3)	Males (4)	(5)	(6)
Mother HS Grad	3.157*** (0.752)	1.925*** (0.725)	1.839** (0.745)	4.558*** (0.952)	3.031*** (0.985)	2.970*** (1.022)
Mother Some College	5.320*** (0.916)	3.855*** (0.885)	3.873*** (0.918)	7.326*** (1.012)	5.319*** (0.992)	5.427*** (1.014)
Mother College Grad	6.227*** (0.833)	4.518*** (0.830)	4.713*** (0.832)	7.100*** (1.125)	4.968*** (1.094)	5.147*** (1.130)
Mother Post College	9.426*** (1.122)	7.327*** (1.100)	7.607*** (1.134)	8.828*** (1.433)	6.495*** (1.380)	6.798*** (1.424)
Mother Not in HH	0.269 (1.497)	-0.092 (1.470)	-0.307 (1.516)	2.015 (1.456)	0.982 (1.414)	1.498 (1.510)
Father HS Grad	1.529* (0.841)	0.802 (0.747)	0.663 (0.765)	2.459** (1.090)	1.558 (1.065)	1.159 (1.080)
Father Some College	4.014*** (1.106)	2.922*** (1.034)	2.578** (1.062)	5.724*** (1.314)	4.311*** (1.246)	3.580*** (1.285)
Father College Grad	4.545*** (0.965)	3.041*** (0.924)	2.935*** (0.924)	5.241*** (1.410)	3.808*** (1.275)	3.256** (1.288)
Father Post College	5.816*** (1.208)	4.143*** (1.183)	3.929*** (1.216)	6.951*** (1.592)	5.530*** (1.402)	4.990*** (1.362)
Father Not in HH	1.128 (0.889)	1.366 (0.824)	1.250 (0.838)	1.757 (1.159)	1.974* (1.097)	1.529 (1.162)
Age in Months		-0.344*** (0.042)	-0.346*** (0.043)		-0.397*** (0.036)	-0.406*** (0.037)
Foreign Born		-5.715*** (1.769)	-5.848*** (1.770)		-7.019*** (1.353)	-7.184*** (1.346)
Log Family Income		0.641* (0.329)	0.683** (0.336)		0.813*** (0.270)	0.825*** (0.301)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	No	Yes	Yes
School Linear TT	No	No	Yes	No	No	Yes
Observations	5650	5650	5650	4703	4703	4703
$R^2$	0.293	0.351	0.378	0.265	0.354	0.392
Equality P Value Moth Coll vs Moth Postcoll	0.000	0.001	0.001	0.113	0.148	0.129
Equality P Value Fath Coll vs Fath Postcoll	0.255	0.325	0.385	0.088	0.061	0.060

Note: This table reports parameter estimates and standard errors (in parentheses) from a regression of the student's Peabody Picture Vocabulary Test (PVT) score on individual characteristics for the in-home sample. All columns include dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Race fixed effects include dummies for Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 5: Variation in Main Variables

(a) Variation in  $MF$  and  $FF$ 

	Females		Males	
	MF	FF	MF	FF
Raw Variation				
Mean	0.145	0.122	0.143	0.119
SD	0.102	0.102	0.097	0.098
Min, Max	0.000, 0.707	0.000, 0.909	0.000, 0.713	0, 0.870
Net of Fixed Effects and School Trends				
Mean	0.000	0.000	0.000	0.000
SD	0.020	0.019	0.021	0.019
Min, Max	-0.109, 0.095	-0.133, 0.165	-0.171, 0.108	-0.072, 0.127
Count	5899	5899	4954	4954

(b) Variation in  $MN$  and  $FN$ 

	Females		Males	
	MN	FN	MN	FN
Raw Variation				
Mean	16.398	14.178	16.139	13.785
SD	15.543	15.090	14.810	14.230
Min, Max	0.00, 86.805	0.00, 80.457	0.00, 86.805	0.00, 80.457
Net of Fixed Effects and School Trends				
Mean	0 .000	0 .000	0 .000	0 .000
SD	2.670	2.095	2.540	2.099
Min, Max	-13.091, 16.843	-16.032, 10.179	-18.211, 14.169	-15.715, 14.704
Count	5899	5899	4954	4954

Note: This table reports the raw and residual (net of fixed effects and time trends) variation in  $MF$ ,  $FF$ ,  $MN$ , and  $FN$ .  $MF$  (respectively,  $FF$ ) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent) in the grade and school.  $MN$  (respectively,  $FN$ ) is the number of male (respectively, female) “high achievers” (those with at least one post-college parent) in the grade and school. Wave IV weights used.

Table 6: Balance Tests for  $MF$  and  $FF$ 

Panel A, Females						
	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months
MF	0.411 (0.592)	0.856 (6.627)	-0.147 (0.111)	-0.329 (0.250)	0.001 (0.200)	4.670 (3.691)
FF	0.245 (0.443)	-4.232 (8.250)	0.134 (0.113)	-0.195 (0.301)	0.162 (0.132)	2.050 (3.921)
Own Parent Post College	0.312*** (0.036)	5.163*** (0.725)	-0.009 (0.007)	-0.018 (0.019)	0.004 (0.019)	-1.103*** (0.288)
Panel B, Males						
	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months
MF	-0.037 (0.666)	3.858 (7.933)	0.107 (0.144)	-0.239 (0.356)	0.328 (0.208)	-2.782 (3.748)
Own Parent Post College	0.277*** (0.042)	4.183*** (0.615)	-0.014 (0.010)	-0.009 (0.019)	0.030* (0.015)	-1.157*** (0.294)
FF	0.370 (0.506)	3.731 (9.972)	-0.0879 (0.207)	-0.557** (0.226)	0.155 (0.134)	-2.980 (5.399)

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of  $MF$  and  $FF$  on individual characteristics. The estimates displayed in each row are for separate regressions in which the dependent variable is the variable name in the column and the independent variable is displayed in the row.  $MF$  (respectively,  $FF$ ) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent) in the grade and school. The regressions of  $FF$  in Panel A and  $MF$  in Panel B include a control for whether the individual has at least one post-college parent. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. All regressions are unweighted. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$



Table 7: High Achievers and Bachelor's Degree Attainment

	Dependent Variable: Bachelor's degree											
	Females (1)	(2)	(3)	(4)	(5)	(6)	Males (7)	(8)	(9)	(10)	(11)	(12)
MF	-1.053*** (0.293)	-1.046*** (0.276)	-1.019*** (0.269)	-0.991*** (0.296)	-1.083*** (0.317)	-0.945*** (0.286)	0.024 (0.428)	0.285 (0.387)	0.288 (0.389)	0.221 (0.365)	0.124 (0.365)	0.110 (0.380)
FF	-0.258 (0.384)	0.176 (0.357)	0.199 (0.348)	0.184 (0.364)	0.291 (0.338)	0.318 (0.331)	0.553 (0.550)	0.306 (0.522)	0.307 (0.526)	0.323 (0.502)	0.201 (0.493)	0.213 (0.498)
Foreign Born		0.089** (0.041)	0.081** (0.040)	0.089** (0.040)	0.087** (0.040)	0.061 (0.040)		0.102** (0.045)	0.099** (0.045)	0.101** (0.045)	0.104** (0.046)	0.068 (0.042)
PVT Score		0.530*** (0.065)	0.233* (0.133)	0.533*** (0.065)	0.535*** (0.064)			0.455*** (0.071)	0.211 (0.168)	0.450*** (0.071)	0.449*** (0.070)	
Age in Months		-0.006*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)	-0.008*** (0.001)		-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.007*** (0.001)
Mother HS Grad		0.060*** (0.022)	0.060*** (0.022)	0.060*** (0.022)	0.059*** (0.022)	0.064*** (0.021)		0.007 (0.021)	0.007 (0.021)	0.009 (0.021)	0.009 (0.021)	0.025 (0.021)
Mother Some College		0.115*** (0.026)	0.114*** (0.026)	0.115*** (0.026)	0.116*** (0.026)	0.138*** (0.027)		0.086*** (0.025)	0.086*** (0.025)	0.088*** (0.025)	0.088*** (0.025)	0.107*** (0.024)
Mother College Grad		0.213*** (0.036)	0.213*** (0.036)	0.213*** (0.036)	0.214*** (0.036)	0.239*** (0.036)		0.098*** (0.032)	0.098*** (0.032)	0.100*** (0.032)	0.099*** (0.032)	0.127*** (0.032)
Mother Post College		0.286*** (0.043)	0.284*** (0.043)	0.287*** (0.042)	0.287*** (0.043)	0.319*** (0.043)		0.202*** (0.040)	0.202*** (0.040)	0.205*** (0.040)	0.202*** (0.040)	0.241*** (0.043)
Mother Not in HH		0.067* (0.039)	0.066* (0.039)	0.066* (0.038)	0.067* (0.038)	0.063 (0.038)		-0.043 (0.041)	-0.045 (0.041)	-0.043 (0.041)	-0.042 (0.040)	-0.029 (0.037)
Father HS Grad		0.053** (0.023)	0.054** (0.023)	0.053** (0.023)	0.052** (0.023)	0.061** (0.024)		0.010 (0.027)	0.011 (0.027)	0.010 (0.028)	0.011 (0.027)	0.017 (0.025)
Father Some College		0.120*** (0.026)	0.119*** (0.025)	0.120*** (0.026)	0.121*** (0.026)	0.137*** (0.025)		0.107*** (0.030)	0.106*** (0.030)	0.106*** (0.030)	0.105*** (0.030)	0.116*** (0.030)
Father College Grad		0.239*** (0.035)	0.240*** (0.035)	0.240*** (0.035)	0.240*** (0.035)	0.254*** (0.035)		0.201*** (0.045)	0.202*** (0.045)	0.201*** (0.045)	0.200*** (0.044)	0.218*** (0.041)
Father Post College		0.301*** (0.052)	0.298*** (0.052)	0.300*** (0.052)	0.298*** (0.052)	0.323*** (0.052)		0.319*** (0.044)	0.319*** (0.044)	0.318*** (0.044)	0.315*** (0.044)	0.317*** (0.047)
Father Not in HH		0.034 (0.023)	0.034 (0.023)	0.034 (0.023)	0.033 (0.023)	0.036 (0.024)		0.042 (0.030)	0.043 (0.030)	0.041 (0.030)	0.042 (0.030)	0.049* (0.028)
Log Family Income		0.046*** (0.012)	0.046*** (0.012)	0.046*** (0.012)	0.046*** (0.012)	0.042*** (0.012)		0.024** (0.011)	0.024** (0.011)	0.023** (0.011)	0.023** (0.011)	0.028** (0.012)
PVT Rank			0.135** (0.058)						0.109 (0.067)			
Fraction Female				0.394* (0.226)	0.438* (0.226)	0.325 (0.217)				-0.540** (0.270)	-0.576** (0.267)	-0.594** (0.265)
Fraction Foreign Born					-0.942 (0.642)	-0.816 (0.627)					1.294** (0.522)	1.478*** (0.500)
Fraction Black					-0.703 (0.522)	-0.606 (0.509)					-0.341 (0.603)	-0.344 (0.592)
Fraction Latino					-0.292 (0.431)	-0.154 (0.414)					-0.120 (0.410)	-0.199 (0.411)
Fraction Asian					0.638 (0.686)	0.546 (0.645)					-1.534* (0.793)	-1.628** (0.786)
Fraction Other Races					-0.431 (0.310)	-0.274 (0.298)					-0.113 (0.250)	-0.126 (0.246)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5899	5650	5649	5650	5650	5898	4954	4703	4703	4703	4703	4948
R <sup>2</sup>	0.187	0.332	0.332	0.332	0.334	0.318	0.199	0.326	0.326	0.326	0.328	0.312
Adjusted R <sup>2</sup>	0.152	0.299	0.299	0.300	0.300	0.285	0.158	0.286	0.286	0.287	0.287	0.273

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) "high achievers" (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Race fixed effects include dummies for Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01

Table 8: High Achievers and Bachelor's Degree Attainment

	Dependent Variable: Bachelor's degree											
	Females (1)	(2)	(3)	(4)	(5)	(6)	Males (7)	(8)	(9)	(10)	(11)	(12)
IHS MN	-0.100*** (0.023)	-0.103*** (0.028)	-0.100*** (0.027)	-0.094*** (0.031)	-0.095*** (0.030)	-0.082*** (0.026)	-0.005 (0.052)	0.027 (0.045)	0.027 (0.046)	0.005 (0.046)	0.004 (0.046)	0.001 (0.047)
IHS FN	0.036 (0.033)	0.058** (0.027)	0.058** (0.026)	0.053* (0.028)	0.050* (0.026)	0.051* (0.027)	0.016 (0.050)	0.001 (0.046)	0.001 (0.046)	0.017 (0.043)	0.035 (0.047)	0.036 (0.051)
Foreign Born		0.089** (0.040)	0.081** (0.040)	0.089** (0.040)	0.087** (0.040)	0.062 (0.040)		0.103** (0.045)	0.100** (0.045)	0.101** (0.045)	0.100** (0.045)	0.064 (0.041)
PVT Score		0.531*** (0.065)	0.235* (0.135)	0.532*** (0.065)	0.532*** (0.064)			0.455*** (0.071)	0.211 (0.169)	0.450*** (0.071)	0.450*** (0.071)	
Age in Months		-0.006*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)	-0.008*** (0.001)		-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.007*** (0.001)
Mother HS Grad		0.060*** (0.022)	0.059*** (0.022)	0.060*** (0.022)	0.058** (0.022)	0.062*** (0.021)		0.008 (0.021)	0.008 (0.021)	0.010 (0.021)	0.007 (0.021)	0.024 (0.021)
Mother Some College		0.115*** (0.026)	0.114*** (0.026)	0.115*** (0.026)	0.115*** (0.026)	0.138*** (0.027)		0.086*** (0.025)	0.086*** (0.025)	0.089*** (0.024)	0.085*** (0.025)	0.104*** (0.024)
Mother College Grad		0.213*** (0.036)	0.212*** (0.036)	0.213*** (0.036)	0.212*** (0.036)	0.238*** (0.036)		0.099*** (0.032)	0.098*** (0.032)	0.101*** (0.032)	0.098*** (0.032)	0.127*** (0.032)
Mother Post College		0.287*** (0.043)	0.285*** (0.043)	0.287*** (0.043)	0.286*** (0.043)	0.318*** (0.043)		0.203*** (0.040)	0.203*** (0.040)	0.205*** (0.040)	0.204*** (0.040)	0.244*** (0.043)
Mother Not in HH		0.067* (0.039)	0.066* (0.038)	0.067* (0.038)	0.065* (0.038)	0.062 (0.038)		-0.043 (0.041)	-0.045 (0.041)	-0.043 (0.041)	-0.044 (0.040)	-0.030 (0.037)
Father HS Grad		0.052** (0.023)	0.052** (0.023)	0.052** (0.023)	0.051** (0.023)	0.060** (0.024)		0.010 (0.027)	0.011 (0.027)	0.010 (0.027)	0.012 (0.027)	0.019 (0.025)
Father Some College		0.119*** (0.026)	0.118*** (0.026)	0.119*** (0.026)	0.117*** (0.026)	0.135*** (0.025)		0.106*** (0.030)	0.106*** (0.030)	0.106*** (0.030)	0.106*** (0.030)	0.117*** (0.030)
Father College Grad		0.240*** (0.035)	0.240*** (0.035)	0.240*** (0.035)	0.238*** (0.035)	0.251*** (0.035)		0.202*** (0.045)	0.203*** (0.045)	0.201*** (0.045)	0.202*** (0.044)	0.220*** (0.041)
Father Post College		0.300*** (0.051)	0.296*** (0.051)	0.299*** (0.052)	0.296*** (0.052)	0.321*** (0.052)		0.319*** (0.044)	0.319*** (0.044)	0.318*** (0.044)	0.316*** (0.044)	0.319*** (0.047)
Father Not in HH		0.034 (0.023)	0.034 (0.023)	0.034 (0.023)	0.031 (0.023)	0.035 (0.024)		0.041 (0.031)	0.042 (0.031)	0.041 (0.031)	0.042 (0.030)	0.049* (0.028)
Log Family Income		0.045*** (0.012)	0.046*** (0.012)	0.045*** (0.012)	0.045*** (0.012)	0.041*** (0.011)		0.024** (0.011)	0.024** (0.011)	0.023** (0.011)	0.023** (0.011)	0.028** (0.012)
PVT Rank			0.134** (0.059)						0.109 (0.067)			
Fraction Female				0.222 (0.248)	0.383 (0.256)	0.261 (0.236)				-0.575** (0.280)	-0.722** (0.290)	-0.747** (0.292)
IHS Number Foreign Born					-0.050** (0.022)	-0.046** (0.021)					0.048** (0.021)	0.049** (0.021)
IHS Number Black					0.004 (0.027)	0.013 (0.026)					-0.032 (0.043)	-0.028 (0.043)
IHS Number Latino					0.022 (0.022)	0.023 (0.019)					-0.009 (0.023)	-0.012 (0.022)
IHS Number Asian					0.021 (0.032)	0.014 (0.030)					-0.060* (0.031)	-0.054* (0.029)
IHS Number Other Race					-0.010 (0.047)	0.008 (0.046)					-0.031 (0.044)	-0.038 (0.039)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5899	5650	5649	5650	5650	5898	4954	4703	4703	4703	4703	4948
R <sup>2</sup>	0.186	0.332	0.332	0.332	0.333	0.318	0.199	0.325	0.326	0.326	0.328	0.312
Adjusted R <sup>2</sup>	0.152	0.300	0.299	0.300	0.300	0.286	0.158	0.285	0.286	0.286	0.287	0.273

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. *MN* (respectively, *FN*) is the number of male (respectively, female) "high achievers" (those with at least one post-college parent). IHS refers to the inverse hyperbolic sine transformation. All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Race fixed effects include dummies for Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01

Table 9: High Achievers and Educational Attainment

	Females		Males	
	(1)	(2)	(3)	(4)
	Vocational or Associate's Degree	High School Graduate	Vocational or Associate's Degree	High School Graduate
MF	1.120*** (0.300)	0.060 (0.110)	0.055 (0.281)	0.284 (0.173)
FF	-0.243 (0.320)	0.038 (0.132)	0.073 (0.240)	-0.114 (0.197)
PVT Score	-0.042 (0.061)	0.220*** (0.044)	-0.076 (0.075)	0.206*** (0.048)
Fraction Female	-0.243 (0.287)	-0.124 (0.110)	0.172 (0.261)	0.120 (0.185)
School, Grade FE	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes
Observations	5650	5650	4703	4703
$R^2$	0.097	0.199	0.115	0.242
Adjusted $R^2$	0.052	0.159	0.062	0.196

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of educational attainment on individual and peer characteristics with the dependent variable listed in the column heading. The variable Vocational or Associate's Degree takes a value of 1 if the individual has a vocational/technical degree from a program lasting 1-3 years or an associate's degree and 0 otherwise. The variable High School Graduate takes a value of 1 if the individual has completed high school and 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) "high achievers" (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father's education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 10: High Achievers and Grades

	Females					Males				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	GPA	Math	Science	English	History	GPA	Math	Science	English	History
MF	-1.043 (0.643)	-2.289*** (0.850)	-2.010** (0.856)	-0.699 (0.850)	0.038 (0.925)	0.543 (0.531)	0.674 (0.820)	-0.347 (0.919)	0.686 (0.799)	0.522 (0.979)
FF	0.222 (0.521)	-0.141 (0.658)	-0.546 (0.858)	0.901 (0.698)	0.352 (0.891)	0.692 (0.728)	-0.839 (0.978)	0.744 (0.904)	1.529 (0.944)	1.282 (1.055)
PVT Score	1.197*** (0.136)	0.710*** (0.169)	1.355*** (0.173)	1.304*** (0.149)	1.559*** (0.179)	1.006*** (0.132)	0.467*** (0.169)	1.173*** (0.190)	0.912*** (0.160)	1.498*** (0.183)
Fraction Female	0.175 (0.450)	0.512 (0.684)	0.633 (0.607)	-0.363 (0.648)	0.109 (0.655)	0.872 (0.598)	0.617 (0.723)	0.468 (0.864)	0.948 (0.906)	2.034** (0.999)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5548	5224	4961	5508	4946	4611	4388	4135	4552	4126
$R^2$	0.295	0.181	0.248	0.252	0.255	0.292	0.239	0.253	0.234	0.257
Adjusted $R^2$	0.259	0.137	0.206	0.214	0.213	0.248	0.189	0.201	0.186	0.206

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of grades on individual and peer characteristics. Grades are based on student reports of their grade in each subject over the previous year, with A=4, B=3, C=2, and D or lower=1. Average GPA reflects the average across these four subjects. *MF* (respectively, *FF*) is the fraction of male (respectively, female) "high achievers" (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father's education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 11: Confidence and Motivation Factor Loadings

(a) Females

	Eigenvalue	Proportion of Variance
Factor 1	1.65	0.55
Factor 2	0.88	0.29
Factor 3	0.47	0.16
Rotated Factor Loadings	Factor 1	
Very Intelligent	0.51	
College Likely	0.85	
Want College	0.82	

(b) Males

	Eigenvalue	Proportion of Variance
Factor 1	1.58	0.53
Factor 2	0.87	0.29
Factor 3	0.54	0.18
Rotated Factor Loadings	Factor 1	
Very Intelligent	0.54	
College Likely	0.82	
Want College	0.79	

Note: This table reports eigenvalues and factor loadings based on factor analysis of the following variables: “Want College”, which equals 1 if the student reports that they want to go to college as a 5 on a scale of 1-5 and equals 0 otherwise; “College likely”, which equals 1 if the student says the likelihood that they will go to college is a 5 on a scale of 1-5 and equals 0 otherwise; and “Very intelligent”, which equals 1 if the student reports that their intelligence level is “moderately above average” or “extremely above average” relative to others their own age and equals 0 otherwise. Factor analysis performed separately for males and females. Wave IV weights used.

Table 12: High Achievers and Confidence and Risky Behaviors

	Females				Males			
	(1) Confidence Index	(2) Risky Index 1	(3) Risky Index 2	(4) Birth Before 18	(5) Confidence Index	(6) Risky Index 1	(7) Risky Index 2	(8) Birth Before 18
MF	-1.337** (0.673)	1.397* (0.720)	0.662 (0.759)	0.467*** (0.164)	0.111 (0.894)	0.175 (0.724)	-0.871 (0.779)	0.111 (0.103)
FF	0.353 (0.712)	-1.346** (0.674)	1.306** (0.607)	0.354* (0.183)	1.378* (0.780)	-1.597** (0.698)	-1.974** (0.914)	-0.179** (0.081)
PVT Score	1.265*** (0.166)	-0.302* (0.169)	-0.364* (0.187)	-0.122*** (0.045)	0.955*** (0.167)	0.094 (0.155)	-0.266 (0.174)	-0.063** (0.028)
Fraction Female	-0.220 (0.540)	-0.247 (0.500)	-0.693 (0.549)	0.000 (0.122)	-0.518 (0.660)	0.573 (0.767)	-0.235 (0.684)	0.106 (0.123)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5631	5513	5513	5548	4685	4505	4505	4610
$R^2$	0.237	0.235	0.174	0.130	0.245	0.291	0.191	0.154
Adjusted $R^2$	0.199	0.196	0.132	0.086	0.199	0.247	0.140	0.102

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of confidence and risky behaviors on individual and peer characteristics. The Confidence Index is the first factor from a factor analysis of three variables measuring self-perceptions of intelligence, desire to go to college, and the likelihood of going to college. The Risky Index 1 (respectively, 2) is the first (respectively, second) factor from a factor analysis of 8 variables measuring risky behavior. First Birth Before 18 takes a value of 1 if the individual has had a child by the time she turns age 18 and 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father’s education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 13: Risky Behavior Factor Loadings

## (a) Females

	Eigenvalue	Proportion of Variance
Factor1	2.97	0.37
Factor2	1.06	0.13
Factor3	0.97	0.12

Rotated Factor Loadings	Factor 1	Factor 2
Any Cigarette	0.61	0.20
Any Alcohol	0.75	0.04
Drunk	0.86	0.03
Binge	0.84	0.00
Any Marijuana	0.58	0.26
Fight	0.07	0.73
Arrested before 18	0.01	0.71
Unprotected Sex	0.35	0.18

## (b) Males

	Eigenvalue	Proportion of Variance
Factor1	2.97	0.37
Factor2	1.12	0.14
Factor3	0.94	0.12

Rotated Factor Loadings	Factor 1	Factor 2
Any Cigarette	0.59	0.17
Any Alcohol	0.75	0.06
Drunk	0.88	0.06
Binge	0.87	0.02
Any Marijuana	0.55	0.23
Fight	0.07	0.73
Arrested before 18	0.04	0.74
Unprotected Sex	0.27	0.28

Note: This table reports eigenvalues and factor loadings based on factor analysis of the following variables: “any alcohol”, which equals 1 if the individual has ever had more than a “couple of sips” of alcohol and equals 0 otherwise; “any cigarettes”, which equals 1 if the individual has ever smoked cigarettes and equals 0 otherwise; “any marijuana”, which equals 1 if the individual has smoked any marijuana in the past 30 days and equals 0 otherwise; “binge drinking”, which equals 1 if the individual has had 5 or more drinks “in a row” in the past year and equals 0 otherwise; “drunk”, which equals 1 if the individual reports being drunk in the past year and equals 0 otherwise; “fight”, which equals 1 if the individual reports getting in a “serious physical fight” in the past year and 0 otherwise; “unprotected sex”, which equals 1 if the individual did not use any form of birth control the most recent time she had sex and 0 otherwise; and arrest before 18, which equals 1 if the individual was arrested before age 18 and 0 otherwise. Factor analysis performed separately for males and females. Wave IV weights used.

Table 14: High Achievers and Bachelor's Degree Attainment: Heterogeneity

	Dependent Variable: Bachelor's Degree											
	Females						Males					
	Below Median PVT (1)	Above Median PVT (2)	Neither Parent College (3)	Parent College (4)	School Below Median Test Scores (5)	School Above Median Test Scores (6)	Below Median PVT (7)	Above Median PVT (8)	Neither Parent College (9)	Parent College (10)	School Below Median Test Scores (11)	School Above Median Test Scores (12)
MF	-1.410*** (0.476)	-0.739 (0.486)	-0.038 (0.431)	-2.138*** (0.546)	-0.144 (0.820)	-1.351** (0.619)	-0.192 (0.668)	0.188 (0.359)	-0.217 (0.503)	0.554 (0.525)	0.943 (0.832)	0.317 (0.419)
FF	1.493*** (0.469)	-0.353 (0.490)	0.697* (0.390)	0.325 (0.679)	0.357 (0.453)	1.156** (0.494)	0.776 (0.707)	0.097 (0.609)	0.344 (0.615)	-0.015 (0.755)	-0.282 (0.531)	0.313 (0.924)
PVT Score	0.465*** (0.119)	0.587*** (0.201)	0.505*** (0.083)	0.601*** (0.142)	0.508*** (0.091)	0.620*** (0.102)	-0.011 (0.104)	0.795*** (0.161)	0.374*** (0.085)	0.596*** (0.164)	0.412*** (0.103)	0.562*** (0.121)
Fraction Female	0.331 (0.264)	0.583 (0.373)	0.348 (0.222)	0.442 (0.579)	0.210 (0.535)	0.420 (0.324)	-0.491 (0.428)	-0.635* (0.346)	-0.525 (0.322)	-0.585 (0.554)	-1.530*** (0.458)	-0.437* (0.249)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2913	2737	3443	1939	2245	2207	2093	2610	2744	1731	1842	1872
$R^2$	0.330	0.357	0.254	0.405	0.303	0.339	0.351	0.368	0.279	0.443	0.336	0.336
Adjusted $R^2$	0.263	0.290	0.192	0.313	0.266	0.297	0.258	0.296	0.203	0.345	0.292	0.286

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. Columns (1)-(2) and (7)-(8) split the sample into below-median and above-median PVT score. Columns (3)-(4) and (9)-(10) split the sample by whether at least one parent has a college degree. Columns (5)-(6) and (11)-(12) split the sample by the fraction of students at the individual's school testing at or above grade level. *MF* (respectively, *FF*) is the fraction of male (respectively, female) "high achievers" (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father's education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 15: Other Outcomes

	Females				Males			
	(1) STEM Major	(2) LFP	(3) Ever Married	(4) Total No. Children	(5) STEM Major	(6) LFP	(7) Ever Married	(8) Total No. Children
MF	-0.202 (0.561)	-0.623** (0.305)	-0.239 (0.457)	1.670** (0.775)	0.115 (0.733)	-0.085 (0.212)	-0.237 (0.430)	-0.248 (0.681)
FF	1.021* (0.522)	0.465* (0.241)	-0.022 (0.467)	-1.428 (0.971)	0.564 (1.316)	-0.078 (0.235)	-0.470 (0.528)	-1.908** (0.877)
PVT Score	0.153 (0.160)	0.102 (0.070)	-0.107 (0.075)	-0.962*** (0.205)	0.301 (0.338)	0.027 (0.057)	-0.013 (0.095)	-0.466*** (0.168)
Fraction Female	0.547 (0.822)	-0.077 (0.260)	0.189 (0.308)	0.616 (0.630)	-3.223*** (1.176)	0.383** (0.155)	-0.034 (0.282)	0.569 (0.599)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	909	4090	5649	5640	584	3275	4698	4678
$R^2$	0.298	0.100	0.202	0.222	0.455	0.161	0.205	0.208
Adjusted $R^2$	0.148	0.061	0.162	0.183	0.267	0.114	0.156	0.159

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of STEM field choice, labor force participation, marriage, and fertility on individual and peer characteristics. STEM major choice is equal to 1 if the individual lists a STEM field as one of her two major/minor fields of study and is equal to 0 otherwise. LFP is equal to 1 if the individual is currently working at least 10 hours per week, is on sick leave or temporarily disabled, is on maternity/paternity leave, or is unemployed and looking for work and is equal to zero otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother’s and father’s education (dummies for high school, some college but no degree, college degree, post college for each parent), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Columns (1) and (5) also include an indicator for whether the individual has completed a bachelor’s degree by Wave III. Sample for STEM field of study restricted to students in grades 10-12 in Wave I who have completed a postsecondary degree by Wave III. Sample for LFP restricted to students in grades 9-12 in Wave I. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$



Table 16: High Achievers and Bachelor's Degree Attainment: Alternative Measures

	Dependent Variable: Bachelor's degree			
	Females (1)	(2)	Males (3)	(4)
MFA	-1.579** (0.644)		0.208 (0.651)	
FFA	-0.198 (0.687)		0.544 (0.859)	
MFB		-1.537** (0.680)		-0.097 (0.783)
FFB		0.511 (0.570)		-0.081 (0.749)
PVT Score	0.535*** (0.064)	0.531*** (0.064)	0.449*** (0.070)	0.449*** (0.070)
Fraction Female	0.307 (0.257)	0.539** (0.209)	-0.611** (0.286)	-0.592** (0.267)
School, Grade FE	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes
Individual Characteristics Controls	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes
Observations	5650	5650	4703	4703
$R^2$	0.333	0.333	0.328	0.328
Adjusted $R^2$	0.300	0.300	0.287	0.287

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. *MFA* (respectively, *FFA*) represents male (respectively, female) "high achievers" (those with at least one post-college parent) as a fraction of all students in the grade. *MFB* (respectively, *FFB*) is represents male (respectively, female) "high achievers" (those with two post-college parents) as a fraction of males (respectively, females) in the grade. All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father's education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

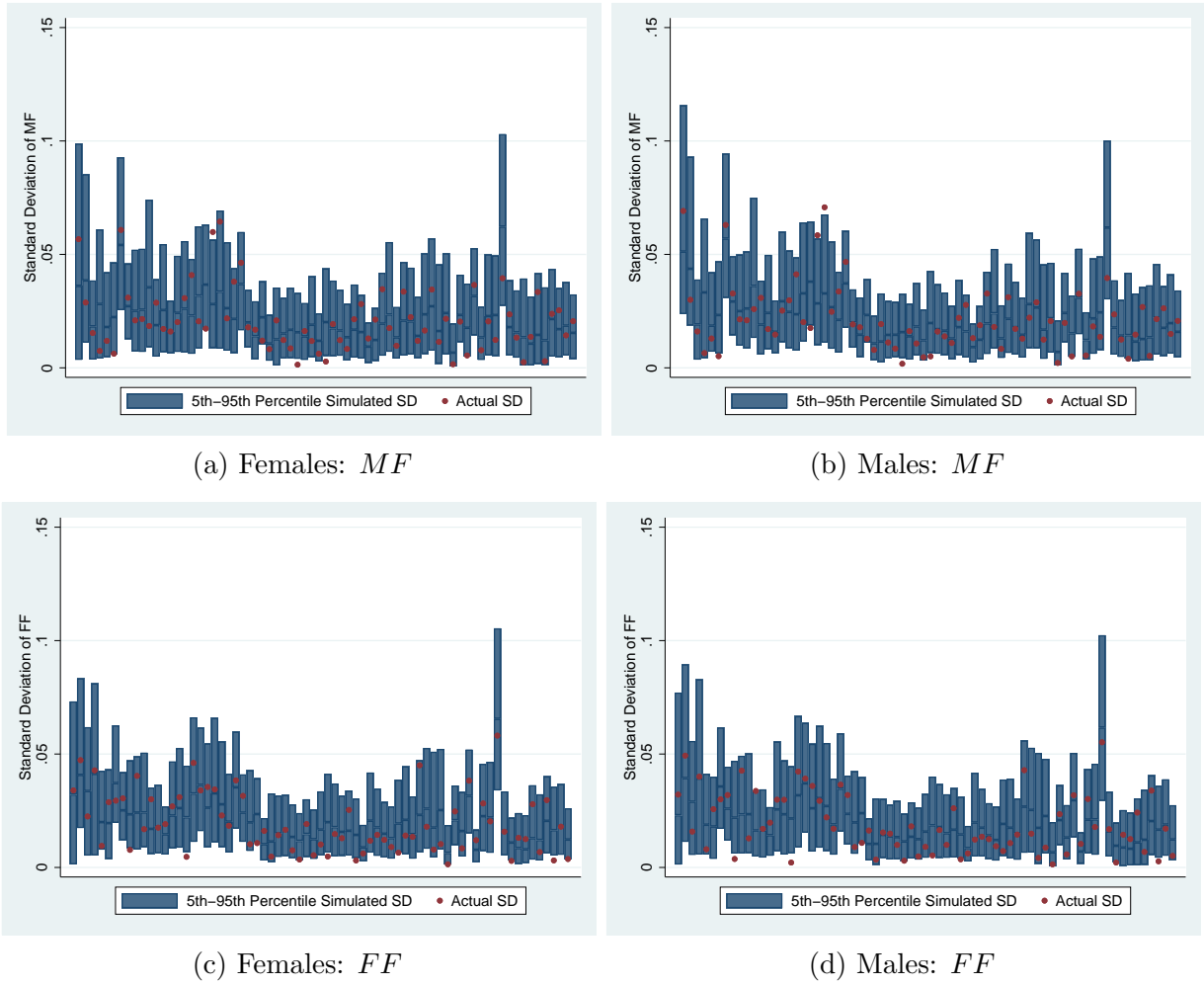
Table 17: High Achievers and Bachelor's Degree Attainment: Robustness

	Dependent Variable: Bachelor's degree							
	Females (1)	(2)	(3)	(4)	Males (5)	(6)	(7)	(8)
MF	-0.851*** (0.272)	-0.823*** (0.247)	-1.148*** (0.355)	-1.001*** (0.325)	0.201 (0.295)	0.178 (0.298)	0.151 (0.465)	0.108 (0.482)
FF	0.047 (0.266)	0.055 (0.262)	0.163 (0.326)	0.194 (0.323)	-0.042 (0.370)	-0.111 (0.365)	0.375 (0.530)	0.389 (0.528)
PVT Score	0.544*** (0.064)		0.527*** (0.067)		0.469*** (0.069)		0.426*** (0.075)	
Fraction Female	0.388*** (0.143)	0.291** (0.143)	0.461** (0.231)	0.343 (0.229)	0.011 (0.214)	-0.001 (0.216)	-0.611** (0.293)	-0.592** (0.292)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	No	No	Yes	Yes	No	No	Yes	Yes
Individual Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5650	5898	5098	5326	4703	4948	4135	4364
$R^2$	0.308	0.292	0.337	0.321	0.289	0.274	0.328	0.313
Adjusted $R^2$	0.289	0.274	0.303	0.288	0.266	0.251	0.286	0.273

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. Columns (1)-(2) (respectively (5)-(6) for males) exclude linear time trends. Columns (3)-(4) (respectively (7)-(8) for males) exclude schools with variation in  $MF$  and  $FF$  outside of the 90 percent confidence interval obtained in the Monte Carlo simulations.  $MF$  (respectively,  $FF$ ) is the fraction of male (respectively, female) "high achievers" (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father's education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. If mother's (respectively, father's) education is missing, all mother's (respectively, father's) education dummies are set to zero and a dummy is included for missing mother's (respectively, father's) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Coefficient on PVT score multiplied by 100.

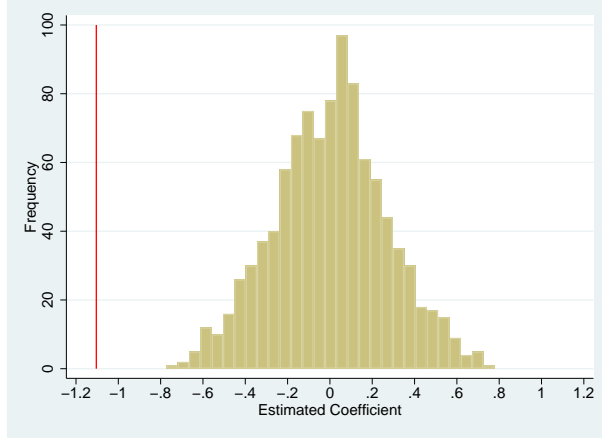
## 12 Figures

Figure 1: Monte Carlo Estimates of  $MF$  and  $FF$

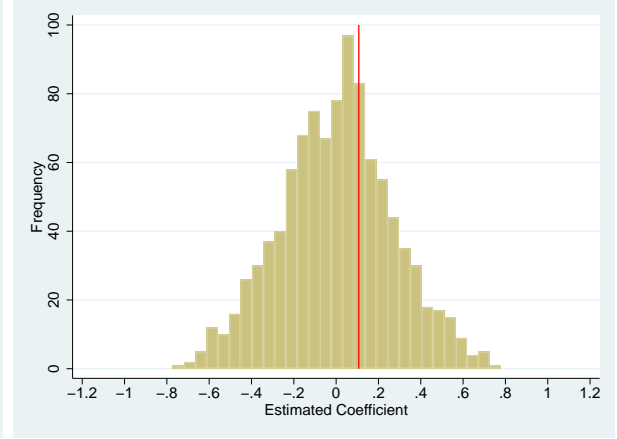


Note: These figures display simulated and actual standard deviations for schools in the sample with at least three grades, with each bar representing a different school. Upper and lower edges of the bar represent the 5th and 95th percentiles respectively of the simulated within-school standard deviation of  $MF$  (or  $FF$ ). The dot represents the empirical standard deviation.

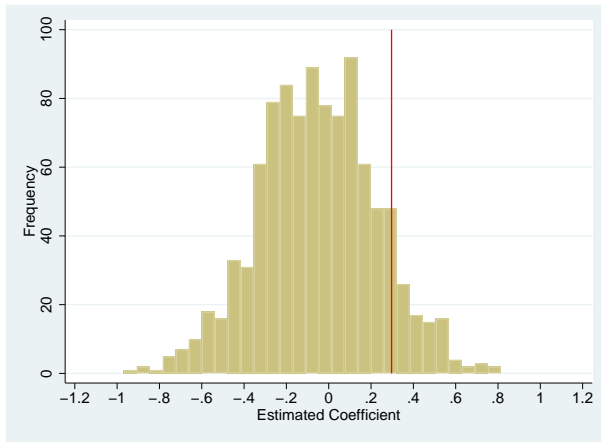
Figure 2: Randomization-Based Inference for  $MF$  and  $FF$



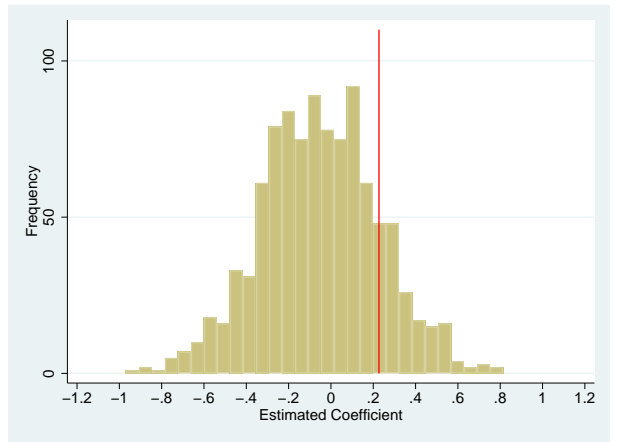
(a) Females:  $MF$



(b) Males:  $MF$



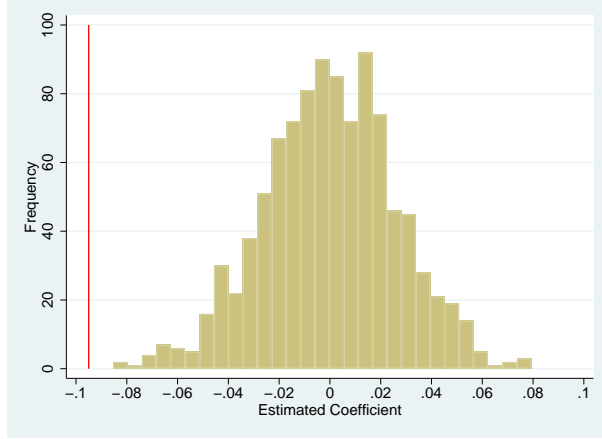
(c) Females:  $FF$



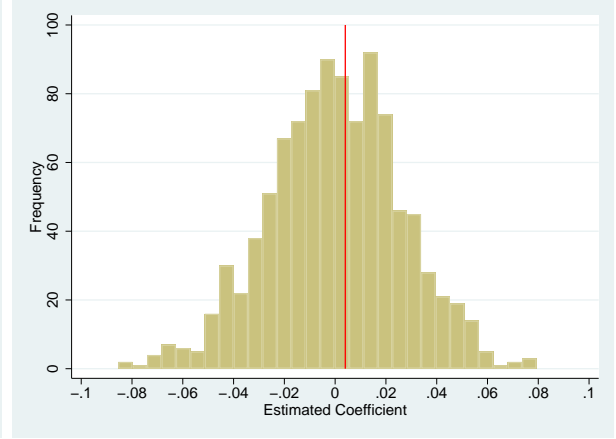
(d) Males:  $FF$

Note: These figures show distribution of coefficients obtained from the final OLS specification in Table 7 while replacing  $MF$  (respectively,  $FF$ ) with the value of  $MF$  (respectively,  $FF$ ) from a random grade in the same school. Red line represents actual estimate obtained in specifications (5) and (11) in Table 7.

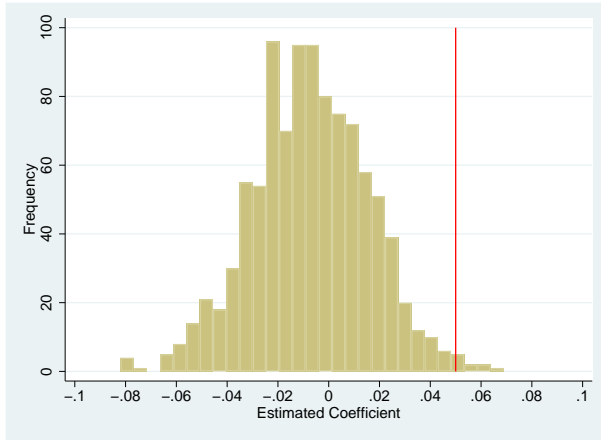
Figure 3: Randomization-Based Inference for  $MN$  and  $FN$



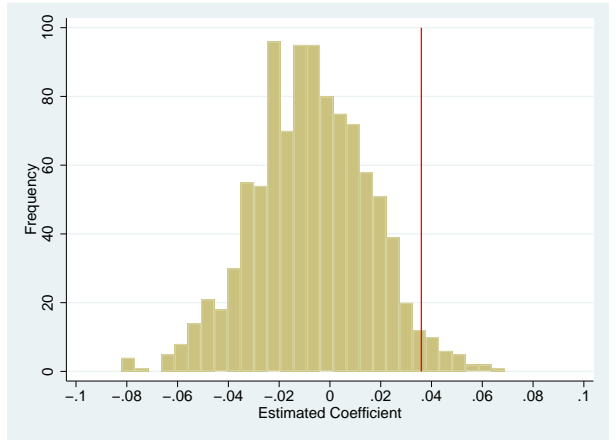
(a) Females:  $MN$



(b) Males:  $MN$



(c) Females:  $FN$



(d) Males:  $FN$

Note: These figures show distribution of coefficients obtained from the final OLS specification in Table 7 while replacing  $MN$  (respectively,  $FN$ ) with the value of  $MN$  (respectively,  $FN$ ) from a random grade in the same school. Red line represents actual estimate obtained in specifications (5) and (11) in Table 8.

## 13 Appendix

### 13.0.1 Add Health: Sample selection

Of the over 15,000 students in grades 7-12 in Wave I and followed through Wave IV, we drop about 4000 students for whom we cannot match information on peer characteristics (because they cannot be identified in both the in-home and in-school survey). We also drop those in a male-only school (58 students), and students without information on our main variables (7 students). We also drop 42 7th and 8th graders from a school that doubles in size between 8th and 9th grade, because the smaller junior high and larger high school populations may be different in characteristics. Lastly, we drop 272 individuals in grades with fewer than 20 students in a grade and a school with only one grade left at the end of this procedure (56 students).

50 schools have grades 9-12, 15 have grades 7-12, 43 have grades 7-8, and the remaining 10 schools have other mixes of grades. There is 1 school that has grades 7-12, but we drop grades 7-8 (see above) and are left with grades 9-12.

### 13.0.2 Imputing Post-College for Parents

We use the in-school survey which records the student's response to the highest level of education attained by their residential father and residential mother and create a dummy variable  $PC_i$  for student  $i$  that takes the value one if either the residential mother or the residential father of student  $i$  has a post-college education, i.e., obtained any education beyond a four-year college degree, and takes the value of 0 otherwise. If a student either does not have a residential father/mother or the information is missing, we impute that parent's level of education using the other parent's education. For example, if the residential father's education is missing, but the residential mother has a high-school education, we impute a value for father post-college by taking the average value of father post-college among students of the same gender within the school who also have a residential mother with a high-school education. If there are no students with equivalent mother's education and non-missing information on father's education, we impute father post-college using the value of father post-college among all students in the school who have a residential mother with a high-school education.<sup>62</sup>

### 13.0.3 Identification Tests for $MN$ and $FN$

Appendix Table 2 provides tests of balance as described in Section 4 of the paper for  $MN$  and  $FN$ . They reveal that there is sufficient residual variation to estimate the effects

---

<sup>62</sup>For the 32 cases for which this procedure didn't work (because there was no other parent with the same education), we impute the missing parent's level of education at the same level as the non-missing parent.

of  $MN$  and  $FN$  and that this residual variation is not strongly correlated with individual characteristics.

Appendix Figure 1 shows the equivalent to Figure 1 for  $MN$  and  $FN$  for females (panels a and c) and males (panels b and d). Again, close to 90 percent of our schools have a standard deviation of  $MN$  and  $FN$  within the 90 percent confidence interval obtained from our simulations for both males and females. Specifically, 90 percent (females) and 89 percent (males) have a standard deviation of  $MN$  falling within the estimated 90 percent confidence interval, and 90 percent (females) and 90 percent (males) have a standard deviation of  $FN$  falling within the estimated 90 percent confidence interval.

# 14 Appendix Tables and Figures

Appendix Table 1: Balance Tests,  $MN$  and  $FN$

Panel A, Females						
	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months
IHS MN	-0.0143 (0.057)	-0.462 (0.822)	0.000 (0.014)	-0.030 (0.028)	0.015 (0.021)	0.232 (0.554)
IHS FN	0.003 (0.038)	-0.464 (0.735)	0.017 (0.012)	-0.023 (0.026)	0.023* (0.013)	-0.197 (0.372)
Own Parent Post College	0.310*** (0.036)	5.170*** (0.715)	-0.009 (0.007)	-0.018 (0.019)	0.004 (0.019)	-1.133*** (0.295)
Panel B, Males						
	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months
IHS MN	0.009 (0.101)	0.859 (1.033)	0.022 (0.017)	0.011 (0.042)	0.026 (0.023)	-0.626 (0.562)
Own Parent Post College	0.278*** (0.043)	4.200*** (0.616)	-0.014 (0.010)	-0.007 (0.019)	0.029* (0.015)	-1.170*** (0.293)
IHS FN	-0.012 (0.040)	-0.208 (0.780)	-0.015 (0.016)	-0.046* (0.024)	0.011 (0.012)	0.147 (0.449)

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of  $MN$  and  $FN$  on individual characteristics. The estimates displayed in each row are for separate regressions in which the dependent variable is the variable name in the column and the independent variable is displayed in the row.  $MN$  (respectively,  $FN$ ) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent) as defined in the text and IHS is the inverse hyperbolic sine transformation. The regressions of  $FN$  in Panel A and  $MN$  in Panel B include a control for whether the individual has at least one post-college parent. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. All regressions are unweighted. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$



Appendix Table 2: High Achievers and Confidence and Motivation

	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
	Very Intelligent	College Likely	Want College	Very Intelligent	College Likely	Want College
MF	-0.288 (0.358)	-0.485 (0.322)	-0.545* (0.297)	0.251 (0.426)	-0.589 (0.399)	0.507 (0.456)
FF	-0.125 (0.407)	0.113 (0.335)	0.289 (0.335)	0.802** (0.404)	0.785 (0.481)	-0.015 (0.380)
PVT Score	0.649*** (0.094)	0.433*** (0.076)	0.317*** (0.072)	0.818*** (0.084)	0.120 (0.086)	0.226** (0.089)
Fraction Female	0.004 (0.229)	-0.064 (0.275)	-0.127 (0.282)	-0.002 (0.283)	-0.324 (0.328)	-0.174 (0.331)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5642	5636	5641	4695	4691	4694
$R^2$	0.167	0.224	0.144	0.212	0.219	0.191
Adjusted $R^2$	0.125	0.185	0.102	0.164	0.171	0.142

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of measures of confidence and motivation on individual and peer characteristics. “Want College” equals 1 if the student reports that they want to go to college as a 5 on a scale of 1-5 and equals 0 otherwise; “College likely” equals 1 if the student says the likelihood that they go to college is a 5 on a scale of 1-5 and equals 0 otherwise. “Very intelligent” equals 1 if the student reports that their intelligence level is “moderately above average” or “extremely above average” relative to others their own age and equals 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father’s education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ .

Appendix Table 3: High Achievers and Risky Behaviors

	Females					Males										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Cigarettes	Alcohol	Drunk	Binge	Marijuana	Fight	Arrest	Unpr Sex	Cigarettes	Alcohol	Drunk	Binge	Marijuana	Fight	Arrest	Unpr Sex
MF	1.004** (0.414)	0.315 (0.307)	0.656* (0.395)	0.433 (0.322)	0.267 (0.249)	0.368 (0.247)	-0.031 (0.124)	0.202 (0.342)	0.618 (0.428)	0.184 (0.330)	-0.217 (0.359)	-0.183 (0.357)	-0.029 (0.303)	-0.532 (0.470)	-0.206 (0.238)	-0.126 (0.197)
FF	-0.088 (0.410)	-0.718* (0.397)	-0.824** (0.361)	-0.491 (0.352)	-0.151 (0.266)	-0.106 (0.248)	0.201* (0.108)	0.720** (0.342)	-0.752 (0.522)	-0.518* (0.295)	-0.454 (0.351)	-0.650 (0.458)	-0.354 (0.377)	-0.902* (0.491)	-0.462* (0.269)	-0.189 (0.272)
PVT Score	-0.209** (0.088)	-0.091 (0.083)	-0.097 (0.086)	-0.120 (0.080)	-0.055 (0.052)	-0.261*** (0.062)	0.021 (0.032)	-0.129** (0.055)	-0.086 (0.088)	0.185** (0.091)	0.042 (0.066)	0.013 (0.072)	0.012 (0.053)	-0.098 (0.083)	0.004 (0.055)	-0.139** (0.065)
Fraction Female	-0.058 (0.264)	-0.324 (0.218)	-0.156 (0.280)	-0.247 (0.257)	0.061 (0.155)	-0.704*** (0.206)	0.012 (0.091)	0.052 (0.222)	0.782* (0.413)	0.202 (0.339)	0.022 (0.350)	-0.013 (0.365)	0.044 (0.214)	-0.036 (0.295)	-0.140 (0.198)	-0.078 (0.174)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5630	5626	5612	5608	5594	5629	5628	5598	4678	4681	4652	4656	4626	4676	4666	4642
R <sup>2</sup>	0.166	0.198	0.191	0.171	0.135	0.156	0.122	0.143	0.173	0.202	0.240	0.261	0.157	0.156	0.159	0.149
Adjusted R <sup>2</sup>	0.124	0.158	0.151	0.130	0.091	0.113	0.078	0.100	0.122	0.154	0.193	0.216	0.106	0.105	0.108	0.097
Dep Var Mean	0.586	0.563	0.275	0.237	0.121	0.218	0.027	0.126	0.574	0.565	0.279	0.275	0.135	0.406	0.111	0.102

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of measures of risky behaviors on individual and peer characteristics. The dependent variables are defined as follows: “any cigarettes” equals 1 if the individual has ever smoked cigarettes and equals 0 otherwise; “any alcohol” equals 1 if the individual has ever had more than a “couple of sips” of alcohol and equals 0 otherwise; “drunk” equals 1 if the individual reports being drunk in the past year and equals 0 otherwise; “binge drinking” equals 1 if the individual has had 5 or more drinks “in a row” in the past year and equals 0 otherwise; “any marijuana” equals 1 if the individual has smoked any marijuana in the past 30 days and equals 0 otherwise; “fight” equals 1 if the individual reports getting in a “serious physical fight” in the past year and 0 otherwise; “arrest before 18” equals 1 if the individual was arrested before age 18 and 0 otherwise; and “unprotected sex,” equals 1 if the individual did not use any form of birth control the most recent time she had sex and 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent) as described in the text. All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father’s education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. Coefficient on PVT score multiplied by 100. Wave IV weights used. Standard errors clustered at the school level. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01.

Appendix Table 4: Confidence and Risky Behaviors with Controls for PVT Rank

	Females				Males			
	(1) Confidence Index	(2) Risky Index 1	(3) Risky Index 2	(4) Birth Before 18	(5) Confidence Index	(6) Risky Index 1	(7) Risky Index 2	(8) Birth Before 18
MF	-1.283* (0.676)	1.325* (0.685)	0.628 (0.760)	0.457*** (0.161)	0.113 (0.912)	0.170 (0.720)	-0.874 (0.786)	0.111 (0.104)
FF	0.413 (0.707)	-1.440** (0.680)	1.261** (0.602)	0.343* (0.185)	1.378* (0.787)	-1.597** (0.698)	-1.974** (0.913)	-0.179** (0.082)
PVT Score	0.471 (0.375)	0.913*** (0.341)	0.210 (0.522)	0.027 (0.097)	0.261 (0.481)	0.777* (0.441)	0.150 (0.422)	-0.099 (0.073)
Fraction Female	-0.340 (0.542)	-0.064 (0.511)	-0.607 (0.562)	0.024 (0.121)	-0.568 (0.658)	0.626 (0.789)	-0.202 (0.689)	0.104 (0.125)
PVT Rank	0.352** (0.153)	-0.540*** (0.148)	-0.255 (0.182)	-0.068* (0.040)	0.309 (0.190)	-0.303* (0.179)	-0.185 (0.171)	0.016 (0.030)
School, Grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5630	5512	5512	5547	4685	4505	4505	4610
$R^2$	0.237	0.239	0.175	0.131	0.246	0.293	0.191	0.154
Adjusted $R^2$	0.199	0.200	0.133	0.087	0.200	0.248	0.140	0.102

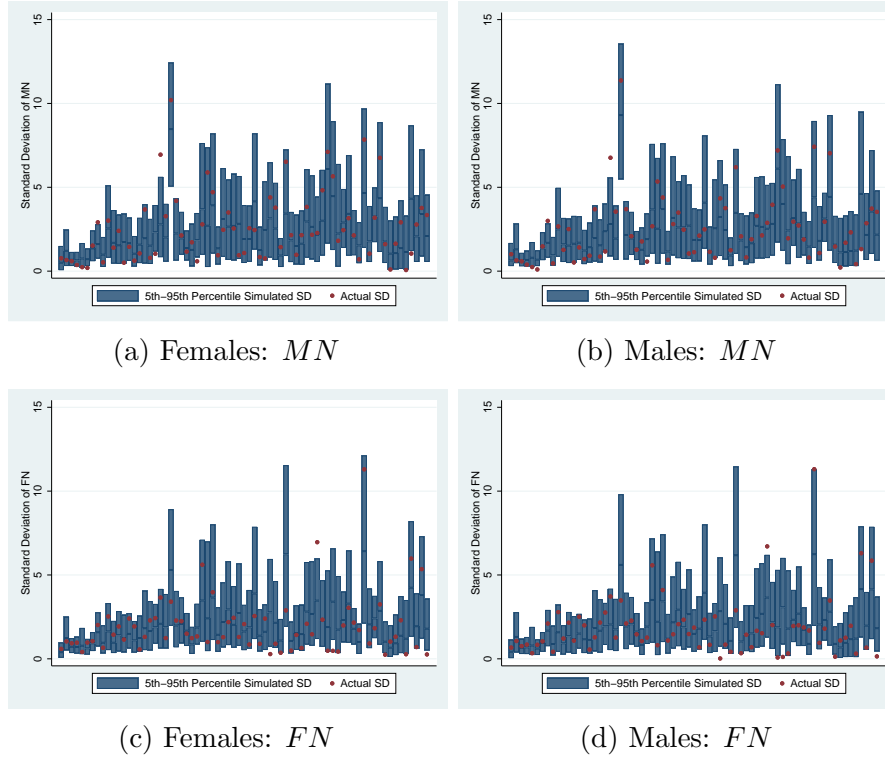
Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of measures of confidence and motivation and risky behaviors on individual and peer characteristics. The Confidence Index is the first factor from a factor analysis of three variables measuring self-perceptions of intelligence, desire to go to college, and likelihood of going to college. The Risky Index 1 (respectively, 2) is the first (respectively, second) factor from a factor analysis of 8 variables measuring risky behaviors. First Birth Before 18 takes a value of 1 if the individual has had a child by the time she turns age 18 and 0 otherwise. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent). All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother’s and father’s education (dummies for high school, some college but no degree, college degree, post college for each parent), and log family income. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races. Coefficient on PVT score multiplied by 100. The PVT percentile rank of the student is calculated by taking the absolute rank of each student relative to others in her grade and school in the in-home sample (with the worst-performing student having a value of 1) and then converting into a percentile. Wave IV weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Appendix Table 5: Sample Attrition

	Dependent Variable: In Wave IV Sample			
	Females (1)	(2)	Males (3)	(4)
MF	0.216 (0.259)		0.333 (0.310)	
FF	0.339 (0.248)		-0.503 (0.433)	
MN		0.029 (0.027)		0.028 (0.032)
FN		0.023 (0.022)		-0.052 (0.032)
PVT Score	0.076 (0.057)	0.076 (0.057)	0.162** (0.068)	0.163** (0.068)
Fraction of Peers who are Female	0.099 (0.178)	0.090 (0.172)	0.341 (0.256)	0.457* (0.274)
School, Grade FE	Yes	Yes	Yes	Yes
School Linear TT	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Peer Characteristics Controls	Yes	Yes	Yes	Yes
Observations	6878	6876	6250	6250
$R^2$	0.152	0.146	0.144	0.145

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of being in the Wave IV sample (conditional on being in Wave I) on individual and peer characteristics. *MF* (respectively, *FF*) is the fraction of male (respectively, female) “high achievers” (those with at least one post-college parent) as described in the text. All columns include a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Individual controls include race dummies (Black, Latino, Asian, and other races), age in months, mother and father’s education (dummies for each parent for high school, some college but no degree, college degree, and post college), and log family income. If mother’s (respectively, father’s) education is missing, all mother’s (respectively, father’s) education dummies are set to zero and a dummy is included for missing mother’s (respectively, father’s) education. If family income is missing, family income is set to the mean value for the school and a dummy is included for missing family income. Peer characteristics controls include fraction foreign born, Black, Latino, Asian, and other races in columns (1) and (3) and the IHS transformation of the count of peers who are foreign born, Black, Latino, Asian, and other races in columns (2) and (4). Wave I weights used. Standard errors clustered at the school level. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ .

Appendix Figure 1: Monte Carlo Estimates of  $MN$  and  $FN$



Note: These figures display simulated and actual standard deviations for schools in the sample with at least three grades, with each bar representing a different school. Upper and lower edges of the bar represent the 5th and 95th percentile respectively of the simulated within-school standard deviation of  $MN$  (or  $FN$ ). The dot represents the empirical standard deviation.