The Asymmetric Gender Effects of High Flyers*

Angela Cools
Raquel Fernández†
Eleonora Patacchini

This version: February 2021

Abstract

Using longitudinal information on a representative sample of U.S. students, we study the effects of exposure to female and male "high flyers" in high school. We identify a causal effect by exploiting quasi-random variation to peers with highly-educated parents across grades within a school. Greater exposure to male high flyers decreases the likelihood that women obtain a bachelor's degree, lowers their math and science grades, decreases their LFP and increases fertility. They show lower levels of self-confidence/aspirations. The effects are found for girls with below median ability and for those with at least one college-educated parent. There are no effects of high flyers of either gender on boys.

Key words: Gender; education; cohort study; high flyers; peers
JEL classification: I21, J16

*Cools: Davidson College; Fernández: NYU, NBER, CEPR, ESOP, and IZA; Patacchini: Cornell University, EIEF, CEPR, and IZA. Fernández thanks the CV Starr Center for financial support. This is a substantially revised version of an earlier working paper. We thank for helpful comments seminar audiences at the Universidad de San Andres, the CEPR Gender Economics Series,the 8th Development Conference of Santiago, Chile, the ESOP conference of Oslo University, the Gender and Labor Market Conference at U. of Chicago, the SEHO Conference, the Catalan Economic Society Conference, CUNY, Wake Forest University, and the SEA Conference.

†Corresponding author. Mailing Address: Department of Economics, New York University. 19 W. 4th Street, New York, NY 10012. Phone: (212) 998-8908. Fax: (212) 995-4186 E-mail: raquel.fernandez@nyu.edu

1 Introduction

A large literature has documented the importance of peer effects in education. These studies have mostly focused on short-run outcomes and how these are affected by various measures of *average* peer quality. Our paper contributes to the literature by studying long-run educational outcomes by gender for a US representative sample of high-school students. Our analysis generates an important and novel finding: male "high flyers" (the term we use to denote students who can be expected to perform very well academically) have a significant negative effect on the propensity of their high school female peers to go on to obtain a college degree. We show that these results stem from male high flyers rather than from exposure to boys who on average perform better. Furthermore, there is an asymmetric gender effect: female high flyers do not affect their male peers.

Our investigation focuses on the exposure to high flyers at a particularly impressionable period of life: high school. We use 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 parent has some *post-college* education – to proxy for a bundle of characteristics that predict that a student will do very well in school. Aggregating this feature 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 flyers of each gender. The main question we ask is how does exposure to high flyers affect long-run human capital. We find a very strong asymmetric gender effect: male high flyers have a statistically and economically significant negative effect on the probability that girls will obtain a bachelor's degree but do not affect boys. Female high flyers, on the other hand, do not affect college outcomes for either gender. Probing further, we show that the negative effect of male high flyers is concentrated in the lower half of the female ability distribution and among those with more-educated parents. Girls exposed to a higher proportion of male high flyers tend to have lower math and science grades in high school and to substitute away from a four-year college degree into a two-year college degree. Furthermore, they have lower labor force participation and higher fertility by the ages of 26-32. Although our data does not allow us to pin down the precise pathway underlying our results, we show that a higher proportion of male high flyers is associated with lower self-confidence/aspirations in girls, reminiscent of the asymmetric gender effects from competition found by Niederle and Vesterlund (2007) and Bordalo, Coffman, Gennaioli, and Shleifer (2019), among others.

Our paper is related to several strands of a literature that study the interaction of peer effects and gender broadly speaking.[1] One strand examines how the proportion of own-gender students affects various outcomes (e.g., Hoxby (2000a); Lavy and Schlosser

[1]See, e.g., Sacerdote (2014) and Cools and Patacchini (in press).

(2011); Black, Devereux, and Salvanes (2013)). Another strand studies how ability composition/peer quality affects academic outcomes (e.g., Carrell, Fullerton, and West (2009); Imberman, Kugler, and Sacerdote (2012)). There are only a handful of papers that, like ours, are concerned with the potentially asymmetric gender effects from high-performing peers. Mouganie and Wang (2020) study high-school students in China and find that high-performing male peers (as defined by their national exam scores in mathematics) reduce the likelihood of girls choosing a science track (relative to an arts track) for the remainder of high school; high-performing female peers have the opposite effect. Exploiting the shuffling of students between primary and secondary schools in England along with subject-specific differences in peer ability, Lavy, Silva, and Weinhardt (2012) find instead that high-achieving male peers (those scoring in the top five percent on national standardized tests) have no significant impact on girls' test scores but negatively impact boys' scores; high-achieving female peers affect girls' test scores positively and have no effect on boys' scores. Feld and Zölitz (2019) exploit the random assignment of first-year students within compulsory courses to teaching sections in a Dutch business school. They show that male peers with a higher average pre-assignment GPA are associated with men taking more mathematical courses; women choose to take fewer mathematical courses and are less likely to choose a mathematically intensive major. Relative to this literature, our contribution is to study longer-run outcomes and to do so for a representative sample of US high schools. Furthermore, we are able to explore the channels by which these effects occur by including measures of self-confidence/aspirations. Lastly, we show that the negative effects stem from those students that we call high flyers rather than from high average quality.

Our paper is also related to those that study the long-run impacts of peers' parental education. Bifulco, Fletcher, and Ross (2011) and Bifulco, Fletcher, Oh, and Ross (2014) use Add Health and find that peer quality, as measured by percentage of students whose mother completed college, improves high-school completion and college attendance but has no effect on bachelor's degree attainment. We show that these results stem, however, from ignoring gender, i.e., from not differentiating between male and female peers with college-educated mothers and from not allowing the effect of these peers to differ by the gender of the student (see Table 5). Furthermore, in a horse race between their variable (which we now gender) and our high flyer measure, we show that only the latter affects long-run education outcomes.

Our paper proceeds as follows. In Section 2, we present the data and sample selection. In Section 3, we detail our empirical model and identification strategy. Section 4 is devoted to the main regression analysis. Section 5 examines heterogeneity in the results, explores pathways, and studies other long-run outcomes.

2 Data and Sample Selection

This analysis uses data from Add Health, 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 (Wave I). Every student on the school roster was asked to complete an in-school questionnaire which included basic questions about the student's demographics and about the characteristics (including educational attainment) of both parents. A randomly selected subsample of about 20,000 students was also interviewed at home, where questions about attitudes and other sensitive topics were asked. 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. Interviewers also issued a survey to the student's residential parent, asking questions about attitudes and family income, among other topics. 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. 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 Wave I, the only wave in which the students' peers were surveyed.

After dropping students who were not followed through Wave IV and those who cannot be matched with peer characteristics, we also eliminated particular grades/schools (e.g., an all-male school or 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.[2] Summary statistics for the Wave IV-weighted sample are reported by sex in Table A1 in the Appendix. As shown, the Add Health sample is broadly similar to the U.S. population as calculated from the ACS.

3 Empirical Model and Identification Strategy

Girls in a grade with a larger proportion of high-performing boys may, ceteris paribus, form a biased gender stereotype that leads them to believe that boys tend to be smarter. Alternatively, girls may be less willing to compete with boys when a larger fraction of them perform exceptionally well. They may react the same way to high-performing girls or they may view competing with them in an altogether different manner. The same possibilities arise for boys: they could perform worse when faced with a greater proportion of high-performing girls or boys. This is, in the end, an empirical question. We explore this question by examining how girls' versus boys' outcomes differ when faced

[2]Table A7 in the Appendix examines whether attrition is a function of the main variables introduced in the next section and concludes that this is not a problem.

with high-performing peers of each gender.

To avoid the reflection problem, we use high parental education as an indicator of whether a student is what we term a "high flyer." Parents' final educational attainment is usually determined before a student enters high school (thus avoiding a two-way interaction between student and parental achievement) and there is a strong, positive relationship between an individual's academic achievement and the educational attainment of their parents. We call a student a "high flyer" if at least one parent has a post-college education (i.e. obtained any professional training beyond a four-year college). We choose post-college education rather than, say, college, because the former is more strongly predictive of high student achievement, for both sexes, than other levels of parental education (this is formally shown in Table A2 of the Appendix). In Section 4 we explore alternative ways to proxy for peers whose background characteristics predict high performance at school.[3]

We estimate the following regression model:

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

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 at least a bachelor's degree. $MaleFrac\_high_{igs}$ (respectively, $FemaleFrac\_high_{igs}$) 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 $MaleFrac\_high_{igs}$ and $FemaleFrac\_high_{igs}$ 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 equal 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. α_g is a grade fixed effect, β_s is a school fixed effect, and $\delta_s g$ is a school-specific linear time trend. 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. Note that all peer variables (e.g., percent from a given ethnicity) are calculated in the same leave-one-out manner. Finally, $\varepsilon_{i,t+1}$ are i.i.d., mean 0 innovations.

Our empirical strategy exploits idiosyncratic variation in exposure to high flyers across different cohorts of high-school students within a given school, a common approach in the literature.[4] The grade fixed-effects control for initial differences across cohorts whereas the

[3] It should be noted here that we cannot control for peers' parental income. More generally, a post-college parent may impart other characteristics to a child in addition to high academic performance, and these may play a role in generating peer effects. Note that we can control for parental income for the students in our sample.

[4] See, e.g., Hoxby (2000a, 2000b); Angrist and Lang (2004); Gould, Lavy, and Paserman (2009); Lavy

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 required is 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. The school linear time trend deals 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. 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. As shown in Table A3 in the Appendix, there is sufficient residual variation in the key peer variables after removing fixed effects and the time trend.

We investigate the validity of the identification strategy, following Lavy and Schlosser (2011), by examining whether the variation in the key peer variables is related to variation in a number of predetermined student characteristics. These regressions include a dummy for whether the student's own parents are high flyers whenever the peer variable is of the same gender as the student (e.g., all regressions for girls in the balance regressions of Table 1 include a dummy for whether their own parents are high-flyers when the female high-flyer peer variable is used) as not doing so mechanically introduces a correlation between the same-gender peer variable and a series of outcomes.[5] As shown in Table 1, 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.[6] A Monte Carlo simulation in the Appendix also lends support to this hypothesis.[7]

4 Results

Table 2 reports the estimation results of the model in equation (1), with the completion of a bachelor's (four-year college) degree as the dependent variable. Standard errors are clustered at the school level. The first five columns are for girls; the last five are for boys.

Column (1) includes only grade and school fixed effects and a school-level linear time trend. Column (2) adds individual controls and parental background. We include age in

and Schlosser (2011); Lavy, Paserman, and Schlosser (2011); Olivetti, Patacchini, and Zenou (2020); and Merlino, Steinhardt, and Wren-Lewis (2019).

[5] See Guryan, Kroft, and Notowidigdo (2009).

[6] 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 Table 1 supports the hypothesis that the model specification identifies an exogenous source of variation.

[7] We test whether the variation in the main variables is "as good as random" by performing Monte Carlo simulations. Figure A1 in the Appendix shows that around 90% of schools have a standard deviation of $MaleFrac\_high$ and $FemaleFrac\_high$ within the 90 percent confidence interval of the simulated deviation when those shares are random draws for both sexes.

months, an indicator for foreign born, race, the student's PVT score, parental education, an indicator for mother or father not in the household, and the log of household income.[8] Note that conditional on student grade and school, age in months is negatively associated with attainment of a bachelor's degree which may 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. Column (3) includes the PVT percentile rank of the student in addition to their PVT score, as this has been shown to matter in the literature.[9] The PVT rank is positively and significantly associated with college graduation for women (but not for men which may be a function of sample size over which rank is calculated). The coefficients on the main variables are unaffected by its inclusion. In column (4) (column (9) for men), we include the fraction of female peers 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 their long-run education outcomes but decreases them for boys. In column (5) (column (10) for men), we include other peer variables: the fractions foreign born, Black, Asian, Latino, and other races.

Across specifications there is a great deal of stability in the magnitude of the estimated coefficients on all variables. A greater fraction of male high flyers decreases the likelihood of a girl obtaining a bachelor's degree. The effect is sizable: using the most complete specification (column (5)), an increase in $MaleFrac\_high$ by a standard deviation (2.0 percentage points) is associated with a 2.2 percentage point decrease in the probability that girls obtain a four-year college degree by Wave IV, a 6% decrease from the mean of 35%. The effect of female high flyers, on the other hand, while positive is statistically insignificant. The results are radically different for boys. Neither male nor female high flyers have a statistically significant effect on boys.

An important potential concern is that variation in the fraction of "high flyers" could be driven by variation in the number of dropouts across grades in a school. To address this concern Table 3 repeats the analysis using the *number* of high flyers rather than the *fraction* of these. As shown, a greater number of male high flyers reduce females'

[8] Recall that all these characteristics are from Wave 1 and not contemporaneous with the outcomes of Wave 4.

[9] Percentile 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. We follow Elsner and Isphording (2017), and first calculate the rank of each student relative to others in her grade and school sample (with the worst-performing student having a rank of 1) before converting into a percentile to render the measure comparable across samples of different sizes. Thus, the percentile rank ranges from 0 (the worst student) to 1 (the best student).

college attainment. This eliminates the concern that a higher fraction of male high flyers is a consequence of a grade previously having many male dropouts (presumably with less-educated parents) which would then lead to a very different interpretation of the results. In fact, the correlation between $MaleFrac\_high$ and the number of boys is 0.13. If the variation were driven by higher dropout rates of marginal boys, we would expect a negative relationship between these variables.

4.1 Robustness and Other Concerns

To increase confidence in the estimation results, we use a simulation exercise to study the likelihood that a result of this magnitude could have occurred by chance (see Athey and Imbens (2017)). We generate randomness in the exposure to high flyers by assigning to each individual the residual from a random grade within the same school (obtained by regressing $MaleFrac\_high$ on a school fixed effect and linear time trend), and adding it to the predicted value for that individual's true grade. The equivalent procedure is used to obtain placebo values for $FemaleFrac\_high$. All other variables (both own and of one's peers) are kept at their true values. Figure A2 shows the estimated coefficients of the regressions when the placebo variables are used and the procedure is repeated 1,000 times.[10] The share of estimates that is larger in absolute value than the actual treatment represents the randomization-based p-value. Only two of the estimates of $MaleFrac\_high$ are larger in absolute value than the the actual treatment coefficient (represented by the line), providing evidence that it is highly unlikely to have occurred by chance.[11] The share of estimates of $FemaleFrac\_high$ larger in absolute value than the line (actual treatment), on the other hand, is 0.39. For boys, the share of estimates larger in absolute value than the estimated coefficient for $MaleFrac\_high$ is 0.76 and it is 0.63 for $FemaleFrac\_high$, indicating their lack of statistical significance. Appendix Table A5 shows that the results are robust to the exclusion of outliers and the removal of the school trend.

Angrist (2014) raises two concerns vis a vis the peer literature. First, as alluded to earlier, there is a mechanical negative correlation between own and peer characteristics as a result of the leave-one-out strategy. Note that this is not a concern for opposite-gender peer effects as, for example, all girls in a given grade and school face the same fraction of male high flyers. As discussed previously, we adjust for the same-gender correlation by including a dummy for own-parent post-college status in the balance tests. We also control for mother and father's education in all other regressions.[12] Alternatively, following Carrell, Hoekstra, and Kuka (2018), we can remove the mechanical correlation by restricting the sample to those students without post-college parents. Table A5 shows that results are robust to this adjustment. A second concern raised by Angrist (2014), and later elaborated

[10]Further details on the methodology can be found in the Appendix.

[11]The p-value is 0.002.

[12]The results are robust to using instead a dummy for parent post-college rather than including mother's and father's education separately.

on by Feld and Zölitz (2017), is that classical measurement errors in the peer variable of interest can lead to an overestimation of peer effects when peers are not randomly assigned. As shown in Table 1, our analysis exploits pseudo-random variation in peer characteristics. Furthermore, Feld and Zölitz (2017) show, both analytically and using Monte Carlo simulations, that in settings where peers are randomly assigned, classical measurement error will only lead to attenuation bias. This is in fact what we see. Following the procedures used by Feld and Zölitz (2017) and Carrell et al. (2018), we add progressively higher levels of measurement error to $MaleFrac\_high$.[13] As shown in Figure A3 in the Appendix, adding measurement error attenuates the estimates for girls and leaves the estimates for boys insignificant. Our results are therefore highly unlikely to be due to measurement error in the data.

4.2 Alternative Measures of Peer Ability

We conclude this section by examining how the main results are affected by alternative measures of peer ability.[14] As mentioned previously, many studies have studied the effects of *average* peer ability. An important question to ask, therefore, is whether the results are driven by peers with a high average or by high flyers. The most natural measure of peer quality in our setting is to use parental years of education to construct a measure of peer quality.[15] We do this, by student gender, in the usual leave-one-out fashion as all the other peer variables. We denote these averages by $MaleAvg$ and $FemaleAvg$. It is worth noting that the Pearson correlation between average parental education and high flyers is relatively low: 0.10 for males and 0.09 for females, net of school fixed effects and time trend. There is therefore still plenty of variation in the key variables: controlling for average parental education decreases the standard deviation of the key variables only slightly to 1.6 percentage points.

Table 4, columns (1) and (3), report the results using the most complete specification but replacing the key variables with $MaleAvg$ and $FemaleAvg$. The results show a negative relationship between $MaleAvg$ and females' long-run educational attainment. By contrast, we find a positive role of female peers' average parental education. There is no effect of either variable on boys' likelihood of completing a bachelor's degree. As shown in column (2), however, when the two sets of key peer variables are included only the

[13]Specifically, for a given amount of measurement error on $MaleFrac\_high$ (e.g., of z percent), we replace z percent of boys' parents' post-college status in the in-school sample with a 0 or 1 based on a binomial distribution with a mean of 14.4 percent (the overall in-school sample mean). We then re-calculate $MaleFrac\_high$ for each individual and implement the regression in columns (5) and (10) of Table 2, keeping all other variables at their true levels.

[14]Appendix Table A6 shows the results from using alternative definitions of the main peer variables (both parents with post college education, only mother post college, or only father post college).

[15]Average parental years of education are computed by assigning 9 years to less than high school degree, 12 years to a high school degree, 14 years to some college, 16 years to a bachelor's degree, and 18 years to a post-college education. If a parent's education is missing or the parent is not in the household, we impute it using the other parent's educational attainment (see the Appendix for details).

proportion of male high flyers and not the male average decreases girls' long-run educational attainment. Moreover, the coefficients on the two competing male peer variables are different from one another at the 1 percent level. $FemaleFrac\_high$ is insignificant but, interestingly, $FemaleAvg$ remains significant, indicating that a higher female average has a positive effect on their female peers' outcomes. There is no statistically significant impact of any of these variables on boys' educational achievement.

Lastly, we can compare our results with those of Bifulco et al. (2014). Using the same data source, they study the effect of peers with college-educated mothers on long-run bachelor's degree attainment. They do not differentiate either the sample or the peer variable by gender. They find that these peers have no long-run consequences on educational attainment. Column 1 in Table 5 replicates their analysis on our entire sample (where $MotherFrac\_Coll$ is the proportion of peers with college educated mothers), using a female dummy for gender but otherwise ignoring the gender of both the peers and the students. We obtain the same result as they did: there is no significant effect on education from peers with college-educated mothers.

We next ask whether this is a consequence of not differentiating the sample by gender. As can be seen in columns (2) and (5), it is not. Next we examine whether gendering the peer variable matters, i.e., differentiating between the proportions of boys vs girls with college-educated mothers ($MaleMotherFrac\_Coll$ vs $FemaleMotherFrac\_Coll$). As can be seen in column (3), the fraction of male peers with college-educated mother has a negative and statistically significant effect on the college outcome of girls; it has an insignificant effect on boys. From this we can conclude that Bifulco et al.'s null result is a consequence of not gendering the key peer variables and the sample. Had they done so, they would have uncovered the negative effects of these boys.

As a last instructive exercise, column (4) includes the high-flyers peer variable alongside the college-educated mothers peer variable. Only the high flyers have a statistically significant effect. Furthermore, these coefficients are statistically significantly different one another. Note that in all cases the peer variables do not affect boys' educational outcomes.

5 Heterogeneous Effects, Self-Confidence, and Further Outcomes

We next examine heterogeneity in the main results. We consider individual ability and family background since they are strongly correlated with the probability of graduating from college. We split the sample according to (i) individual PVT score (at-or-below the median versus above the median) and (ii) parents' education levels (neither parent with any kind of college degree versus at least one parent with degree).[16]

As can be seen in columns (1) and (2) of Table 6, higher levels of $MaleFrac\_high$ 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 $MaleFrac\_high$ (0.016 for girls with below-median PVT, 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 the coefficient is half the magnitude and statistically insignificant. A higher proportion of female high flyers, 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 flyers would have essentially a zero net effect on these girls. For girls with an above-median PVT score, there is no statistically significant effect from female high flyers. This result is consistent with research suggesting that lower-ability females may be particularly positively influenced by higher-performing friends.[17] As can be seen in columns (5) and (6), there is no statistically significant effect of the proportion of high flyers of either gender on boys.

Turning to parental education (columns (3) and (4)), it is clear that the negative impact of male high flyers is concentrated among girls with a college-educated parent; there is no effect on the other group of girls. The coefficient is twice the one we obtained previously for the entire sample. A one standard-deviation increase in $MaleFrac\_high$ (2.1 percentage points for this sample) leads to a 4.5 percentage point decrease in college completion, roughly a 7% decrease on a mean of 61%. This suggests that the negative impact is precisely on those girls who, from a family-background perspective, would be most likely to attend and graduate from college. The impact of $FemaleFrac\_high$, on the other hand, is positive for girls whose parents do not have college degree. For this group, a one standard deviation increase in $FemaleFrac\_high$ is associated with a 1.2 percentage point increase in the probability of obtaining a college degree.[18] Lastly, note that the equivalent sample split for boys (columns 5-8) once again show no significant effect of either male or female high flyers.

Table 7 replicates the exercises of Table 4 for girls with below-median PVT (upper panel) and either parent with a college degree (lower panel) to show that high flyers rather than high averages are responsible for the results. For below-median PVT girls, there is no significant relationship between average male parental education and their educational

[16] 48% of girls (38% of boys) with an above median PVT score obtain a bachelor's degree as do 61% of girls (48% of boys) with at least one college-educated parent.

[17] See Hahn, Islam, Patacchini, and Zenou (2020).

[18] We perform randomization-based inference tests indicating that the heterogeneity results are unlikely to be obtained by chance (we obtain p-values of 0.008 and less than 0.001 for girls with below-median PVT and those with at lesat one college-educated parent, respectively).

attainment. Male high flyers continue to have a significant, negative impact even after controlling for the male average.[19] For girls with a college-educated parent, column (1) shows a negative, significant impact of average male parental education. However, this peer variable becomes insignificant when male high flyers are included; the latter is negative and significant as usual.

Why might greater exposure to a higher proportion of male high flyers decrease girls' college attainment? One possibility is that they negatively affect 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 female high flyers. 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. Note that the effect must be gender specific as boys are not negatively affected.

In column (1) of Table 8 we repeat the main specification with the student's GPA in Wave I as the dependent variable, followed by the student's grades in math and science (columns 2 and 3). As shown, a greater proportion of male high flyers is associated with a negative but statistically insignificant impact on girls' GPA and a significant negative effect on their math and science grades (a one standard-deviation increase in $MaleFrac\_high$ decreases math and science grades by about 5 percent of a standard deviation). There is no statistically significant effect of either peer variable on boys' GPA or grades in math or science. These results are suggestive of findings in Bordalo et al. (2019). 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 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.

We next turn to psychological mechanisms such as self-confidence and aspirations/ambition more directly related to college attendance. Girls may become more discouraged or think themselves less competent in the presence of male high flyers. The richness of our data allows us to look more closely at this conjecture. We 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 a 5 and 0 otherwise. 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 5 and 0 otherwise. Finally, from the question "Compared with other people your age, how intelligent are

[19] It is interesting to note that average male parental education is for the first time negative and marginally significant for boys.

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 either "moderately above average" or "extremely above average" and 0 otherwise.[20] Since the above indicators of confidence and motivation are highly correlated, we use factor analysis to reduce the dimensionality of the dependent variables. 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. This factor explains 55% of the variance. By construction, the index has a mean of 0 and a standard deviation of 1. Table A8 in the Appendix shows the variance explained by the factors, their eigenvalues, and the factor loadings for factor 1. As shown, the measures of desire to attend college and perceived likeliness of attending college load most strongly.

Column (4) in Table 8 (column (12) for boys) reports the results of the main specification with the confidence index as the dependent variable. As shown in the table, male high flyers decrease girls self-confidence (a one standard deviation decreases their self confidence by about 3 percent of a standard deviation); it has no effect on boys. The proportion of female high flyers is positive but statistically insignificant for girls; it is positive and statistically significant for boys, a rather curious result.

Performing the same heterogeneity analysis as before but using the self-confidence index as the dependent variable, the second panel of Table 6 shows that the negative effect of male high flyers is on girls with below-median PVT scores and those with at least one college-educated parent. Recall that these same two groups of girls were negatively affected when the dependent variable was a bachelor's degree, suggesting that a loss in confidence/ambition may be a significant mechanism.[21] Lastly, note that boys with below median PVT and those with college-educated parents are the ones who are (positively) affected by $FemaleFrac\_high$.

Given that greater exposure to male high flyers decreases the probability that girls obtain a bachelor's degree, we next ask what level of education they attain. A natural alternative is a 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 lower than a bachelor's degree. In Table 8, column (5) (column (13) for boys), we use the most complete specification of Table 2 but with the dependent variable equal to 1 if the individual has a vocational or associate's degree as their highest level of education.[22] As

[20]In our sample, 75% of girls and 68% of males rate the amount that they want to go to college as a 5; 61% of girls and 50% boys answer 5 for college likely; 33% of females and 35% of males rate their intelligence as "moderately above average" or "extremely above average."

[21]Table A9 in the Appendix shows that male high flyers and not a higher male average is responsible for the decrease in girls' self-confidence/aspirations.

[22]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. The variable is coded as 0 if

shown, girls exposed to a greater proportion of male high flyers 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 magnitude of the coefficient is similar and the sign is opposite to that one obtained for the completion of a bachelor's degree. This indicates that male high flyers influence the decision about the type of degree to pursue rather than the decision to pursue any education after high school.[23]

We next turn to decisions in adulthood including labor force participation (LFP), marriage, and fertility. For LFP, 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 equal zero otherwise.[24] Marriage and fertility are based on questions to all sample respondents in Wave IV about whether they have ever been married and the total number of (non-deceased) biological children they have.

The results are shown in columns (6)-(8) (columns (14)-(16) for boys) of Table 8. A one standard deviation increase in $MaleFrac\_high$ (0.023) is associated with a 1.4 percentage point decline in female LFP. In contrast, a one standard deviation increase in $FemaleFrac\_high$ (0.021) is associated with higher LFP of a slightly smaller magnitude. There is no impact of male high flyers on a woman's probability of having ever been married by Wave IV, but a one standard deviation increase in $MaleFrac\_high$ (0.020) increases a woman's total number of biological children by 0.03, a 3 percent increase over its mean of 1.08. There is no effect of $FemaleFrac\_high$ on males' LFP or marriage, but a one standard deviation increase in $FemaleFrac\_high$ decreases a man's total number of biological children by 0.04, a 5% decrease on a mean of 0.72. The decrease in women's LFP and their higher fertility may be a consequence of their education outcomes, showing the longer-run effects of greater exposure to male high flyers.

6 Conclusion

Our analysis generates an important and novel result: exposure to male high flyers decreases the likelihood that women obtain a bachelor's degree, lowers their math and science grades and, in the longer run, decreases their labor force participation and increases fertility. The girls affected are those in the bottom half of the ability distribution and those with at least one college-educated parent. Those girls also show lower levels of self-confidence/aspirations. There is no effect on male outcomes from high flyers of either

the individual reports a degree from a vocational program lasting less than 1 year.

[23]We found no significant impact of "high flyers" on high-school completion rates for either girls or boys.

[24]Employment is based on the question: "are you currently working for pay at least 10 hours per week?". Individuals who report still working at their first full-time job are not asked this question so we code them as employed. For this question we exclude those in the military or prison in Wave IV and restrict the sample to individuals who were in 9th-12th grades in Wave I. This ensures that they are 28-32 years old by Wave IV, and thus likely to have completed all schooling.

gender. These findings highlight the importance of differentiating both sample and key peer variables by gender.

Our analysis raises the important and policy-relevant question of whether prior results in the literature are truly driven by high averages or by the upper tail as we find in our paper. Our findings indicate that a possible policy response to the negative effect of male high flyers on women's educational attainment is to have an equal (or greater) proportion of female high flyers. In the future, in addition to exploring the effect of high flyers vs high averages in other data sets and contexts, it would be of great interest to obtain data that allowed one to distinguish whether the negative effects on women are generated mainly by student interactions/perceptions, by how teachers react to these male students, or by how the parents interact with the teachers/school.[25]

[25]There is evidence that teachers reactions to boys vs girls depends on the socio-economic status of the male student. See, e.g., Musto (2019).

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. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

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., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, *109*(3), 739–73.

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.

Cools, A., and Patacchini, E. (in press). Peer Effects in Education. In *Handbook of the Economics of Education.* B. McCall (ed.) Routledge.

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.

Feld, J., and Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387–428.

Feld, J., and Zölitz, U. (2019). Peers from Venus and Mars: High-Achieving Men Affect Major Choices and Labor Market Outcomes. Working Paper.

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.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American*

Economic Journal: Applied Economics, *1*(4), 34–68.

Hahn, Y., Islam, A., Patacchini, E., and Zenou, Y. (2020). Friendship and Female Education: Evidence from a Field Experiment in Bangladeshi Primary Schools. *The Economic Journal*, *130*(627), 740–764.

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.

Lavy, V., Silva, O., and Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics*, *30*(2), 367–414.

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. (2020). High-performing peers and female STEM choices in school. *Journal of Labor Economics*, *38*(3), 805–841.

Musto, M. (2019). Brilliant or bad: The gendered social construction of exceptionalism in early adolescence. *American Sociological Review*, *84*(3), 369–393.

Niederle, M., and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Olivetti, C., Patacchini, E., and Zenou, Y. (2020). Mothers, peers, and gender-role identity. *Journal of the European Economic Association*, *18*(1), 266–301.

Sacerdote, B. (2014). Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics*, *6*(1), 253-272.

Tables

Table 1: Balance Tests for $MaleFrac\_high$ and $FemaleFrac\_high$

Panel A, Females	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months	Grade Size	# Girls	# Boys
MaleFrac_high	0.403	0.737	-0.150	-0.319	-0.002	4.699	-18.301	-37.340	19.039
	(0.596)	(6.665)	(0.111)	(0.250)	(0.200)	(3.689)	(59.790)	(38.077)	(42.149)
Observations	5899	5650	5899	5899	5899	5898	5899	5899	5899
Adjusted R^2	0.924	0.200	0.025	0.063	0.409	0.898	0.990	0.989	0.986
FemaleFrac_high	0.251	-4.175	0.136	-0.196	0.161	1.987	86.698	29.075	57.623
	(0.444)	(8.270)	(0.114)	(0.302)	(0.133)	(3.923)	(61.849)	(34.769)	(38.684)
Own Parent Post College	0.313***	5.172***	-0.009	-0.018	0.004	-1.105***	1.572	0.532	1.040
	(0.036)	(0.725)	(0.007)	(0.019)	(0.019)	(0.288)	(0.994)	(0.509)	(0.646)
Observations	5899	5650	5899	5899	5899	5898	5899	5899	5899
Adjusted R^2	0.925	0.213	0.025	0.063	0.409	0.898	0.990	0.989	0.986

Panel B, Males	Log Family Income	PVT Score	Mother Not in HH	Father Not in HH	Black	Age in Months	Grade Size	# Girls	# Boys
MaleFrac_high	-0.025	3.841	0.107	-0.243	0.329	-2.718	-21.939	-27.715	5.776
	(0.667)	(7.972)	(0.144)	(0.357)	(0.209)	(3.767)	(61.391)	(32.222)	(49.002)
Own Parent Post College	0.277***	4.189***	-0.014	-0.009	0.030*	-1.162***	-1.095	-1.003*	-0.092
	(0.042)	(0.615)	(0.010)	(0.019)	(0.015)	(0.295)	(1.077)	(0.596)	(0.749)
Observations	4954	4703	4954	4954	4954	4948	4954	4954	4954
Adjusted R^2	0.910	0.182	0.003	0.054	0.409	0.871	0.990	0.989	0.986
FemaleFrac_high	0.368	3.625	-0.086	-0.557**	0.157	-2.994	104.059	32.375	71.683
	(0.505)	(9.975)	(0.207)	(0.226)	(0.134)	(5.404)	(67.540)	(33.648)	(46.325)
Observations	4954	4703	4954	4954	4954	4948	4954	4954	4954
Adjusted R^2	0.909	0.174	0.003	0.055	0.409	0.871	0.991	0.989	0.986

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of individual or cohort characteristics on $MaleFrac\_high$ and $FemaleFrac\_high$. The estimates in the rows separated by the double lines are for separate regressions in which the dependent variable is the variable name in the column and the independent variables are displayed in the rows. $MaleFrac\_high$ (respectively, $FemaleFrac\_high$) is the fraction of male (respectively, female) high flyers (those with at least one post-college parent) in the grade and school. The regressions of $FemaleFrac\_high$ in Panel A and $MaleFrac\_high$ in Panel B include a control for whether the individual has at least one post-college parent. All columns include grade fixed effects, school fixed effects, and school linear time trends. 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 2: High Flyers and Bachelor's Degree Attainment

	Dependent Variable: Bachelor's degree									
	Females (1)	(2)	(3)	(4)	(5)	Males (6)	(7)	(8)	(9)	(10)
MaleFrac_high	-1.063***	-1.052***	-1.025***	-0.997***	-1.090***	0.028	0.290	0.292	0.226	0.129
	(0.294)	(0.276)	(0.269)	(0.296)	(0.317)	(0.428)	(0.387)	(0.389)	(0.365)	(0.365)
FemaleFrac_high	-0.258	0.176	0.199	0.184	0.291	0.553	0.307	0.307	0.323	0.202
	(0.385)	(0.357)	(0.348)	(0.365)	(0.338)	(0.551)	(0.523)	(0.527)	(0.502)	(0.493)
Foreign Born		0.089**	0.081**	0.089**	0.087**		0.102**	0.099**	0.101**	0.104**
		(0.041)	(0.040)	(0.040)	(0.040)		(0.045)	(0.045)	(0.045)	(0.046)
PVT Score		0.530***	0.233*	0.533***	0.535***		0.455***	0.211	0.450***	0.449***
		(0.065)	(0.133)	(0.065)	(0.064)		(0.071)	(0.168)	(0.071)	(0.070)
Age in Months		-0.006***	-0.005***	-0.006***	-0.006***		-0.005***	-0.005***	-0.005***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Mother HS Grad		0.060***	0.060***	0.060***	0.059***		0.007	0.007	0.009	0.009
		(0.022)	(0.022)	(0.022)	(0.022)		(0.021)	(0.021)	(0.021)	(0.021)
Mother Some College		0.115***	0.114***	0.115***	0.116***		0.086***	0.086***	0.088***	0.088***
		(0.026)	(0.026)	(0.026)	(0.026)		(0.025)	(0.025)	(0.025)	(0.025)
Mother College Grad		0.213***	0.213***	0.213***	0.214***		0.098***	0.098***	0.100***	0.099***
		(0.036)	(0.036)	(0.036)	(0.036)		(0.032)	(0.032)	(0.032)	(0.032)
Mother Post College		0.286***	0.284***	0.287***	0.287***		0.202***	0.202***	0.205***	0.202***
		(0.043)	(0.043)	(0.042)	(0.043)		(0.040)	(0.040)	(0.040)	(0.040)
Mother Not in HH		0.067*	0.066*	0.066*	0.067*		-0.043	-0.045	-0.043	-0.042
		(0.039)	(0.039)	(0.038)	(0.038)		(0.041)	(0.041)	(0.041)	(0.040)
Father HS Grad		0.053**	0.054**	0.053**	0.052**		0.010	0.011	0.010	0.011
		(0.023)	(0.023)	(0.023)	(0.023)		(0.027)	(0.027)	(0.028)	(0.027)
Father Some College		0.120***	0.119***	0.120***	0.121***		0.107***	0.106***	0.106***	0.105***
		(0.026)	(0.025)	(0.026)	(0.026)		(0.030)	(0.030)	(0.030)	(0.030)
Father College Grad		0.239***	0.240***	0.240***	0.240***		0.201***	0.202***	0.201***	0.200***
		(0.035)	(0.035)	(0.035)	(0.035)		(0.045)	(0.045)	(0.045)	(0.044)
Father Post College		0.301***	0.298***	0.300***	0.298***		0.319***	0.319***	0.318***	0.315***
		(0.052)	(0.052)	(0.052)	(0.052)		(0.044)	(0.044)	(0.044)	(0.044)
Father Not in HH		0.034	0.034	0.034	0.033		0.042	0.043	0.041	0.042
		(0.023)	(0.023)	(0.023)	(0.023)		(0.030)	(0.030)	(0.030)	(0.030)
Log Family Income		0.046***	0.046***	0.046***	0.046***		0.024**	0.024**	0.023**	0.023**
		(0.012)	(0.012)	(0.012)	(0.012)		(0.011)	(0.011)	(0.011)	(0.011)
PVT Rank			0.135**					0.109		
			(0.058)					(0.067)		
Fraction Female				0.394*	0.438*				-0.540**	-0.575**
				(0.226)	(0.226)				(0.269)	(0.267)
Fraction Foreign Born					-0.941					1.294**
					(0.642)					(0.522)
Fraction Black					-0.703					-0.340
					(0.521)					(0.603)
Fraction Latino					-0.294					-0.120
					(0.430)					(0.410)
Fraction Asian					0.633					-1.534*
					(0.686)					(0.793)
Fraction Other Races					-0.433					-0.113
					(0.310)					(0.250)
Observations	5899	5650	5649	5650	5650	4954	4703	4703	4703	4703
R^2	0.187	0.332	0.332	0.332	0.334	0.199	0.326	0.326	0.326	0.328
Adjusted R^2	0.152	0.299	0.299	0.300	0.300	0.158	0.286	0.286	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. $MaleFrac\_high$ (respectively, $FemaleFrac\_high$) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, and a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Columns (2)-(5) and (7)-(10) include race 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: High Flyers and Bachelor's Degree Attainment

	Dependent Variable: Bachelor's degree									
	Females					Males				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IHS MaleNumb_high	-0.101***	-0.104***	-0.100***	-0.095***	-0.095***	-0.004	0.027	0.028	0.006	0.004
	(0.023)	(0.028)	(0.027)	(0.031)	(0.030)	(0.052)	(0.045)	(0.046)	(0.046)	(0.046)
IHS FemaleNumb_high	0.036	0.058**	0.058**	0.053*	0.050*	0.016	0.001	0.001	0.016	0.035
	(0.033)	(0.027)	(0.026)	(0.028)	(0.026)	(0.050)	(0.046)	(0.046)	(0.043)	(0.047)
Foreign Born		0.089**	0.081**	0.089**	0.087**		0.102**	0.100**	0.101**	0.100**
		(0.040)	(0.040)	(0.040)	(0.040)		(0.045)	(0.045)	(0.045)	(0.045)
PVT Score		0.531***	0.235*	0.532***	0.532***		0.455***	0.211	0.450***	0.450***
		(0.065)	(0.135)	(0.065)	(0.064)		(0.071)	(0.169)	(0.071)	(0.071)
Age in Months		-0.006***	-0.005***	-0.006***	-0.006***		-0.005***	-0.005***	-0.005***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Mother HS Grad		0.060***	0.059***	0.060***	0.058**		0.008	0.008	0.010	0.007
		(0.022)	(0.022)	(0.022)	(0.022)		(0.021)	(0.021)	(0.021)	(0.021)
Mother Some College		0.115***	0.114***	0.115***	0.115***		0.086***	0.086***	0.089***	0.085***
		(0.026)	(0.026)	(0.026)	(0.026)		(0.025)	(0.025)	(0.024)	(0.025)
Mother College Grad		0.213***	0.212***	0.213***	0.212***		0.099***	0.098***	0.101***	0.098***
		(0.036)	(0.036)	(0.036)	(0.036)		(0.032)	(0.032)	(0.032)	(0.032)
Mother Post College		0.287***	0.285***	0.287***	0.286***		0.203***	0.203***	0.205***	0.204***
		(0.043)	(0.043)	(0.043)	(0.043)		(0.040)	(0.040)	(0.040)	(0.040)
Mother Not in HH		0.067*	0.066*	0.067*	0.065*		-0.043	-0.045	-0.043	-0.044
		(0.039)	(0.038)	(0.038)	(0.038)		(0.041)	(0.041)	(0.041)	(0.040)
Father HS Grad		0.052**	0.052**	0.052**	0.051**		0.010	0.011	0.010	0.012
		(0.023)	(0.023)	(0.023)	(0.023)		(0.027)	(0.027)	(0.027)	(0.027)
Father Some College		0.119***	0.118***	0.119***	0.117***		0.106***	0.106***	0.106***	0.106***
		(0.026)	(0.026)	(0.026)	(0.026)		(0.030)	(0.030)	(0.030)	(0.030)
Father College Grad		0.240***	0.240***	0.240***	0.238***		0.202***	0.203***	0.201***	0.202***
		(0.035)	(0.035)	(0.035)	(0.035)		(0.045)	(0.045)	(0.045)	(0.044)
Father Post College		0.300***	0.296***	0.300***	0.297***		0.319***	0.319***	0.318***	0.316***
		(0.051)	(0.051)	(0.052)	(0.052)		(0.044)	(0.044)	(0.044)	(0.044)
Father Not in HH		0.034	0.034	0.034	0.031		0.041	0.042	0.041	0.042
		(0.023)	(0.023)	(0.023)	(0.023)		(0.031)	(0.031)	(0.031)	(0.030)
Log Family Income		0.045***	0.046***	0.045***	0.045***		0.024**	0.024**	0.023**	0.023**
		(0.012)	(0.012)	(0.012)	(0.012)		(0.011)	(0.011)	(0.011)	(0.011)
PVT Rank			0.134**					0.109		
			(0.059)					(0.067)		
Fraction Female				0.222	0.383				-0.573**	-0.720**
				(0.247)	(0.256)				(0.280)	(0.290)
IHS Number Foreign Born					-0.050**					0.048**
					(0.022)					(0.021)
IHS Number Black					0.004					-0.032
					(0.027)					(0.043)
IHS Number Latino					0.022					-0.009
					(0.022)					(0.023)
IHS Number Asian					0.021					-0.060*
					(0.032)					(0.031)
IHS Number Other Race					-0.010					-0.031
					(0.047)					(0.044)
Observations	5899	5650	5649	5650	5650	4954	4703	4703	4703	4703
R^2	0.186	0.332	0.332	0.332	0.333	0.199	0.325	0.326	0.326	0.328
Adjusted R^2	0.152	0.300	0.299	0.300	0.300	0.158	0.285	0.286	0.286	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. *MaleNumb_high* (respectively, *FemaleNumb_high*) is the number of male (respectively, female) "High Flyers" (those with at least one post-college parent). IHS refers to the inverse hyperbolic sine transformation. All columns include school fixed effects, grade fixed effects, school linear time trends, and a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Columns (2)-(5) and (7)-(10) include race 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 4: Alternative Measures of Peer Achievement and Bachelor's Degree Attainment

	Dependent Variable: Bachelor's degree			
	Females		Males	
	(1)	(2)	(3)	(4)
MaleAvg	-0.093*	0.004	-0.004	-0.033
	(0.048)	(0.058)	(0.051)	(0.062)
FemaleAvg	0.072**	0.095**	0.008	-0.016
	(0.035)	(0.041)	(0.049)	(0.050)
MaleFrac_high		-1.162***		0.291
		(0.362)		(0.441)
FemaleFrac_high		-0.091		0.294
		(0.391)		(0.554)
Observations	5650	5650	4703	4703
R^2	0.333	0.334	0.328	0.328
Adjusted R^2	0.300	0.301	0.287	0.287
Equality P Value MaleAvg vs MaleFrac_high		0.004		0.499
Equality P Value FemaleAvg vs FemaleFrac_high		0.655		0.591

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. *MaleAvg* (respectively, *FemaleAvg*) is the average parents' years of education for male (respectively, female) peers. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including female, 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. Equality P values are the results of Wald tests for equality of coefficients. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table 5: Comparison with Prior Literature

	Dependent Variable: Bachelor's degree						
	All (1)	Females (2)	Females (3)	Females (4)	Males (5)	Males (6)	Males (7)
MotherFrac_Coll	-0.150	-0.304			0.094		
	(0.274)	(0.296)			(0.398)		
MaleMotherFrac_Coll			-0.499**	-0.107		0.067	0.018
			(0.243)	(0.242)		(0.300)	(0.291)
FemaleMotherFrac_Coll			0.256	0.217		0.041	-0.027
			(0.302)	(0.322)		(0.344)	(0.334)
MaleFrac_high				-1.022***			0.117
				(0.331)			(0.342)
FemaleFrac_high				0.169			0.217
				(0.373)			(0.489)
Female	0.067***						
	(0.013)						
Observations	10353	5650	5650	5650	4703	4703	4703
R^2	0.294	0.332	0.332	0.334	0.328	0.328	0.328
Adjusted R^2	0.275	0.299	0.299	0.300	0.287	0.287	0.287
Equality P Value MaleMotherFrac_Coll vs MaleMotherFrac_High				0.055			0.839
Equality P Value FemaleMotherFrac_Coll vs FemaleMotherFrac_High				0.936			0.700

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. *MaleMotherFrac_Coll* (respectively, *FemaleMotherFrac_Coll*) is the fraction of male (respectively, female) peers with a college-educated mother. *MaleMotherFrac_high* (respectively, *FemaleMotherFrac_high*) is the fraction of male (respectively, female) peers with a post-college mother. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including female, 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. Equality P values are the results of Wald tests for equality of coefficients. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table 6: High Flyers and Bachelor's Degree Attainment: Heterogeneity

	Dependent Variable: Bachelor's Degree							
	Females Below Med PVT (1)	Above Med PVT (2)	Neither Parent Coll (3)	Parent Coll (4)	Males Below Med PVT (5)	Above Med PVT (6)	Neither Parent Coll (7)	Parent Coll (8)
MaleFrac_high	-1.423***	-0.744	-0.046	-2.151***	-0.167	0.187	-0.212	0.567
	(0.475)	(0.487)	(0.431)	(0.549)	(0.665)	(0.359)	(0.503)	(0.523)
FemaleFrac_high	1.487***	-0.351	0.696*	0.325	0.776	0.098	0.345	-0.006
	(0.469)	(0.491)	(0.390)	(0.681)	(0.710)	(0.609)	(0.615)	(0.754)
Observations	2913	2737	3443	1939	2093	2610	2744	1731
R^2	0.330	0.357	0.254	0.405	0.351	0.368	0.279	0.443
Adjusted R^2	0.263	0.290	0.192	0.313	0.258	0.296	0.203	0.345
Dep Var Mean	0.213	0.475	0.235	0.607	0.152	0.376	0.185	0.479

	Dependent Variable: Confidence Index							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MaleFrac_high	-3.307**	-0.320	-0.966	-1.626*	-0.329	0.220	0.366	-0.775
	(1.373)	(0.893)	(0.871)	(0.875)	(1.602)	(0.995)	(1.334)	(1.360)
FemaleFrac_high	1.545	-0.623	-0.292	0.383	3.237**	1.362	1.299	2.844**
	(1.456)	(0.999)	(1.240)	(1.044)	(1.398)	(1.100)	(1.385)	(1.408)
Observations	2898	2733	3433	1935	2081	2604	2734	1727
R^2	0.275	0.254	0.198	0.321	0.301	0.274	0.239	0.332
Adjusted R^2	0.202	0.176	0.132	0.216	0.200	0.191	0.159	0.214

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment and the confidence index on individual and peer characteristics. In the upper panel, the dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. In the lower panel, the dependent variable is the Confidence Index, which 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. Columns (1)-(2) and (5)-(6) split the sample into below-median and above-median PVT score. Columns (3)-(4) and (7)-(8) split the sample by whether at least one parent has a college degree. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including fraction female, 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. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table 7: Alternative Measures of Peer Achievement for Subgroups of Students

(a) Below-Median PVT Scores

	Dependent Variable: Bachelor's degree			
	Females		Males	
	(1)	(2)	(3)	(4)
MaleAvg	-0.074	0.010	-0.127	-0.208*
	(0.060)	(0.077)	(0.093)	(0.114)
FemaleAvg	0.142***	0.080	0.003	-0.099
	(0.043)	(0.049)	(0.072)	(0.078)
MaleFrac_high		-1.443**		0.897
		(0.620)		(0.859)
FemaleFrac_high		1.126*		1.283
		(0.578)		(0.857)
Observations	2913	2913	2093	2093
R^2	0.327	0.331	0.353	0.356
Adjusted R^2	0.260	0.263	0.260	0.262
Equality P Value MaleAvg vs MaleFrac_high		0.033		0.240
Equality P Value FemaleAvg vs FemaleFrac_high		0.089		0.127

(b) Either Parent with a College Degree

	Dependent Variable: Bachelor's degree			
	Females		Males	
	(1)	(2)	(3)	(4)
MaleAvg	-0.245**	-0.020	0.160	0.174
	(0.110)	(0.140)	(0.106)	(0.145)
FemaleAvg	0.159	0.222	0.001	0.010
	(0.114)	(0.135)	(0.109)	(0.112)
MaleFrac_high		-2.081***		-0.118
		(0.667)		(0.738)
FemaleFrac_high		-0.410		-0.076
		(0.813)		(0.827)
Observations	1939	1939	1731	1731
R^2	0.402	0.407	0.444	0.444
Adjusted R^2	0.310	0.315	0.346	0.345
Equality P Value MaleAvg vs MaleFrac_high		0.008		0.729
Equality P Value FemaleAvg vs FemaleFrac_high		0.485		0.922

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of bachelor's degree attainment on individual and peer characteristics. Panel (a) restricts the sample to those with below-median PVT scores and panel (b) restricts the sample to those with at least one college-educated parent. The dependent variable is equal to 1 if the individual has completed a bachelor's (four-year college) degree and 0 otherwise. *MaleAvg* (respectively, *FemaleAvg*) is the average parents' years of education for male (respectively, female) peers. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including female, 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. Equality P values are the results of Wald tests for equality of coefficients. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table 8: Other Outcomes

	Females								Males							
	(1) GPA	(2) Math	(3) Science	(4) Confidence Index	(5) Two Year Degree	(6) LFP	(7) Ever Married	(8) Total No. Children	(9) GPA	(10) Math	(11) Science	(12) Confidence Index	(13) Two Year Degree	(14) LFP	(15) Ever Married	(16) Total No. Children
MaleFrac_high	-1.035	-2.287***	-1.998**	-1.331*	1.119***	-0.625**	-0.243	1.667**	0.553	0.692	-0.341	0.096	0.047	-0.085	-0.231	-0.242
	(0.644)	(0.852)	(0.859)	(0.674)	(0.301)	(0.304)	(0.458)	(0.774)	(0.531)	(0.822)	(0.920)	(0.894)	(0.281)	(0.213)	(0.432)	(0.683)
FemaleFrac_high	0.216	-0.158	-0.546	0.346	-0.239	0.465*	-0.029	-1.440	0.692	-0.844	0.760	1.376*	0.076	-0.078	-0.474	-1.897**
	(0.523)	(0.657)	(0.860)	(0.714)	(0.320)	(0.241)	(0.467)	(0.972)	(0.729)	(0.979)	(0.903)	(0.781)	(0.239)	(0.235)	(0.528)	(0.877)
Observations	5548	5224	4961	5631	5650	4090	5649	5640	4611	4388	4135	4685	4703	3275	4698	4678
R^2	0.295	0.181	0.248	0.237	0.097	0.100	0.202	0.222	0.292	0.239	0.253	0.245	0.115	0.161	0.205	0.208
Adjusted R^2	0.259	0.137	0.206	0.199	0.052	0.061	0.162	0.183	0.248	0.189	0.201	0.199	0.062	0.114	0.156	0.159
Dep Var Mean	2.925	2.762	2.945	0.001	0.196	0.849	0.547	1.075	2.735	2.679	2.768	0.005	0.149	0.934	0.454	0.718

Note: This table reports parameter estimates and standard errors (in parentheses) for regressions of GPA, math grade, science grade, the confidence index, two-year degree, labor force participation, marriage, and fertility on individual and peer characteristics. Grades are based on student reports of their grade in each subject, with A=4, B=3, C=2, and D or lower=1. GPA reflects the average across four subjects (math, science, english, and history). 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 variable Two Year 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 as the highest degree and 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. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including fraction female, 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. 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

For Online Publication

Appendix

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.

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.[26]

[26] For the 29 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.

Imputing Average Parents' Years of Education

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 convert each parents' level of attainment into years of education as follows: less than high school equals 9 years, high school degree equals 12 years, some college equals 14 years, college degree equals 16 years, and post-college education equals 18 years. 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's years of education by taking the average value of father's years of education 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's years of education using the average value among all students in the school who have a residential mother with a high-school education.[27]

Placebo Test

We use the simulation exercise described in Athey and Imbens (2017) to study the likelihood that a result of this magnitude could have occurred by chance by generating randomness in the exposure of individuals to "high-flyer" students. We first separate the value of $MaleFrac\_high$ for each grade in each school into two components: the predicted value based on the school fixed effect and linear time trend, and the residual. We then assign to each individual in the sample the residual from a random grade within the same school, and add it to the predicted value for that individual's true grade to generate random deviations from the trend in $MaleFrac\_high$. All other variables (both own and those of one's peers) are kept at their true levels. This procedure is repeated 1,000 times, running the fullest specification of Table 2 using the placebo $MaleFrac\_high$ and the actual $FemaleFrac\_high$. The equivalent procedure is repeated to obtain placebo values of $FemaleFrac\_high$.

The distributions of the estimated coefficients on $MaleFrac\_high$ and $FemaleFrac\_high$, by gender, are shown in Figure A2, with the vertical line in each graph indicating the estimated treatment effect we obtained in Table 2 and the share of estimates that is larger in absolute value than the dashed line (actual treatment) representing the randomization-based p-value.

[27] For the 29 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.

Appendix Tables and Figures

Table A1: 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 A2: Parental Education and Child GPA Over 3.5: In-Home Sample

	Dependent Variable: Grade Point Average Over 3.5							
	Females				Males			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mother HS Grad	0.048**	0.036*	0.027	0.027	0.025	0.021	0.010	0.009
	(0.019)	(0.020)	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.024)
Mother Some College	0.079***	0.065**	0.042*	0.037	0.055**	0.049*	0.025	0.023
	(0.025)	(0.025)	(0.024)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)
Mother College Grad	0.140***	0.116***	0.084***	0.089***	0.078**	0.064**	0.039	0.036
	(0.029)	(0.031)	(0.030)	(0.031)	(0.031)	(0.031)	(0.031)	(0.033)
Mother Post College	0.257***	0.229***	0.169***	0.168***	0.174***	0.161***	0.126***	0.127***
	(0.044)	(0.045)	(0.044)	(0.045)	(0.042)	(0.043)	(0.043)	(0.043)
Mother Not in HH	0.008	0.006	0.014	0.023	-0.058**	-0.055**	-0.058**	-0.056*
	(0.030)	(0.030)	(0.032)	(0.034)	(0.029)	(0.028)	(0.029)	(0.030)
Father HS Grad	0.036	0.027	0.016	0.021	0.023	0.010	0.001	0.000
	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Father Some College	0.111***	0.095***	0.073**	0.086***	0.098***	0.083***	0.069**	0.066**
	(0.027)	(0.026)	(0.028)	(0.030)	(0.030)	(0.030)	(0.029)	(0.031)
Father College Grad	0.193***	0.173***	0.148***	0.148***	0.091**	0.076**	0.059*	0.048
	(0.033)	(0.033)	(0.031)	(0.032)	(0.036)	(0.034)	(0.033)	(0.033)
Father Post College	0.198***	0.175***	0.150***	0.156***	0.171***	0.147***	0.137***	0.136***
	(0.043)	(0.042)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Father Not in HH	0.010	0.019	0.004	0.010	-0.003	0.004	-0.007	-0.013
	(0.021)	(0.020)	(0.021)	(0.022)	(0.023)	(0.024)	(0.025)	(0.025)
Age in Months		-0.004***	-0.002**	-0.002**		-0.004***	-0.003***	-0.002**
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Foreign Born		0.079**	0.125***	0.116***		0.048	0.075*	0.080*
		(0.037)	(0.036)	(0.037)		(0.036)	(0.041)	(0.041)
Log Family Income		0.017**	0.009	0.009		0.015*	0.012	0.011
		(0.008)	(0.009)	(0.009)		(0.008)	(0.008)	(0.008)
PVT Score			0.692***	0.669***			0.410***	0.415***
			(0.067)	(0.071)			(0.067)	(0.070)
Observations	5788	5788	5548	5548	4851	4848	4611	4611
R^2	0.139	0.150	0.183	0.213	0.129	0.151	0.174	0.211
Adjusted R^2	0.118	0.129	0.161	0.174	0.104	0.125	0.147	0.164
Equality P Value Moth Coll vs Moth Postcoll	0.007	0.008	0.041	0.059	0.016	0.015	0.029	0.025
Equality P Value Fath Coll vs Fath Postcoll	0.907	0.959	0.954	0.849	0.043	0.077	0.053	0.036

Note: This table reports parameter estimates and standard errors (in parentheses) from a regression on the in-home sample in which the dependent variable is equal to 1 if the student has a GPA strictly over 3.5. 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 grade and school fixed effects and a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year. Columns (2)-(4) and (6)-(8) include dummies for Black, Latino, Asian, and other races. Columns (4) and (8) include school linear time trends. 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 A3: Variation in *MaleFrac_high* and *FemaleFrac_high*

	Females		Males	
	MaleFrac_high	FemaleFrac_high	MaleFrac_high	FemaleFrac_high
Raw Variation				
Mean	0.144	0.122	0.143	0.119
SD	0.102	0.102	0.096	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

Note: This table reports the raw and residual (net of fixed effects and time trends) variation in *MaleFrac_high* and *FemaleFrac_high*. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent) in the grade and school. Wave IV weights used.

Table A4: Parental Education and Child GPA: In-School Sample

	Dependent Variable: Grade Point Average					
	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
Mother HS Grad	0.124***	0.099***	0.097***	0.100***	0.077***	0.074***
	(0.016)	(0.014)	(0.014)	(0.019)	(0.018)	(0.018)
Mother Some College	0.231***	0.198***	0.197***	0.205***	0.177***	0.176***
	(0.019)	(0.017)	(0.017)	(0.024)	(0.022)	(0.022)
Mother College Grad	0.307***	0.258***	0.257***	0.272***	0.232***	0.229***
	(0.018)	(0.018)	(0.017)	(0.025)	(0.024)	(0.024)
Mother Post College	0.345***	0.295***	0.292***	0.330***	0.292***	0.287***
	(0.024)	(0.023)	(0.023)	(0.025)	(0.023)	(0.024)
Mother Not in HH	-0.010	-0.013	-0.014	-0.001	-0.001	-0.004
	(0.020)	(0.017)	(0.017)	(0.021)	(0.019)	(0.020)
Father HS Grad	0.067***	0.047***	0.046***	0.073***	0.059***	0.061***
	(0.019)	(0.018)	(0.018)	(0.020)	(0.019)	(0.019)
Father Some College	0.229***	0.200***	0.198***	0.203***	0.179***	0.181***
	(0.023)	(0.021)	(0.021)	(0.023)	(0.022)	(0.022)
Father College Grad	0.273***	0.232***	0.230***	0.294***	0.261***	0.265***
	(0.021)	(0.021)	(0.021)	(0.024)	(0.024)	(0.023)
Father Post College	0.386***	0.339***	0.337***	0.378***	0.342***	0.346***
	(0.029)	(0.027)	(0.027)	(0.029)	(0.028)	(0.028)
Father Not in HH	-0.077***	-0.052***	-0.052***	-0.059***	-0.035*	-0.032*
	(0.017)	(0.016)	(0.016)	(0.018)	(0.017)	(0.017)
Foreign Born		0.157***	0.155***		0.138***	0.139***
		(0.022)	(0.021)		(0.023)	(0.023)
Age in Years		-0.152***	-0.150***		-0.141***	-0.141***
		(0.011)	(0.011)		(0.009)	(0.009)
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	38832	38832	38832	37694	37694	37694
R^2	0.170	0.202	0.213	0.160	0.187	0.195
Adjusted R^2	0.167	0.198	0.207	0.156	0.183	0.188
Equality P Value Moth Coll vs Moth Postcoll	0.022	0.022	0.030	0.001	0.000	0.000
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 on the in-school sample in which the dependent variable is equal to grade point average, which is calculated based on self-reported student grades in math, science, english, and history where A=4, B=3, C=2, and D or lower=1. All columns include grade and school fixed effects. Columns (2)-(3) and (5)-(6) include dummies for Black, Latino, Asian, and other races. Columns (3) and (6) include school linear time trends. 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 foreign born status or school age is missing, variables are set equal to zero and a dummy is included for missing foreign born status or age, respectively. Regressions are unweighted. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table A5: High Flyers and Bachelor's Degree Attainment: Robustness

	Dependent Variable: Bachelor's degree					
	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
MaleFrac_high	-0.853***	-1.035***	-0.975***	0.199	0.158	0.209
	(0.273)	(0.347)	(0.365)	(0.296)	(0.471)	(0.424)
FemaleFrac_high	0.047	0.199	0.382	-0.043	0.369	0.410
	(0.266)	(0.336)	(0.352)	(0.370)	(0.527)	(0.533)
Observations	5650	4977	4952	4703	4056	4097
R^2	0.308	0.333	0.282	0.289	0.328	0.274
Adjusted R^2	0.289	0.299	0.241	0.266	0.285	0.223

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. Column (1) (respectively (4) for males) excludes linear time trends. Column (2) (respectively (4) for males) excludes schools with variation in *MaleFrac_high* and *FemaleFrac_high* outside of the 90 percent confidence interval obtained in the Monte Carlo simulations. Column (3) (respectively (6) for males) excludes those with a post college mother or father. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including fraction female, foreign born, Black, Latino, Asian, and other races. Columns (2), (3) and (5), and (6) include school linear time trends. 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 A6: High Flyers and Bachelor's Degree Attainment: Alternative Measures

	Dependent Variable: Bachelor's degree					
	Females (1)	(2)	(3)	Males (4)	Males (5)	(6)
MaleMotherFrac_high	-1.032**			0.369		
	(0.411)			(0.455)		
FemaleMotherFrac_high	-0.438			0.342		
	(0.470)			(0.491)		
Male FatherFrac_high		-1.252***			-0.188	
		(0.374)			(0.563)	
FemaleFatherFrac_high		0.664*			0.027	
		(0.339)			(0.574)	
MaleBothFrac_high			-1.526**			-0.064
			(0.681)			(0.783)
FemaleBothFrac_high			0.508			-0.115
			(0.570)			(0.749)
Observations	5650	5650	5650	4703	4703	4703
R^2	0.333	0.334	0.333	0.328	0.328	0.328
Adjusted R^2	0.300	0.301	0.300	0.287	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. *MaleMotherFrac_high* (respectively, *FemaleMotherFrac_high*) represents male (respectively, female) "High Flyers," measured using only the post-college status of mothers, as a fraction of males (respectively, females) in the grade. *MaleFatherFrac_high* (respectively, *FemaleFatherFrac_high*) uses only the post-college status of fathers. *MaleBothFrac_high* (respectively, *FemaleBothFrac_high*) represents male (respectively, female) "High Flyers" (those with two post-college parents) as a fraction of males (respectively, females) in the grade. *MaleFrac_Coll* (respectively, *FemaleFrac_Coll*) represents male (respectively, female) "High Flyers" (those with at least one college parent) as a fraction of males (respectively, females) in the grade. All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including fraction female, 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. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

Table A7: Sample Attrition

	Dependent Variable: In Wave IV Sample			
	Females		Males	
	(1)	(2)	(3)	(4)
MaleFrac_high	0.196		0.306	
	(0.254)		(0.308)	
FemaleFrac_high	0.323		-0.487	
	(0.250)		(0.433)	
IHS MaleNumb_high		0.027		0.026
		(0.026)		(0.032)
IHS FemaleNumb_high		0.022		-0.051
		(0.022)		(0.032)
Observations	6878	6876	6250	6250
R^2	0.155	0.149	0.146	0.146
Adjusted R^2	0.121	0.114	0.107	0.107

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. $MaleFrac\_high$ (respectively, $FemaleFrac\_high$) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent) as described in the text. $MaleNumb\_high$ (respectively, $FemaleNumb\_high$) is the number of male (respectively, female) "High Flyers" (those with at least one post-college parent). IHS refers to the inverse hyperbolic sine transformation. All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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 female, foreign born, Black, Latino, Asian, and other races in columns (1) and (3) and the fraction female 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.

Table A8: 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

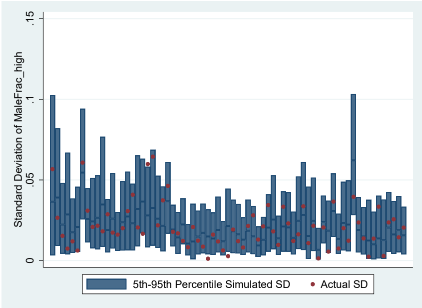
Note: This table reports eigenvalues and factor loadings based on factor analysis of the following variables: "Want College," which equals 1 if the student reports that they want to go to college as a 5 on a scale of 1-5 and equals 0 otherwise; "College likely," which equals 1 if the student says the 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 A9: Alternative Measures of Peer Achievement and Confidence

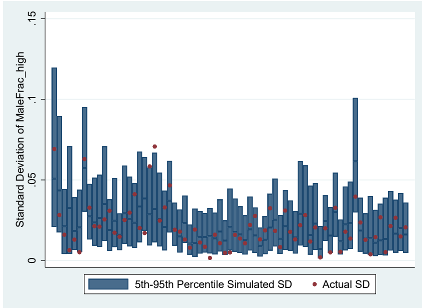
	Dependent Variable: Confidence Index			
	Females		Males	
	(1)	(2)	(3)	(4)
MaleAvg	-0.034	0.145	-0.091	-0.223
	(0.090)	(0.110)	(0.126)	(0.171)
FemaleAvg	0.117	0.177**	0.030	-0.131
	(0.077)	(0.084)	(0.120)	(0.141)
MaleFrac_high		-2.077***		1.208
		(0.784)		(1.145)
FemaleFrac_high		-0.445		2.082**
		(0.868)		(0.952)
Observations	5631	5631	4685	4685
R^2	0.237	0.238	0.245	0.246
Adjusted R^2	0.198	0.199	0.199	0.200
Equality P Value MaleAvg vs MaleFrac_high		0.010		0.260
Equality P Value FemaleAvg vs FemaleFrac_high		0.499		0.036

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 the Confidence Index, which 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. *MaleAvg* (respectively, *FemaleAvg*) is the average parents' years of education for male (respectively, female) peers. *MaleFrac_high* (respectively, *FemaleFrac_high*) is the fraction of male (respectively, female) "High Flyers" (those with at least one post-college parent). All columns include school fixed effects, grade fixed effects, school linear time trends, a dummy for whether Wave I interview took place in 1994-1995 or 1995-1996 school year, race dummies (Black, Latino, Asian, and other races), age in months, PVT score, 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. All columns also include controls for peer characteristics including female, 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. Equality P values are the results of Wald tests for equality of coefficients. Wave IV weights used. Standard errors clustered at the school level. * p<0.1 ** p<0.05 *** p<0.01

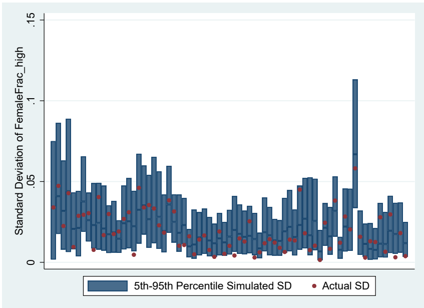
Figure A1: Monte Carlo Estimates of $MaleFrac\_high$ and $FemaleFrac\_high$



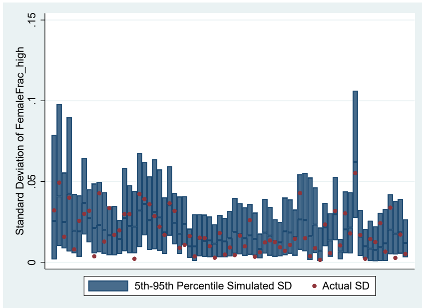
(a) Females: $MaleFrac\_high$



(b) Males: $MaleFrac\_high$



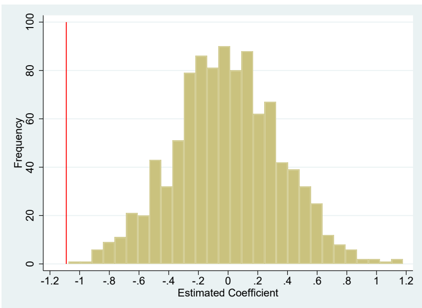
(c) Females: $FemaleFrac\_high$



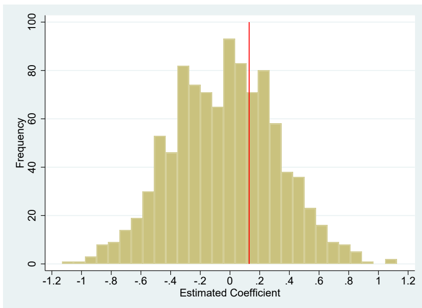
(d) Males: $FemaleFrac\_high$

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 $MaleFrac\_high$ and $FemaleFrac\_high$). The dot represents the empirical standard deviation.

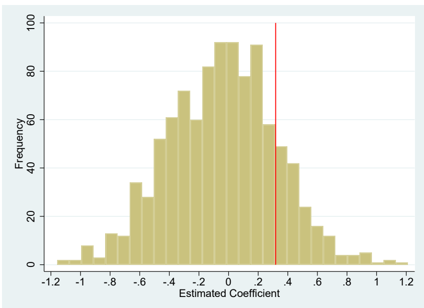
Figure A2: Randomization-Based Inference for $MaleFrac\_high$ and $FemaleFrac\_high$



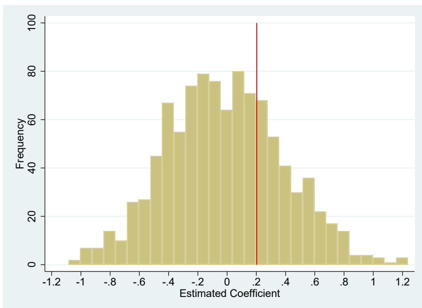
(a) Females: $MaleFrac\_high$



(b) Males: $MaleFrac\_high$



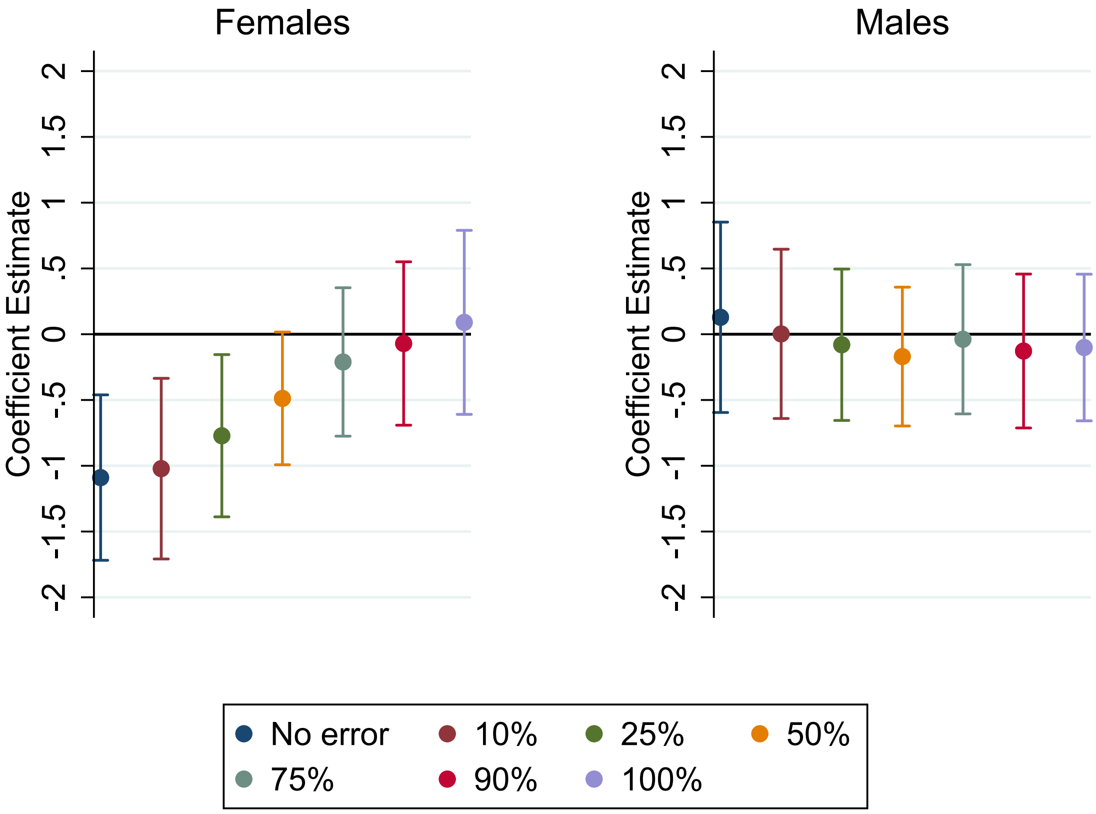
(c) Females: $FemaleFrac\_high$



(d) Males: $FemaleFrac\_high$

Note: These figures show distribution of coefficients obtained from the final OLS specification in Table 2 while replacing $MaleFrac\_high$ (respectively, $FemaleFrac\_high$) with the value of $MaleFrac\_high$ (respectively, $FemaleFrac\_high$) from a random grade in the same school. Red line represents actual estimate obtained in specifications (5) and (10) in Table 2.

Figure A3: Estimates of $MaleFrac\_high$ with Measurement Error



Note: These figures show coefficients and 95 percent confidence intervals obtained for $MaleFrac\_high$ with measurement error based on the OLS specifications in columns (5) and (10) of Table 2. For a given amount of measurement error on $MaleFrac\_high$, we replace parents' post college status for the given percent of boys in the in-school sample with a 0 or 1 based on a binomial distribution with a mean equal to the final sample mean. We then re-calculate $MaleFrac\_high$ for each individual and implement the regression in columns (5) and (10) of Table 2, keeping all other variables at their true levels.