

LIKE FATHER, LIKE DAUGHTER (UNLESS THERE IS A SON): SIBLING SEX COMPOSITION AND WOMEN'S STEM MAJOR CHOICE IN COLLEGE

UMUT OGUZOGLU[†] and SERKAN OZBEKLIK^{*}

September 2016

Abstract

We investigate the potential role of fathers in females' decision to choose a science, technology, engineering and mathematics (STEM) major in college. The main innovation of our paper is to analyze how sibling sex composition affects the probability of being a STEM major in college for females whose fathers are in a STEM occupation. Using the National Longitudinal Survey of Youth 1979 (NLSY79), we find that, for females, having brother(s) significantly decreases the likelihood of choosing a STEM major in college when their fathers are also in a STEM occupation. The inclusion of variables pertaining to respondents' attitudes toward traditional gender roles, birth order, and the presence of an older brother as well as frequently used cognitive skill measures does not change the results. Thus, the observed effect appears to be driven by change in the college major preferences of females. We replicate the analysis using a more recent data set from the U.S. and data from Australia, and find similar results. Our findings suggest that fathers are much more likely to transmit occupation-specific tastes and preferences to their daughters in the absence of a son potentially, contributing to the persistence of the gender gap in STEM majors in college.

JEL Codes: J24, J16, I23

[†] Associate Professor of Economics, University of Manitoba, Research Fellow, IZA. E-mail: umut.oguzoglu@umanitoba.ca. ^{*} Corresponding Author. Associate Professor of Economics, Claremont McKenna College. E-mail: serkan.ozbeklik@cmc.edu. The authors thank Heather Antecol, David Bjerck, Christian Dustmann, Henry Farber, Stefanie Fischer, Lisa Kahn, Steven Levitt, Stephen Machin, Alex Mas, Chris Neilson, Imran Rasul, Basit Zafar, and the workshop participants at the 2016 Society of Labor Economist Meeting, the 2016 Southern California Conference on Applied Microeconomics, and seminar participants at Deakin University, Princeton University, and Melbourne Institute of Applied Economic and Social Research.

1. Introduction

In the second part of the 20th century, women made substantial gains in educational attainment and labor market outcomes. In the United States, about 57 percent of bachelor's and 63 percent of master's degrees were conferred on females in 2010, up from 35 and 32 percent in 1960, respectively (National Center for Education Statistics 2012). College educated women now make up about half of the high-skilled labor force (U.S. Department of Commerce 2011). Despite these gains, college-educated women who work full-time earn only 73 percent of what men earn in the respective category (U.S. Bureau of Labor Statistics 2013). It has been argued that a significant portion of the gender gap in earnings among college graduates stems from the differences in the sectors in which men and women are employed (Brown and Corcoran 1997; Weinberger 1999, AAUW Educational Foundation 2007, Goldin 2014). In the U.S., this occupational sorting is the most evident, especially for college graduates, in science, technology, engineering, and mathematics (STEM) occupations. For example, in 2009, women constituted only 24 percent of the STEM workforce in the U.S. (U.S. Department of Commerce 2011). While women have made significant gains in other professional jobs between 1989 and 2009 and earn much larger STEM job premiums than men do (U.S. Department of Commerce 2011), the fraction of women employed in STEM jobs rose by only 3 percentage points.¹ Although the absence of women in STEM occupations relative to their fraction in high-skilled jobs has attracted much attention, the factors behind this gender gap remains elusive as identifying these factors is not a straightforward exercise. Occupational outcomes are shaped not only by an individual's ex-ante

¹ In the same time period, women's fraction among the employed increased from 38 to 51 percent in non-managerial business and finance jobs, from 43 to 55 percent in medicine and dentistry, and from 30 to 37 percent in managerial jobs.

occupational preferences but also by labor market conditions, employers' attitudes towards hiring women, and women's own satisfaction with the work environment and work-life balance.²

Not surprisingly, the gender gap in STEM employment is also mirrored in the gender differences in the number of college-educated workers with a STEM degree. In 2009, only about 27 percent of the 9.2 million workers with a STEM degree were females (U.S. Department of Commerce 2011). Moreover, this disparity is not a characteristic only of older cohorts. As late as 2006, 15 percent of female and 29 percent of male first-year college students picked a STEM major as their intended major (Hill et al. 2010). Furthermore, even though the percentage of STEM degrees awarded to women has gone up since the 1960s, only about 39 percent (26 percent if one excludes the biology major, for which females have been the majority of degree earners) of bachelor's degrees earned in STEM fields went to female college graduates in the same year (See Figure 1). As STEM education is the main gateway to STEM careers with high earnings potential, it is crucial to understand the factors affecting gender differences in choosing a STEM major in college. We argue that college major choice is a better metric than occupational choice in understanding gender differences in preferences for STEM careers. This is because major choice is less likely to be impacted by labor market conditions or employers' prejudice against hiring women in male dominated fields and, therefore, more likely than occupational choice to reflect differences in individuals' preferences.

Underrepresentation of women in STEM fields has been a focal point of the recent policy debate on the gender earnings gap in the U.S. Several policy initiatives have recently been

² In the United States, for example, a smaller fraction of women with a STEM degree choose a career in a STEM field and, when they do, they are more likely to leave the labor market than men because of reasons ranging from workplace environment to family responsibilities (Hewlett et al. 2008, and Frehill et al. 2009, Hill et al. 2010, Hunt 2016).

introduced in the U.S., aiming at improving women’s interest and performance in STEM fields. In a recent fact sheet released by the White House (White House, 2013) the importance of “[w]orking with teachers, businesses, philanthropists, foundations, non-profits, scientists, and engineers...” is emphasized to achieve this goal. In this paper, we present strong causal evidence that fathers’ involvement should also be an essential part of these initiatives. To the best of our knowledge, this is the first paper that causally links the gender gap in STEM major choice in college to parental involvement.

There is an extensive literature investigating gender differences in math and science achievement (e.g. Hyde et al. 1990, Hedges and Novell (1995), Lee et al. 2007, Hyde et al. 2008, Guiso et al. 2008, Fryer and Levitt 2010, Bharadwaj et al. 2012). A few papers have also analyzed whether pre-college academic achievement and skills have any role in explaining the corresponding differences in college major choice in the U.S. and concluded that the achievement and skill differences generally measured by test scores are not the driving factors behind the gender gap in college major choice. (Turner and Bowen 1999 and Arcidiacono 2004).³ In fact, Arcidiacono (2004) and Zafar (2013) found that gender differences in college majors can mainly be explained by differences in preferences and tastes between male and female students. Consistent with the findings of these two papers, Carrell et al. (2010) found that having larger fraction of introductory math and science courses taught by female professors has a large and significant positive impact on the likelihood of graduating with a STEM degree for females—especially for those in the upper quartile of math achievement distribution—arguing that “whatever it is about female professors that affects women in their first-year math and

³ The only exception is a recent working paper by Speer (2016) who argues that differences in pre-college ASVAB scores explain a significant portion of the gender gap in college STEM fields.

science courses, it is not something that changes retention rates but rather something that changes their *preferences* for math and science.” In identifying one of the factors that might explain part of the gender gap in STEM careers, our paper is further related to two other strands of the literature. First, numerous papers have argued that parents significantly influence their children’s educational and occupational choices, and that this influence starts early on (Eccles and Hoffman 1984, Eccles 1994 among many others). Moreover, a part of this influence is due to intergenerational transmission of educational preferences and labor market skills from parents to their children (e.g. Solon 1999, Bjorklund and Jantti 2009, Bjorklund and Salvanes 2011, Black and Devereux 2011). Second, it has been argued that sibling sex composition might affect educational attainment of females, though evidence in favor of this relationship is mixed at best. In particular, Butcher and Case (1994) show that women who have brothers receive significantly more education on average than women raised with any sister. Kaestner (1997), on the other hand, used more recent cohorts to investigate the significance of having a brother on women’s educational attainment and found no effect for white females and a negative effect for black females. Finally, Hauser and Kuo (1998) used larger surveys covering the whole 20th century and concluded that there is very little evidence that sibling sex composition had any impact on women’s educational attainment.

In this paper, we investigate whether gender-biased intergenerational transmission of occupation-specific tastes and preferences differs between females with different sibling sex compositions; thus, the paper indirectly contributes to explaining the differences between males’ and females’ decisions to choose a STEM major in college. The main innovation of our paper is to analyze how sibling sex composition affects the probability of being a STEM major in college for females whose fathers are also in a STEM occupation. This effect may work through several

channels. One possibility is that having a son may affect fathers' occupation-specific investment in daughters' human capital, or change fathers' influence on daughters' tastes and preferences. Alternatively, having a brother may alter women's perception of gender roles or competitiveness. As a result, the likelihood of choosing a STEM major for females with a male sibling(s) might be different from that of females without one.⁴

We find that females who have brother(s) are, depending on the specification and comparison group used, 15-28 percentage points less likely to choose a STEM major than females with only sister(s) if their fathers are also in a STEM occupation. For females whose fathers *do not* have a STEM job, we *do not* find a similar effect. The inclusion of variables pertaining to respondents' attitudes toward traditional gender roles, birth order, and the presence of an older brother as well as frequently used cognitive measures, which may be affected by the transmission of human capital and skills from fathers, does not change the results. Therefore, our results appear to be driven by female college major preferences rather than other potential mechanisms such as human capital and skills. We also replicate our analysis using a more recent data set from the United States and data from Australia, and find similar results. These findings tentatively suggest that fathers are more likely to pass on their occupation-specific preferences to their daughters in the absence of a son. Therefore, part of the gender gap in choosing a STEM major in college may potentially be attributed to intergenerational transmission of occupation-specific tastes and preferences from fathers to sons and daughters.

⁴ We explain these mechanisms in more detail below.

2. Sibling Sex Composition, Father's Occupation and Major Choice in College

In this section, we consider several mechanisms to explain how sibling sex composition may affect the relationship between father's occupation and daughter's college major choice. Our discussion will abstract away the effect of the number of siblings and mother's occupation on this relationship. Furthermore, since the effect is assumed to operate through the relationship between father's occupation and daughter's choice of a college major, we should not expect to see a similar pattern if the father does not have a STEM job.

For simplicity, consider two families, Family A and Family B, and suppose that both fathers in these families work in STEM jobs. Suppose also that each family has two children: Family A has two girls, and Family B has a girl and a boy. If the sibling's gender has an impact on the relationship between father's occupation and daughter's college major choice, the likelihood of choosing a STEM major for girls in Family A and Family B would differ, everything else being constant. The first mechanism through which we can see this difference is that the father in Family B might be more likely to pass on occupation-specific human capital and tastes and preferences to his daughters in the absence of a son. If fathers favor sons over daughters and are more involved in raising them on average (Lundberg 2005; Dahl and Moretti 2008), they can transmit these skills and tastes more to their sons. In the absence of a brother, females might get more attention from their fathers and more exposure to his occupation-specific tastes and preferences.

The second mechanism we consider is that the differences in the perceptions of females raised in Family A and Family B regarding traditional gender roles and attitudes might cause a

difference in the likelihood of choosing a “gender-appropriate” major in college. For the females in Families A and B, the effect of this mechanism on STEM major choice in college is ambiguous. On the one hand, the female in Family A might be more likely to internalize traditional gender roles and choose a more “gender-appropriate” major, as a result of differences in parental investment and expectations between her and her brother. Alternatively, females with older brothers have been shown to exhibit more “masculine” traits,⁵ and thus might be more likely to choose a “masculine” major (such as a STEM major) in college.

Finally, it has been documented that females are more prone to avoiding competition than males (Niederle and Vesterlund 2007). If siblings compete for parental investments, particularly for occupation-specific investment from their father, having a male sibling might discourage females from developing STEM-specific human capital and tastes and preferences. This mechanism predicts that the female in Family A is more likely to choose a STEM major in college than the female in Family B.

3. Data and Sample Selection

Our main data source is the National Longitudinal Survey of Youth 1979 – 2012 (NLSY79). NLSY79 is a nationally representative sample of American youth who were between 14 and 22 years old when first surveyed in 1979.⁶ These respondents were interviewed annually through 1994, and, since 1994, the survey has been conducted biannually. NLSY79 contains unique

⁵ See Stoneman, Brody, and MacKinnon (1986) for a review of this literature.

⁶ It is surprising that besides NLS79 and its offshoot, the National Longitudinal Study of Youth, Children and Young Adults (NLSCYA), we failed to find any large and nationally representative U.S. dataset that includes detailed information on father’s occupation, respondent’s college major choice, and the sex of each of the respondent’s siblings. A notable exception is PSID, which contains all information we seek but not the necessary sample size.

information on respondents' family background, educational history, and labor market experience. More importantly for this paper, father's occupation, gender and age of siblings, and field of study at the most recent college attended are provided.

3.1 STEM Major in College, Father's Occupation and Sibling Sex Composition

Since 1979, the NLSY79 has collected information on the major field of study at the most recent college attended. Starting from 1984, this information was extended to up to 3 most recent colleges. In our sample, an individual is identified as choosing a STEM major if she reported a STEM major in any of the years that she was interviewed. According to our definition, STEM fields are Agricultural Sciences (e.g. Agronomy, Soil Science, Animal Science and Food Sciences) Biological Sciences, Computer and Information Sciences, Engineering, Mathematics, Physical Sciences, Interdisciplinary Biological and Physical Sciences, and Interdisciplinary Engineering and Other Disciplines.⁷

We use occupation of the father in the beginning of the survey to identify fathers who work in STEM occupations in 1978. STEM occupations are selected according to the definition provided by the U.S. Census Bureau⁸. The list contains 3 categories: STEM occupations, STEM-related occupations, and non-STEM occupations. We choose only the STEM category to identify STEM occupations.

⁷ A full list of fields of study available in NLSY79 can be accessed at <https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-4-fields-study>.

⁸ A complete list from the Census Bureau can be accessed at <http://www.census.gov/people/io/files/STEM-Census-2010-occ-code-list.xls>.

In 1994, the NLSY79 included detailed information about up to 13 siblings, including gender and birth order. From this information, we define following three categories of sibling sex composition: having only brothers, having only sisters, and having both brothers and sisters.

3.2 Opinion on Gender Roles

We make extensive use of the information on *family attitudes*. This is a set of 8 questions that ask respondents whether they strongly disagree, disagree, agree, or strongly agree with phrases that describe traditional gender roles such as “*a woman’s place is in the home, not the office or shop*”, or “*women are much happier if they stay at home and take care of their children.*” We collapse each of the questions to a dummy variable that is equal to one if an individual agrees or strongly agrees with the traditional role described in the phrase⁹. We also create an *attitudes index* that ranges from zero (strongly disagrees with the traditional roles) to 8 (strongly agrees with the traditional roles) by simply summing over 8 dummy variables.

3.3 Test Scores

We use standardized age-adjusted Armed Forces Qualification Test (AFQT) scores in order to control for difference in cognitive skills. In separate analyses we replace AFQT scores with individuals’ scores (also standardized age-adjusted) from one or all of the following 8 subsections of Armed Services Vocational Aptitude Battery (ASVAB): General Science, Arithmetic Knowledge, Numerical Operations, Coding Speed, Auto and Shop Knowledge, Mathematics Knowledge, Mechanical Comprehension, Electronics Information. All test scores are normalized by subtracting mean score of the NLSY79 sample from the individual score, and dividing it by

⁹ When a given statement, such as “a working wife feels more useful than one who doesn't hold a job,” contradicts the traditional gender roles, the corresponding dummy variable is equal to one if a respondent strongly disagrees or disagrees with the statement.

standard deviation of the entire random sample. The normalized scores are then standardized by running a linear regression of normalized score on age dummies. The residuals from these models are retained as the standardized age-adjusted scores.

3.4 Sample Selection

Our sample includes all individuals who were 17 years old or older during any NLSY cycle. All observations with missing information on college major, father occupation, or relevant sibling characteristics are removed from the sample. We also exclude all single child respondents. The final sample includes 2,005 women. Table 1 shows descriptive statistics for the full sample of females as well as for the samples of females by sibling sex composition.

3.5 STEM Families and Enrollment in NLSY79

Before moving on to our econometric analysis, let us first briefly discuss patterns concerning STEM major enrolments in our data. By doing so, we set the stage for a more thorough analysis pertaining to the intergenerational aspect of choosing a STEM major. Table 2 provides information on major choice by gender and father's occupation in NLSY79. In our sample, 48 percent of males and 26 percent of females who were enrolled in college ever chose a STEM major. Both males and females with fathers employed in a STEM occupation are more likely to choose a STEM major in college than the respondents with fathers in a non-STEM occupation in our sample. However, males are almost twice as likely to choose a STEM major as females irrespective of their fathers' occupation, which is not far from the national statistics on the likelihood of STEM major choice by sex in the 1980s.

Table 3 focuses only on the females in our sample and shows the fraction of females who chose a STEM major in college by sibling sex composition type and whether the father had a

STEM occupation. In the full sample, this fraction is relatively close among females with different sibling sex composition. While 31 percent of the females who have only sisters chose a STEM major, the corresponding fraction for females who have only brothers is 23 percent. But, as displayed in columns (2) and (3), when we break the sample by whether or not the father is employed in a STEM job, an interesting pattern emerges. While the gap between the corresponding fractions of females whose fathers were not employed in a STEM occupation shrinks even more, for females with fathers working in STEM jobs, the presence of a brother appears to become very important in choosing a STEM major: only about 24 percent of the females who have any brother and 17 percent of the females who have only brother(s) chose a STEM major, whereas slightly less than 48 percent of the females who have only sisters chose a STEM major.

4. Estimation and Results

4.1 Methodology

To quantify the impact of sibling sex composition on choosing a STEM major for females in families where the father has a STEM occupation, and to control for potentially unobservable confounding variables (e.g. the endogeneity of number and sex composition of children among families), we first estimate the following difference-in-differences (DD) regression:

$$STEM_i = \beta_0 + \beta_1 FATHER_i^S + \beta_2 BRO_i + \beta_3 FATHER_i^S \times BRO_i + \beta_4 CSM_i + X_i' \delta + \varepsilon_i \quad (1)$$

where $STEM_i$ is equal to one if female i is a STEM major; BRO_i is one if female i has any brother; $FATHER_i^S$ takes the value of one if female i 's father has ever been employed in a STEM

occupation; CSM_i is the cognitive skill measure proxied by either AFQT or a subject specific ASVAB score; X_i is a vector of individual (age, race, and ethnicity dummies, a third degree polynomial of number of siblings, age difference with the closest brother, age difference with the closest sister, indicators for whether she is the first child, whether she has an older brother, whether she lived in the U.S. at age 14, whether she lived in an urban area at age 14, whether she lived with her parents at age 14, and the attitudes index) and parental controls (STEM occupation dummy for the mother, a missing occupation indicator for the mother, logarithm of family income in 1978, whether mother worked at age 14, whether father worked at age 14, whether mother and father are still alive, father's and mother's immigration status, indicators for mother's and father's highest degree of education). Our main assumption in the analysis is that the potential unobservable differences affecting women's college major choice between families with children who have various sex compositions are, on average, the same regardless of whether the father is in a STEM or non-STEM occupation. We are mainly interested in the size and statistical significance of the estimate of the coefficient β_3 , which captures the impact of having any brother on the likelihood of choosing a STEM major for the females with fathers employed in a STEM occupation relative to that of females with fathers not employed in a STEM occupation. We also expect the estimate of β_2 to be closer to zero and not statistically significant; if the impact of sibling sex composition on the choice of a STEM major for females operates mainly through its effect on the level and intensity of father's own occupation-specific investment on his daughter, one should not observe a similar effect for the group of females whose fathers are *not* in STEM occupations (which is captured by the estimate of β_2).

We also allow for having brother(s) only or both sister(s) and brother(s) to have distinct effects by estimating the following regression specification:

$$STEM_i = \alpha_0 + \beta_1 FATHER_i^S + \alpha_2 BROonly_i + \alpha_3 FATHER_i^S \times BROonly_i + \alpha_4 BROandSIS_i + \alpha_5 FATHER_i^S \times BROandSIS_i + \alpha_6 CSM_i + X_i' \delta + \varepsilon_i \quad (2)$$

where $BROonly_i$ is one if female i only has brother(s) and zero otherwise; and $BROandSIS_i$ is the corresponding dummy variable for females with both sister(s) and brother(s). Everything else is defined as in equation (1). Our parameters of interest in equation (2) are the coefficient estimates of α_3 and α_5 ; the estimate of α_3 captures the impact of having only brother(s) on the likelihood of females' STEM major choice, while that of α_5 summarizes the corresponding impact for females with both brother(s) and sister(s).

4.2 Threats to Identification

In the absence of a credible control group, our identification strategy would depend on the assumption that there are no unobservables that are correlated with both the sibling sex composition of females and their major choice. Although we cannot directly test it, our use of females without fathers in a STEM occupation as our control group weakens our identification assumption; even if there is an unobservable that affects both the sibling sex composition and the females' major choice, it will not bias our estimates if the effect is the same on average for females with and without STEM fathers, our main assumption as noted in section 4.1. In other words, if females with non-STEM fathers are a credible control group, our DD estimation will eliminate this source of potential bias.

To check the credibility of our control group, we conduct several exercises. First, as noted in section 2, since the effect is assumed to operate through the relationship between father's occupation and daughter's choice of a college major, having brother(s) should not affect the

likelihood of choosing a STEM major for females whose fathers do not have a STEM job. In Table 4, we present the results from the regressions where we include only the females with non-STEM fathers in our estimation sample. Panel A shows the results when we use any brother treatment (as in equation 1) and Panel B when we separately estimate the effect of having brother(s) only and having both brother(s) and sister(s) (as in equation 2). Although the estimates of any brother coefficient (brother only and brother and sister coefficients) in Panel A (Panel B) are negative—albeit very small in magnitude—and marginally significant without any controls, adding the full set of controls both reduces the magnitude of the coefficients and makes them not statistically significant. These results suggest that there is little, if any, evidence that sibling sex composition has any influence on whether a female chooses a STEM major in non-STEM families.

Next, we check for any differences in fertility patterns between STEM and non-STEM families. Different fertility patterns between these two groups of families might be signs of differences in preferences pertaining to child rearing that in turn might affect the major choice of females in college. Column 1 of Table 5 presents the estimated probability difference in having a son between STEM families and non-STEM families, while column 2 (3) shows the estimated difference in the likelihood of the second (third) child, conditional on having a daughter (two daughters). None of the differences are large and statistically significant.

4.3 Results

We first estimate equation (1) and present our results in panel A of Table 6. Column (1) shows the results from the specification without any parental or individual controls. The interaction

coefficient in this specification is negative and both statistically (but only marginally) and economically significant: having any brother(s) reduces the probability of choosing a STEM major in college by about 19 percentage points for females in families where the father is employed in a STEM job. Next, we add a frequently used a measure of cognitive skills, AFQT score, to see whether the interaction coefficient exhibits any change. If the fathers' influence is mostly through change in the cognitive skills of daughters, the inclusion of this measure should have a significant impact on the interaction coefficient, reducing it towards zero. When we add the standardized age-adjusted individual AFQT scores, however, our coefficient of interest exhibits almost no change (column 2 of Table 6).^{10,11} We then add parental and individual controls, and the estimate of the interaction coefficient again stays almost the same (column 3 of Table 6). In panel B, we present our results from the regression specification where we allow for having only brother(s) and having brother(s) and sister(s) to have different effects on the likelihood of females' STEM major choice. When we estimate equation (2) without any individual and parental controls, the interaction coefficient for females with only brother(s) in this specification increases to -0.248, which indicates that having only brother(s) decreases the probability of choosing a STEM major in college by almost 25 percentage points; the corresponding coefficient for females with brother(s) and sister(s) goes down to -0.151 and loses its statistical significance (p-value of 0.115), even though we cannot reject the equality of these two interaction coefficients. As in panel A, adding all our controls does not make any difference to our estimate of the interaction coefficient for having brother(s) and sister(s), while making the interaction coefficient for having only brother(s) slightly more negative (-0.276).

¹⁰ As expected, higher AFQT scores are associated with a higher likelihood of choosing a STEM major.

¹¹ In a recent working paper, Speer (2016) argues that pre-college skills as measured by ASVAB and not AFQT scores can explain a significant portion of the gender gap in STEM major choice. In the context of our paper, however, using subject specific ASVAB - rather than AFQT- scores does not change our results (see Table A5).

Having shown that, for females, having a brother has a large and negative impact on the likelihood of choosing a field of study similar to the occupation of their fathers, we next attempt to identify the potential mechanism for this impact. Although our analysis provides only tentative answers, it might still shed some light on the results presented in the first three columns. We first add being the first child and having an older brother dummies. If being the first child or having an older brother affects female competitiveness, the interaction coefficient might at least partially capture this correlation. Column 4 of Table 6, which presents our results from this specification, indicates that the interaction coefficient estimate is not sensitive to the inclusion of these variables either in panel A or panel B. We then include our attitude index as described in the data section and present the results in column 5. Our estimate of the interaction coefficients in panels A and B exhibit almost no change when this variable is included. Therefore, depending on the specification and the comparison group used, our results show that females who have only sisters are 15-28 percentage points more likely to choose a STEM major than females with different sibling sex composition if their fathers are also in a STEM occupation. Although these values seem quite large, they are surprisingly consistent with the upper bound estimates of Carrell et al. (2010), which showed that the highest ability women who were exclusively taught by women professors in their introductory math and science classes would be 26 percentage more likely to major in STEM fields than those who were exclusively taught by male faculty.

One potentially important caveat to the analysis thus far is the fact that we have only 115 females with STEM fathers in our estimation sample, of which 29 females have no brothers and 86 females have at least one brother. The small sample size used in the estimations may raise concern regarding the reliability of our results. In the next section, we attempt to address this concern using two additional datasets. The National Longitudinal Study of Youth, Children and

Young Adults (NLSCYA), which is an offshoot of NLSY79, gives us an opportunity to replicate the main analysis for similar but more recent data from the U.S. The Household Income and Labour Dynamics in Australia (HILDA), a nationally representative survey of Australians, provides us the largest sample size that contains all the information necessary for our analysis, and a different setting to test the father-daughter link in STEM major choice.

5. Evidence from Other NLSCYA and HILDA and Other Robustness Checks

5.1 National Longitudinal Study of Youth, Children and Young Adults

To investigate whether our main findings hold for the recent generation of youth in the U.S., we use a sample drawn from the National Longitudinal Study of Youth, Children and Young Adults (NLSCYA). NLSCYA collects information from the biological children of the women in the NLSY79. Note that NLSCYA is not a nationally representative sample of the comparable cohort in the U.S. However, with information almost identical to what is available in the NLSY79 survey, it is the ideal dataset to test the robustness of our main results.

In 1994, NLSCYA introduced the young adult section, where all children ages 15 and older are interviewed using questionnaires modeled after the NLSY79. Therefore, all key variables described previously are also available in the NLSCYA. One exception is how parents' occupation is collected. Unlike NLSY79, where father's occupation is reported only once (in the first cycle), NLSCYA collects father's occupation in all of the cycles since 1994. We identify a STEM father if the father of the respondent worked in a STEM job in any of the available years. In order to create a STEM mother indicator, we matched the NLSCYA sample with information about their mothers' from the NLSY79. We define STEM mothers as those who worked at least

once in a STEM occupation since 1994. Another difference between the two datasets is that college major in NLSCYA is slightly less detailed than that in NLSY79. More specifically, it is not possible to identify science fields within the Agriculture/Nature Resources major or within the Interdisciplinary Studies major.¹² We follow the sample selection rule we used to select the NLSY79 sample. The final NLSCYA sample consists of 1,038 women who report their major choices in college. NLSCYA contains a limited amount of family background characteristics when respondents were young. We augment this information by merging the NLSCYA sample with their mothers' information from NLSY79.

5.2 Household, Income and Labour Dynamics in Australia

The second dataset comes from the first twelve waves (years 2001 - 2012) of the Household, Income and Labour Dynamics in Australia (HILDA) survey. The HILDA survey is a nationally representative survey of the Australian population with detailed information, including labor market history, socio-demographic characteristics, including family background, life events and educational choices.¹³ We restrict the HILDA sample to women between 17 and 60 years old in 2012 who reported field of study and with no missing information. The final HILDA sample consists of 4,488 women.

Information on father's occupation for each respondent (current occupation for fathers who currently work, and past occupation for those who are retired or deceased) is available in all waves of HILDA. We use the occupational information from all waves up to and including wave 12 to identify STEM fathers. HILDA wave 12 contains information on the main field of study of

¹² For the NLSCYA sample, we include these two fields in the non-STEM major. The estimation results using NLSY79 sample do not change significantly when Agricultural Sciences and Interdisciplinary Sciences are not considered STEM fields.

¹³ See Watson and Wooden (2004) for the further details of this survey.

those who have completed or are currently enrolled in a post-secondary institution. We use this information to create a STEM major indicator.¹⁴ Family characteristics that are included are parents' employment and marital status when the respondent was 14, an indicator for each parent currently alive, immigration status and racial background, age as of 2012, a third degree polynomial of number of siblings, indicators for being the first sibling and having an older brother, age difference between the closest sibling, and STEM mother indicator.¹⁵ In waves 5, 8 and 11, HILDA collected information about respondents' attitudes toward many aspects of family life and gender issues, comparable to the opinions on gender roles available in NLSY79 and NLSCYA. For example, respondents are asked to give their opinion about the following statement: "*It is better for everyone involved if the man earns the money and the woman takes care of the home and children.*"¹⁶ We use data from wave 11 and sum over 6 relevant questions to create an Attitudes Index for HILDA. The Attitudes Index ranges from 6 (strongly disagrees with traditional gender roles) to 42 (strongly agrees with traditional gender roles).¹⁷

5.3 Main Differences among NLSY79, NLSYCYA, HILDA Samples

Before we report regression results using the alternative data sources, it is noteworthy to reiterate the differences between these data sets and NLSY79. First, NLSY79 is representative sample of U.S. youth, and NLSCYA, while not being a representative sample, is still a youth-focused survey. HILDA, on the other hand, is representative of the entire Australian population, and in

¹⁴ Note that the HILDA information on parents' occupation and field of study is less detailed than that of NLSY79 and NLSCYA. See Data Appendix for details.

¹⁵ See Table A2 for the full list of individual and family control variables included in the HILDA analysis.

¹⁶ Answer ranges from 1 (Strongly disagree) to 7 (Strongly agree).

¹⁷ The mean values of the variables used in the analyses for NLSCYA and HILDA are presented in the appendix tables A1 and A2, respectively.

order to retain sufficient sample size, we restrict the sample to 17–60 year olds.¹⁸ Second, cognitive score measures are not available in HILDA. Third, father’s occupation and college major measures are the most detailed in the NLSY79, slightly less so in NLSCYA, and considerably less so in HILDA. Finally, college major is collected longitudinally by NLS surveys, while it is reported only once, in wave 12 of HILDA.

5.4. Results from NLSCYA and HILDA

In Table 5 we provide results from the estimation of equations (1) and (2) using NLSCYA and HILDA. Panel A reports model results from estimating equation (1), while panel B reports the corresponding results for equation (2). For each sample, we present findings with and without individual and family characteristics.¹⁹

The results are largely consistent with our main findings from NLSY79. In the NLSCYA sample, women whose fathers worked in a STEM job are 16 percentage points less likely to choose a STEM major if they grew up with any brother compared to women with only sisters and STEM fathers. This impact is unchanged when individual and family characteristics are added. However, for both specifications, the interaction coefficient is statistically insignificant due to large standard errors. In HILDA, the corresponding impact is 8 percentage points when we do not control for individual and family characteristics and 7 percentage points when we do. Both of these interaction coefficients are statistically significant.

When we estimate regression specifications given in equation (2), the interaction coefficient of brother-only dummy and STEM father indicator is negative, large and statistically significant.

¹⁸ The magnitude of HILDA results does not change significantly when the sample is restricted to 17-40 year olds, although we lose significance.

¹⁹ Our analyses pertaining to threats to identification for NLSCYA and HILDA are presented in the Appendix tables A3 and A4. The results are similar to the ones from NLSY79, and provide further evidence on the credibility of females with non-STEM fathers as our control group.

Similar in magnitude to NLSY79 results, in NLCYA, having only brother(s) reduces the probability of choosing a STEM major in college by almost 22 percentage points compared to women with sisters only. In HILDA, the impact is also significant but not as large. When no controls are added, the estimated interaction coefficient implies an increase in probability of STEM major choice of around 10 percentage points; this effect drops to about 9 percentage points when the controls are included. As for the coefficient of having brother and sister dummy and STEM father indicator, the effects are smaller (yet still sizeable in magnitude) and not statistically significant in either dataset with the full set of controls.

The difference in the coefficient estimates using HILDA and the U.S. datasets might be attributable to at least three reasons. First, due to cultural and/or institutional differences between the two countries, fathers' influences on females' college major choice might be less significant in Australia than in the U.S. Second, differences in datasets may explain some of these differences. Both parents' occupation and major choice in HILDA are collected using broader categories than what is available in NLSY79 and NLSCYA.²⁰ Therefore, coefficients from the HILDA estimates may be biased downward due to measurement error. Also, HILDA covers a much wider range of birth cohorts (females aged 17-60 in 2012). Finally, estimates using HILDA (where, unlike NLSY79 and NLSCYA, we can observe the major with which one graduated from college) may be picking up fathers' influence on college completion as well as on choice of major. Therefore, smaller estimates can be explained if fathers' influence on college completion with a STEM degree is weaker than it is on college major.²¹

²⁰ Most notably, interdisciplinary and other engineering and agricultural and food science fields of study cannot be identified in HILDA. Moreover, father's occupation is recorded using 2-digit occupational coding in HILDA compared to 3-digit coding in NLSY79 and NLSCYA (more recent years of NLSCYA used 4-digit coding).

²¹ In fact, when we restrict the HILDA sample to those aged older than 24 who are not in full time education, the coefficient estimates decrease around 1 percentage point.

5.5 Other Robustness Checks

In this subsection, we perform several sensitivity checks using our main data sets, NLSY79. First, we limit the number of siblings to fewer than four, and then to fewer than three; the results are presented in Tables 8 and 9. The coefficient estimates from both of the robustness checks are similar to (if not larger than) those presented in Tables 6 and 7. We also use several different ways to define STEM major and occupation and run our main regressions with these new variables. Moreover, we restrict the sample to only whites and re-run our regressions. Our results are very similar to those presented in the paper.

6. Conclusion

Underrepresentation of women in STEM fields and its role in the gender earnings gap have attracted considerable attention in recent policy debate in the U.S. Several policy initiatives introduced since 2009, such as the Educate to Innovate campaign, the Invest in Innovation (I3) fund, and the National Science Foundation's Career-Life Balance Initiative, aim at improving women's interest and performance in STEM fields. In this paper, we present strong evidence that parents' involvement should also be an essential part of these initiatives. To the best of our knowledge, this is the first paper that provides causal evidence regarding intergenerational transmission of tastes and preferences as a potential contributing factor to the gender gap in STEM major choice.

We investigate the role of fathers in explaining the differences between males and females in the choice of a STEM major in college. The main innovation of the paper is to analyze how sibling sex composition affects the probability of being a STEM major in college

for females with fathers in STEM occupations. We argue that, for women, growing up with brothers can limit or alter the transmission of occupational-specific preferences that can be passed on from fathers. Other possible mechanisms are that sibling sex composition can define or strengthen females' perception of what is a "*gender appropriate*" career, or can affect females' competitiveness and, thus, their career choices. We find empirical evidence in support of the role of sibling sex composition from several data sources. Our main findings from the NLSY79 sample suggest that, depending on the specification and comparison group used, females who have brother(s) and have fathers employed in a STEM occupation are 15-28 percentage points less likely to choose a STEM major than females with sister(s) only, and the effect appears to be driven by change in college major preferences of females. The inclusion of respondents' attitudes toward traditional gender roles, birth order, the presence of an older brother, cognitive skill measures and other individual and family characteristics do not change our results. We also show that the results are robust to how STEM occupations are defined and to the sample restriction based on family size and race. Finally, findings using NLSCYA and HILDA provide further evidence of the importance of fathers as a significant driver of female STEM major choice in college for the U.S. and Australia.

There are some fruitful areas for future research. First, it would be useful to analyze whether the intergenerational effect on the gender gap in STEM major choice in college we presented here exists in occupational choice by men and women. Second, conducting a similar analysis for countries with different cultural norms and educational institutions than the U.S. and Australia might shed some light on the generalizability of our results. Finally, replicating our analysis with a larger dataset (if there is one) which includes detailed information on parental occupation, college major choice of females and their siblings' education and sex would be

beneficial. With a larger dataset, it might even be possible to check whether the effects we find in this paper for fathers exist for mothers in STEM occupations as well.

References

- American Association of University Women Educational Foundation. 2007. *Behind the Pay Gap*. Washington, DC.
- Arcidiacono, P. 2004. "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121: 343-375.
- Bharadwaj P., G. De Giorgi, D. Hansen, and C. Neilson. 2012. "The Gender Gap in Mathematics: Evidence from Low- and Middle-Income Countries," NBER Working Paper #18464.
- Bjorklund, A. and Jantti, M. 2009. "Intergenerational Income Mobility and the Role of Family Background," in W. Salverda, B. Nolan & T. Smeeding, eds, *Oxford Handbook of Economic Inequality*, Oxford: Oxford University Press: 491-521.
- Bjorklund, A. and Salvanes, K. G. 2011. "Education and Family Background: Mechanisms and Policies," in *Handbook of the Economics of Education*, 3 edn, Elsevier: 201-247.
- Black, S. E. and Devereux, P. J. 2011. "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, 4 edn., Elsevier: 1487-1541.
- Brown, C. and M. Corcoran. 1997. "Sex-Based Differences in School Content and the Male-Female Wage Gap," *Journal of Labor Economics*, 15: 431-465.
- Butcher, K. and A. Case. 1994. "The Effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics* 109: 531-563.
- Carrell, S.E., M.E. Page, and J.E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *Quarterly Journal of Economics* 125: 1101-1144.
- Dahl, G. and E. Moretti. 2008. "The Demand for Sons," *Review of Economic Studies*, 75: 1085-1120.
- Eccles, J.S. 1994. "Understanding Women's Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices," *Psychology of Women Quarterly*, 18: 585-609.
- Eccles, J.S. and L.W. Hoffman. 1984. "Sex Roles, Socialization, and Occupational Behavior," in H.W. Stevenson and H.E. Siegel (eds.) *Child Development Research and Social Policy*, Chicago: University of Chicago Press.
- Frehill, L. M., Di Fabio, N., Hill, S., Trager, K., & Buono, J. 2008. "Women in Engineering: A Review of the 2007 Literature," *SWE Magazine*, 54(3): 6-30.
- Fryer, R. G., and S. D. Levitt. 2010. "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2: 210-240.

- Guiso, L. F. Monte, P. Sapienza and L. Zingales. 2008. "Culture, Gender, and Math," *Science*, 320: 1164-1165.
- Hauser, R. M. and H. D. Kuo. 1998. "Does the Gender Composition of Sibships Affect Women's Educational Attainment?" *Journal of Human Resources*, 33: 644-657.
- Hewlett, S. A., Buck Luce, C., Servon, L. J., Sherbin, L., Shiller, P., Sosnovich, E., and Sumberg, K. 2008. "The Athena Factor: Reversing the Brain Drain in Science, Engineering and Technology, Harvard Business Review Research Report, Boston: Harvard Business Publishing.
- Hill, C., C. Corbett, and A. St. Rose. 2010. "Why So Few? Women in Science, Technology, Engineering, and Mathematics," AAUW, Washington, DC.
- Hyde, J.S., E. Fennema, and S. Lamon. 1990. "Gender Differences in Mathematics Performance: A Meta-Analysis," *Psychological Bulletin*, 107: 139-155.
- Hyde, J.S., S.M. Lindberg, M.C. Linn, A.B. Ellis, and C.C. Williams. 2008. "Gender Similarities Characterize Math Performance," *Science*, 321: 494-495.
- Hunt, Jennifer. 2016. "Why Do Women Leave Science and Engineering?" *Industrial and Labor Relations Review*, forthcoming.
- Kaestner, R. 1997. "Are Brothers Really Better? Sibling Sex Composition and Educational Achievement Revisited," *Journal of Human Resources*, 32: 250-284.
- Lee, J., W. Grigg, and G. Dion. 2007. *The Nation's Report Card: Mathematics 2007*. Washington DC: National Center for Education Statistics.
- Lundberg, S. 2005. "Sons, Daughters, and Parental Behavior," *Oxford Review of Economic Policy*, 21: 340-356.
- National Center for Education Statistics. 2012. "The Digest of Education Statistics 2011" NCES 2012-001, U.S. Department of Education.
- Niederle, M. and L. Vesterlund. 2007. "Do Women Shy away from Competition? Do Men Compete too Much?," *Quarterly Journal of Economics*, 122: 1067-1101.
- Ma, Yingyi. 2011. "College Major Choice, Occupational Structure and Demographic Patterning by Gender, Race and Nativity." *The Social Science Journal*, 48(1): 112-129.
- Solon, G. 1999. "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, 3 edn, Elsevier: 1761-1800.
- Speer, J. 2016. "Gender Gap in College Major: Revisiting the Pre-College Factors." Mimeo, University of Memphis.

Stoneman, Z., G. Brody, and C. MacKinnon. 1986. "Same Sex and Cross-Sex Siblings: Activity Choices, Roles, Behavior and Gender Stereotyping," *Sex Roles*, 15: 495-511.

Turner, S.E. and Bowen, W.G., 1999. "The Changing (Unchanging) Gender Gap," *Industrial and Labor Relations Review*, 52(2): 289-313.

The White House, Office of Science and Technology Policy. (2013) "Fact Sheet: Women and Girls in Science, Technology, Engineering, and Math (STEM)," Retrieved from: <https://www.whitehouse.gov/administration/eop/ostp/women>.

U.S. Bureau of Labor Statistics. 2013. "Highlights of Women's Earnings in 2012," Bureau of Labor Statistics Reports, Washington, DC.

U.S. Department of Commerce. 2011. "Women in STEM: A Gender Gap to Innovation," Executive Summary, Washington, DC.

Watson, N., and Wooden, M. (2004). "HILDA Survey Four Years on," *The Australian Economic Review*, 37(3): 343-349.

Weinberger, C. J. 1999. "Mathematical College Majors and the Gender Gap in Wages," *Industrial Relations*, 38: 407-413.

Zafar, B. 2013. "College Major Choice and Gender Gap," *Journal of Human Resources*, 48(3): 545-595.

Data Appendix

STEM Fields for Parents' Occupation

NLSY79, NLSCYA and HILDA use different occupational coding. In the case of NLSCYA, there are also variations within the survey years since different occupational coding is utilized for different years. We summarize across and within survey differences in occupational coding below. Note that, in order to identify fathers who work in STEM jobs, we follow STEM occupation definition provided by the US Census Bureau²² for all data sources as closely as possible. Using 2010 Census occupational coding, the US Census Bureau breaks down occupations into the following 3 categories: STEM occupations, STEM-related occupations, and non-STEM occupations. We choose only the STEM category to identify STEM occupations.

NLSY79

In NLSY79 parents' occupation is collected only once at the beginning of the survey. Parents' occupation is coded according to the 3-digit 1970 Census occupational code.

NLSCYA

In NLSCYA, father's occupation is collected for young adults (aged 15 and older) bi-annually since 1994. We identify STEM fathers as those who have worked at least once in a STEM job since 1994. NLSCYA use different occupational coding depending on the survey years. These

²² A complete list from the Census Bureau can be accessed at <http://www.census.gov/people/io/files/STEM-Census-2010-occ-code-list.xls>.

can be summarized as follows: for survey years between 1994 and 2000, 3-digit 1970 Census occupational code; for the survey year 2002, 3-digit 2000 Census occupational code; for survey years between 2004 and 2012, 4-digit 2000 Census occupational code.

We merge occupation of mothers from the original NLSY79 sample (3-digit 1970 Census occupational code) to identify STEM mothers. STEM mothers are defined as those who have worked at least once in a STEM job since 1994.

HILDA

In HILDA, information on parents' current and previous jobs is collected for all survey years that the respondents were interviewed. HILDA uses 2-digit International Standard Classification of Occupation-88 (ISCO-88) and 2-digit Australian and New Zealand Standard Classification of Occupations (ANZSCO 2006) to classify occupation. Both of these coding classifications are much less detailed than what is available in the NLS surveys.

We create a STEM father indicator for the HILDA sample using all survey years and the 2-digit ISCO-88 coding, although using ANZSCO 2006 coding or both ISCO-88 and ANZSCO 2006 together does not change the results in any significant manner.

Available occupations are categories based on 2-digit ISCO-88 coding in HILDA are (STEM occupations in **bold**):

Armed forces

Legislators, senior officials and managers n.e.c

Legislators and senior officials

Corporate managers

General managers

Professionals n.e.c

Physical, mathematical and engineering science professionals

Life science and health professional

Teaching professionals

Other professionals
Technicians and associate professionals n.e.c
Physical and engineering science associate professionals
Life science and health associate professionals
Teaching associate professionals
Other associate professionals
Clerks n.e.c
Office clerks
Customer service clerks
Service workers and shop and market sales workers n.e.c
Personal and protective services workers
Models, salespersons and demonstrators
Skilled agricultural and fishery workers n.e.c
Market-oriented skilled agricultural and fishery workers
Subsistence agricultural and fishery workers
Craft and related trade workers n.e.c
Extraction and building trade workers
Metal, machinery and related trades workers
Precision, handicraft, printing and related trades workers
Other craft and related trades workers
Plant and machine operators and assemblers n.e.c
Stationary plant and related operators
Machine operators and assemblers
Drivers and mobile plant operators
Elementary occupations n.e.c
Sales and services elementary occupations
Agricultural, fishery and related labourers
Labourers in mining, construction, manufacturing, transport

STEM Fields for College Majors

NLSY79

NLSY79 provides a 4-digit Field of Study in College since the beginning of the survey. Starting from 1984, this information was extended to the 3 most recent colleges. In our sample, an individual is identified as choosing a STEM major if she reported a STEM major in any of the years that she was interviewed. According to our definition, STEM fields are Agricultural

Sciences (e.g. Agronomy, Soil Science, Animal Science and Food Sciences), Biological Sciences, Computer and Information Sciences, Engineering, Mathematics, Physical Sciences, Interdisciplinary Biological and Physical Sciences and Interdisciplinary Engineering, and Other Disciplines.²³

NLSYCYA

Major in college in NLSYCYA has been collected bi-annually since 1994. The coding of the majors is slightly less detailed than NLSY79. Available codes are as follows (STEM fields in **bold**)

None, General Studies, Undecided

Agriculture/Natural Resources

Architecture/Environ Design

Area Studies

Biological Sciences

Pre-Med

Pre-Vet

Pre-Dental

Business Management

Communications

Computer/Information Science

Education

Engineering

Fine and Applied Arts

Foreign Languages

Nursing

Other Health Professions

Home Economics

Pre-Law

English

Philosophy

Mathematics

²³ A full list of fields of study available in NLSY79 can be accessed at

<https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-4-fields-study>.

Physical Sciences

Psychology
Anthropology
Archaeology
Economics
History
Political Science/Government
Sociology
Criminology
Theology/Religious Studies
Interdisciplinary Studies
Other Field (Specify)
Ethnic Studies

We identify a STEM major in NLSYCYA if a respondent reported a STEM major in any of the available survey years.

HILDA

In HILDA, the field of study of highest post-school qualification is collected only in wave 12. Available majors in HILDA are much less detailed than what is available in NLS surveys; moreover, unlike NLS surveys, we do not observe previous college degrees. In order to capture individuals with previous STEM degrees prior to entering medical school, we include medicine as a STEM field in HILDA.

Field of study categories available in HILDA are listed below (STEM fields in **bold**):

Natural and physical sciences
Information technology
Engineering and related technologies
Architecture and building
Agriculture, environment and related studies
Medicine

Nursing
Other health-related (e.g. Pharmacy, Dental studies, Rehabilitation)
Education
Management and commerce (e.g. Accounting, Business, Sales)
Law
Society and culture (e.g. Economics, Political science)
Creative arts
Food, hospitality and personal services
Other (please specify)

Figure 1 - Percentage of BA STEM Degrees Awarded to Females, 1966-2006

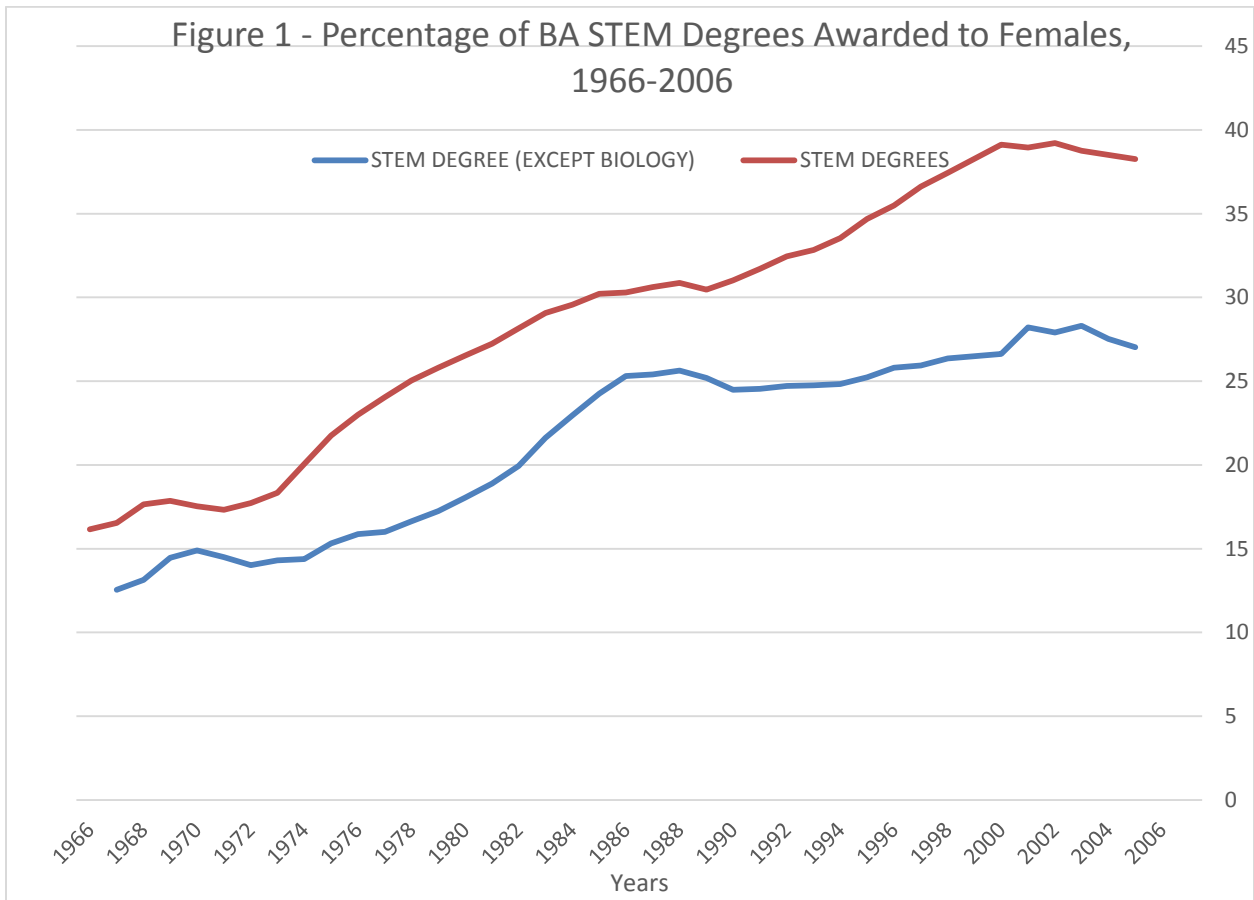


Table 1: Mean of Individual and Family Characteristics by Sibling Gender Composition, Female

	(1) Sisters Only	(2) Sisters & Brothers or Brothers only	(3) Brothers Only	(4) Sisters & Brothers
STEM Father	0.0861	0.0516	0.0721	0.0450
STEM Mother	0.0119	0.0018	0	0.00237
Hispanic	0.151	0.17	0.149	0.177
Black	0.184	0.24	0.154	0.268
White or Asian	0.665	0.589	0.697	0.555
at age 14:				
Lived in U.S.	0.973	0.985	0.99	0.983
Lived in urban	0.783	0.801	0.803	0.800
Lived with				0.796
parents	0.825	0.811	0.856	
Mother worked	0.605	0.555	0.604	0.539
Father worked	0.896	0.899	0.91	0.895
Father is immigrant	0.11	0.0911	0.0995	0.0885
Father is alive now	0.335	0.314	0.353	0.302
Mother is				0.100
immigrant	0.134	0.103	0.109	
Mother is alive now	0.475	0.436	0.453	0.431
Mother Occupation				
Missing	0.318	0.372	0.328	0.385
Mother went				0.221
college	0.315	0.245	0.321	
Father went college	0.401	0.312	0.388	0.288
Log(Income) in				
1978	7.926	7.769	7.648	7.807
AFQT Score	0.479	0.253	0.464	0.304
Number of Siblings	2	3.713	1.848	4.306
is first Child	0.415	0.247	0.41	0.195
has Older Brother	0	0.626	0.59	0.637
Age Diff. with				
Closest Sibling	3.411	3.850	3.244	4.044
Attitudes Index	1.947	1.821	1.677	1.867
Observations	337	1,668	402	1,266

Note: Sample excludes observations with missing college major and missing father occupation. AFQT is normalized using mean and standard deviation of NLSY79 sample, then age-standardized.

Table 2: STEM Major Choice by Gender & Father's Occupation

	Male	Female
Sample:		
All	0.479	0.263
STEM Father	0.590	0.304
Non-STEM Father	0.472	0.260
Observations	1,742	2,005

Note: Statistics represent fraction of observations who chose STEM major in college. Sample excludes observations with missing major and father occupation.

Table 3: STEM Major Choice by Sibling Gender Composition & Father's Occupation, Female

	Father's Occupation		
	All	STEM Father	Non-STEM Father
Sibling Gender Composition:			
Sisters & Brothers or Brothers only	0.252	0.244	0.253
Sisters & Brothers	0.255	0.263	0.254
Sisters Only	0.314	0.482	0.298
Brothers Only	0.228	0.172	0.233
Observations	2,005	115	1,890

Note: Statistics represent fraction of observations who chose STEM major in college. Sample excludes observations with missing college major and father occupation.

Table 4: Non-STEM Families, NLSY79

PANEL A :				
	(1)	(2)	(3)	(4)
Any Brother	-0.0464*	-0.0374	-0.0373	-0.0381
	(0.0272)	(0.0293)	(0.0335)	(0.0336)
Is first child			-0.00475	-0.00456
			(0.0317)	(0.0317)
Has Older Brother			-0.00103	-0.000539
			(0.0309)	(0.0309)
Attitudes Index				-0.00343
				(0.00625)
Individual & Family Controls	NO	YES	YES	YES
PANEL B :				
	(1)	(2)	(3)	(4)
Brother Only	-0.0649*	-0.0595*	-0.0590	-0.0600
	(0.0336)	(0.0338)	(0.0385)	(0.0386)
Brother & Sister	-0.0407	-0.0170	-0.0210	-0.0219
	(0.0279)	(0.0332)	(0.0364)	(0.0364)
Is first child			-0.00140	-0.00119
			(0.0318)	(0.0318)
Has Older Brother			0.00355	0.00406
			(0.0311)	(0.0311)
Attitudes Index				-0.00348
				(0.00625)
Individual & Family Controls	NO	YES	YES	YES
Observations	1890	1890	1882	1882

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Characteristics that are included in columns 2 - 4 are listed in Table 3, models in column (2) - (4) also include quadratic and cubic terms of number of siblings.

Table 5: STEM-Father Effect on Sibling Gender Composition, NLSY79

Panel A: Without Family Controls

	Probability of Son	Probability of 2 nd Children after daughter	Probability of 3 rd Children after 2 daughters
STEM Father	0.000952 (0.0254)	-0.0304 (0.0464)	-0.0289 (0.0624)

Panel B: With Family Controls

	Probability of Son	Probability of 2 nd Children after daughter	Probability of 3 rd Children after 2 daughters
STEM Father	-0.00312 (0.0263)	-0.0492 (0.0485)	0.00999 (0.0674)
Observations	8980	2258	697

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Family controls that are included in Panel B results are: STEM occupation dummy for the mother, a missing occupation indicator for the mother, logarithm of family income in 1978, whether mother worked at age 14, whether father worked at age 14, whether mother and father are still alive, father's and mother's immigration status, indicators for mother's and father's highest degree of education

Table 6: Sibling Composition Models, NLSY79

PANEL A :	(1)	(2)	(3)	(4)	(5)
Any Brother	-0.0402 (0.0283)	-0.0338 (0.0284)	-0.0293 (0.0299)	-0.0350 (0.0339)	-0.0359 (0.0339)
STEM Father	0.184* (0.0965)	0.154 (0.0978)	0.159 (0.0982)	0.160 (0.0986)	0.163* (0.0985)
Any Brother x STEM Father	-0.187* (0.108)	-0.185* (0.109)	-0.191* (0.109)	-0.192* (0.110)	-0.196* (0.110)
AFQT		0.0387*** (0.0111)	0.0515*** (0.0135)	0.0514*** (0.0135)	0.0500*** (0.0137)
Is first child				0.00101 (0.0314)	0.00102 (0.0314)
Has Older Brother				0.00752 (0.0302)	0.00787 (0.0302)
Attitudes Index					-0.00369 (0.00593)
Individual & Family Controls	NO	NO	YES	YES	YES
PANEL B :	(1)	(2)	(3)	(4)	(5)
Brother Only	-0.0628* (0.0342)	-0.0607* (0.0341)	-0.0535 (0.0341)	-0.0628 (0.0388)	-0.0638 (0.0388)
STEM Father	0.184* (0.0965)	0.153 (0.0979)	0.159 (0.0980)	0.160 (0.0983)	0.163* (0.0983)
Brother Only x STEM Father	-0.248** (0.121)	-0.251** (0.123)	-0.272** (0.124)	-0.272** (0.124)	-0.276** (0.124)
Brother & Sister	-0.0332 (0.0291)	-0.0252 (0.0291)	-0.00300 (0.0336)	-0.0113 (0.0365)	-0.0121 (0.0365)
Brother & Sister x STEM Father	-0.151 (0.115)	-0.146 (0.115)	-0.150 (0.116)	-0.150 (0.116)	-0.155 (0.116)
AFQT		0.0403*** (0.0111)	0.0512*** (0.0135)	0.0511*** (0.0135)	0.0496*** (0.0137)
Is first child				0.00561 (0.0315)	0.00565 (0.0315)
Has Older Brother				0.0145 (0.0305)	0.0149 (0.0305)
Attitudes Index					-0.00385 (0.00594)
Individual & Family Controls	NO	NO	YES	YES	YES
Observations	2,005	2,005	2,005	1,997	1,997

Note: Sample only includes college graduates. Characteristics that are included in columns 3 - 5 are listed in Table 3, models in column (3) - (5) also include quadratic and cubic terms of number of siblings.

Table 7: Evidence from Other Datasets

PANEL A :	NLSCYA		HILDA	
	(1)	(2)	(3)	(4)
Any Brother	0.00799 (0.0206)	0.00420 (0.0273)	-0.00982 (0.0105)	0.00849 (0.0131)
STEM Father	0.164 (0.110)	0.147 (0.109)	0.135*** (0.0375)	0.109*** (0.0368)
Any Brother x STEM Father	-0.163 (0.119)	-0.158 (0.121)	-0.0825** (0.0410)	-0.0732* (0.0399)
Is first child		0.00334 (0.0283)		-0.000797 (0.0126)
Has Older Brother		0.0230 (0.0312)		-0.0162 (0.0129)
Attitudes Index		-0.00311 (0.0136)		-0.0000291 (0.000439)
Individual & Family Controls	NO	YES	NO	YES
PANEL B :	NLSCYA		HILDA	
	(1)	(2)	(3)	(4)
Brother only	0.0213 (0.0247)	0.0109 (0.0300)	0.00127 (0.0128)	0.00925 (0.0152)
STEM Father	0.164 (0.110)	0.146 (0.109)	0.135*** (0.0375)	0.109*** (0.0368)
Brother only x STEM Father	-0.219* (0.123)	-0.210* (0.123)	-0.0962** (0.0480)	-0.0895* (0.0466)
Brother & Sister	-0.00240 (0.0223)	-0.00287 (0.0335)	-0.0152 (0.0109)	0.00855 (0.0140)
Brother& Sister x STEM Father	-0.117 (0.131)	-0.115 (0.135)	-0.0760* (0.0425)	-0.0660 (0.0414)
Is first child		0.00310 (0.0280)		-0.000585 (0.0126)
Has Older Brother		0.0211 (0.0313)		-0.0159 (0.0131)
Attitudes Index		-0.00305 (0.0136)		-0.0000262 (0.000439)
Individual & Family Controls	NO	YES	NO	YES
Observations	1,038	1,025	4,488	4,476

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See Table 1, Table A1 and Table A2 for Individual & Family Controls that were added in NLSY79, NLSCYA and HILDA models respectively. All models also include quadratic and cubic terms of number of siblings. NLSYA and HILDA models do not include a cognitive skill measure (e.g. AFQT or ASVAB).

Table 8: Sibling Composition Models, Respondents with less than 4 Siblings

PANEL A :	NLSY79	NLSCYA	HILDA
	(1)	(2)	(3)
Any Brother	-0.0563 (0.0383)	0.00796 (0.0295)	0.0120 (0.0143)
STEM Father	0.156 (0.0999)	0.148 (0.109)	0.114*** (0.0386)
Any Brother x STEM Father	-0.163 (0.116)	-0.179 (0.120)	-0.0694 (0.0432)
Is first child	-0.00639 (0.0370)	0.000942 (0.0321)	-0.00666 (0.0145)
Has Older Brother	0.000575 (0.0388)	0.0224 (0.0373)	-0.0221 (0.0164)
Attitudes Index	-0.0109 (0.00762)	-0.00468 (0.0191)	-0.000425 (0.000536)
Individual & Family Controls	YES	YES	YES
PANEL B :	NLSY79	NLSCYA	HILDA
	(1)	(2)	(3)
Brother only	-0.0673 (0.0429)	0.0131 (0.0317)	0.0160 (0.0167)
STEM Father	0.157 (0.0998)	0.147 (0.109)	0.114*** (0.0386)
Brother only x STEM Father	-0.239* (0.127)	-0.210* (0.124)	-0.0955** (0.0487)
Brother & Sister	-0.0435 (0.0422)	0.000905 (0.0367)	0.00824 (0.0153)
Brother& Sister x STEM Father	-0.102 (0.132)	-0.147 (0.135)	-0.0518 (0.0467)
Is first child	-0.00359 (0.0373)	0.000607 (0.0316)	-0.00676 (0.0145)
Has Older Brother	0.00618 (0.0395)	0.0209 (0.0372)	-0.0226 (0.0168)
Attitudes Index	0.00394 (0.00361)	-0.00486 (0.0191)	-0.000422 (0.000536)
Individual & Family Controls	YES	YES	YES
Observations	1,228	903	3,343

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See Table 1, Table A1 and Table A2 for Individual & Family Controls that were added in NLSY79, NLSCYA and HILDA models respectively. All models also include quadratic and cubic terms of number of siblings. NLCYA and HILDA models do not include a cognitive skill measure (e.g. AFQT or ASVAB).

Table 9: Sibling Composition Models, Respondents with less than 3 Siblings

PANEL A :	NLSY79	NLSCYA	HILDA
	(1)	(2)	(3)
Any Brother	-0.0358 (0.0459)	-0.00513 (0.0324)	0.00481 (0.0160)
STEM Father	0.171 (0.114)	0.189 (0.120)	0.127*** (0.0419)
Any Brother x STEM Father	-0.222 (0.135)	-0.227* (0.132)	-0.0978** (0.0475)
Is first child	-0.0306 (0.0472)	0.0104 (0.0357)	-0.00272 (0.0167)
Has Older Brother	-0.00690 (0.0547)	0.0455 (0.0439)	-0.0180 (0.0196)
Attitudes Index	-0.0172* (0.00911)	0.00701 (0.0233)	0.000110 (0.000634)
Individual & Family Controls	YES	YES	YES
PANEL B :	NLSY79	NLSCYA	HILDA
	(1)	(2)	(3)
Brother only	-0.0630 (0.0505)	0.00426 (0.0343)	0.0121 (0.0185)
STEM Father	0.172 (0.113)	0.188 (0.121)	0.127*** (0.0419)
Brother only x STEM Father	-0.299** (0.139)	-0.295** (0.124)	-0.104** (0.0526)
Brother & Sister	-0.00349 (0.0523)	-0.0235 (0.0438)	-0.00712 (0.0179)
Brother& Sister x STEM Father	-0.0936 (0.177)	-0.129 (0.164)	-0.0896* (0.0533)
Is first child	-0.0205 (0.0477)	0.0106 (0.0351)	-0.00378 (0.0167)
Has Older Brother	0.0142 (0.0562)	0.0424 (0.0434)	-0.0210 (0.0203)
Attitudes Index	-0.0181** (0.00921)	0.00686 (0.0232)	0.000125 (0.000633)
Individual & Family Controls	YES	YES	YES
Observations	810	702	2480

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See Table 1, Table A1 and Table A2 for Individual & Family Controls that were added in NLSY79, NLSCYA and HILDA models respectively. All models also include quadratic and cubic terms of number of siblings. NLCYA and HILDA models do not include a cognitive skill measure (e.g. AFQT or ASVAB).

APPENDIX:

Table A1 – Mean of Individual and Family Characteristics by Sibling Gender Composition, NLSCYA

	(1)	(2)	(3)	(4)
	Sister only	Brother & Sister	Bro & Sist. or Brother only	Brother only
STEM Father	0.0588	0.0535	0.0548	0.0565
STEM Mother	0.0487	0.0778	0.0633	0.0449
Number of Siblings	1.562	3.114	2.385	1.452
is first Child	0.452	0.323	0.415	0.533
has Older Brother	0	0.505	0.488	0.467
Age Difference with Closest Sibling	3.890	4.723	4.521	4.262
Age in 2012	27.57	28.18	28.14	28.08
Hispanic	0.202	0.244	0.226	0.202
Black	0.456	0.430	0.410	0.384
White	0.342	0.326	0.364	0.414
at age 14:				
Father worked	0.745	0.672	0.706	0.749
Mother worked	0.348	0.417	0.416	0.413
Mother was married	0.326	0.337	0.339	0.341
Father is alive now	0.217	0.256	0.257	0.259
Mother immigrant	0.107	0.0744	0.0809	0.0893
Parents together now	0.00368	0.00233	0.00392	0.00595
Mother went college	0.401	0.356	0.363	0.372
Father went college	0.0515	0.0326	0.0444	0.0595
Number of Siblings	1.562	3.114	2.385	1.452
Mother occupation missing	0.0300	0.0778	0.0594	0.0359
Attitudes Index	0.147	0.242	0.222	0.196
Observations	272	430	766	336

Table A2 – Mean of Individual and Family Characteristics by Sibling Gender Composition, HILDA

	(1) Sister only	(2) Brother & Sister	(3) Bro & Sist. or Brother only	(4) Brother only
STEM Father	0.131	0.128	0.123	0.115
STEM Mother	0.113	0.120	0.116	0.107
Australian Born	0.781	0.769	0.773	0.781
Immigrant: ESB	0.0866	0.0753	0.0782	0.0843
Immigrant: NESB	0.115	0.129	0.124	0.115
ATSI	0.0173	0.0269	0.0248	0.0202
Age at 2012	36.05	39.17	38.40	36.82
Number of Siblings	1.701	3.811	3.115	1.664
Age Difference btw Closest Sibling	3.172	4.910	4.407	3.356
is first Child	0.440	0.261	0.307	0.401
has Older Brother	0	0.591	0.593	0.599
Father Unemployed more than 6 months while growing up	0.134	0.146	0.141	0.131
Mother Occupation Missing	0.146	0.187	0.172	0.141
Mother Immigrant	0.331	0.333	0.339	0.352
Father Immigrant	0.377	0.350	0.352	0.354
Mother went University	0.338	0.332	0.348	0.376
Father went University	0.317	0.349	0.362	0.388
Mother is alive now	0.908	0.879	0.885	0.896
Father is alive now at age 14:	0.789	0.718	0.746	0.801
Father worked	0.952	0.931	0.937	0.949
Mother Worked	0.709	0.571	0.603	0.671
Lived with Parents	0.849	0.771	0.793	0.837
Attitudes Index	15.94	16.06	15.97	15.77
Observations	982	2,376	3,515	1,139

Note: ESB is English Speaking background, NESB is non-English speaking Background, ATSI is Aboriginal or Torres Strait Islander.

Table A3: Non-STEM Families, NLSCYA and HILDA

PANEL A :				
	NLSCYA	NLSCYA	HILDA	HILDA
Any Brother	0.00799 (0.0210)	0.00448 (0.0277)	-0.00982 (0.0105)	0.00382 (0.0129)
Is first child		-0.000993 (0.0287)		0.00780 (0.0125)
Has Older Brother		0.0248 (0.0314)		-0.0151 (0.0128)
Attitudes Index		-0.00590 (0.0165)		-0.0000547 (0.000457)
Individual & Family Controls	NO	YES	NO	YES
PANEL B :				
	(1)	(2)	(3)	(4)
Brother Only	0.0213 (0.0243)	0.00833 (0.0297)	0.00127 (0.0128)	0.0121 (0.0153)
Brother & Sister	-0.00240 (0.0231)	-0.00217 (0.0333)	-0.0152 (0.0109)	-0.00184 (0.0133)
Is first child		-0.00182 (0.0288)		0.00594 (0.0125)
Has Older Brother		0.0240 (0.0315)		-0.0176 (0.0132)
Attitudes Index		-0.00587 (0.0165)		-0.0000568 (0.000457)
Individual & Family Controls	NO	YES	NO	YES
Observations	980	967	3,926	3,916

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample only includes non-STEM father families.

Table A4: STEM-Father Effect on Sibling Gender Composition, NLSCYA and HILDA

Panel A: NLSCYA - With Family Controls

	Probability of Son	Probability of 2 nd Children after daughter	Probability of 3 rd Children after 2 daughters
STEM Father	-0.00638 (0.0417)	-0.0371 (0.0549)	-0.00786 (0.0787)
Observations	3086	1551	646

Panel B: HILDA - With Family Controls

	Probability of Son	Probability of 2 nd Children after daughter	Probability of 3 rd Children after 2 daughters
STEM Father	-0.0192 (0.0163)	-0.00237 (0.0176)	-0.0193 (0.0205)
Observations	8,749	6,276	3,527

Note: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results that exclude Family Controls are not significantly different than the reported results above.

Table A5 Panel A - ASVAB Scores, up to 13 siblings

PANEL A:	GS	AR	NO	CS	AS	MK	MC	EI	ALL
Any Brother	-0.0352 (0.0342)	-0.0341 (0.0341)	-0.0309 (0.0343)	-0.0346 (0.0342)	-0.0320 (0.0342)	-0.0350 (0.0342)	-0.0407 (0.0343)	-0.0314 (0.0342)	-0.0419 (0.0342)
STEM Father	0.161* (0.0946)	0.159 (0.0972)	0.188* (0.0980)	0.186* (0.0972)	0.173* (0.0957)	0.148 (0.0980)	0.153 (0.0966)	0.159* (0.0944)	0.135 (0.0956)
Any Brother x STEM Father	-0.209** (0.106)	-0.194* (0.109)	-0.210* (0.110)	-0.208* (0.109)	-0.202* (0.108)	-0.184* (0.110)	-0.201* (0.108)	-0.197* (0.106)	-0.190* (0.107)
ASVAB Score	0.0706*** (0.0147)	0.0668*** (0.0128)	0.0281** (0.0136)	0.0375*** (0.0124)	0.0522*** (0.0180)	0.0606*** (0.0124)	0.0819*** (0.0145)	0.0643*** (0.0150)	- -
PANEL B :	GS	AR	NO	CS	AS	MK	MC	EI	ALL
Brother Only	-0.0627 (0.0394)	-0.0609 (0.0394)	-0.0599 (0.0396)	-0.0630 (0.0395)	-0.0613 (0.0396)	-0.0628 (0.0394)	-0.0650 (0.0397)	-0.0583 (0.0395)	-0.0656* (0.0395)
STEM Father	0.160* (0.0943)	0.158 (0.0970)	0.187* (0.0977)	0.185* (0.0970)	0.172* (0.0954)	0.148 (0.0978)	0.152 (0.0964)	0.159* (0.0942)	0.134 (0.0954)
Brother Only x STEM Father	-0.298** (0.120)	-0.279** (0.123)	-0.288** (0.123)	-0.288** (0.123)	-0.287** (0.121)	-0.269** (0.123)	-0.289** (0.122)	-0.278** (0.119)	-0.281** (0.122)
Brother & Sister	-0.0120 (0.0367)	-0.0114 (0.0366)	-0.00675 (0.0369)	-0.0108 (0.0368)	-0.00742 (0.0367)	-0.0116 (0.0367)	-0.0197 (0.0368)	-0.00879 (0.0368)	-0.0213 (0.0368)
Brother & Sister x STEM Father	-0.162 (0.113)	-0.149 (0.115)	-0.167 (0.116)	-0.166 (0.115)	-0.155 (0.115)	-0.139 (0.116)	-0.154 (0.114)	-0.153 (0.113)	-0.141 (0.113)
ASVAB Score	0.0706*** (0.0147)	0.0665*** (0.0128)	0.0276* (0.0136)	0.0371*** (0.0124)	0.0530*** (0.0179)	0.0605*** (0.0125)	0.0813*** (0.0145)	0.0636*** (0.0150)	- -
Observations	1937	1937	1937	1937	1937	1937	1937	1937	1937

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0$. All ASVAB scores are standardized for each yearly birth cohort. GS: General Science, AR: Arithmetic Reasoning, NO: Numerical Operations, CS : Coding Speed, AS: Auto and Shop Knowledge, MK: Mathematics Knowledge, MC: Mechanics Comprehension, EI: Electronics Information, All of ASVAB Math and Science Scores together. We do not report ASVAB scores coefficients in the last column to conserve space.