I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15483

Do STEM Students Vote?

D’Wayne Bell
Jing Feng
John B. Holbein
Jonathan Smith

AUGUST 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15483

Do STEM Students Vote?

D'Wayne Bell
Harvard University

Jing Feng
College Board

John B. Holbein
University of Virginia

Jonathan Smith
Georgia State University and IZA

AUGUST 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Do STEM Students Vote?*

For decades, pundits, politicians, college administrators, and academics have lamented the dismal rates of civic engagement among students who enroll in courses and eventually major in science, technology, engineering, and mathematics (i.e., STEM) fields. However, the research supporting this conclusion has faced distinct challenges in terms of data quality. Does STEM actually decrease the odds that young people will be actively involved in democracy? This paper assesses the relationship between studying STEM and voting. To do so, we create a dataset of over 23 million students in the U.S. matched to national validated voting records. The novel dataset is the largest known individual-level dataset in the U.S. connecting high school and college students to voting outcomes. It also contains a rich set of demographic and academic variables, to account for many of the common issues related to students' selection into STEM coursework. We consider two measures of STEM participation—Advanced Placement (AP) Exam taking in high school and college major. Using both measures, we find that, unconditionally, STEM students are slightly more likely to vote than their non-STEM peers. After including the rich set of controls, the sign reverses and STEM students are slightly less likely to vote than their non-STEM peers. However, these estimated relationships between STEM and voting are small in magnitude—about the same effect size as a single get-out-the-vote mailer—and we can rule out even very modest causal effects of marginally more STEM coursework on voting for the typical STEM student. We cannot rule out modest effects for a few subfields. Our analyses demonstrate that, on average, marginally more STEM coursework in high school and college does not contribute to the dismally low participation rates among young people in the U.S.

JEL Classification: I21, I23, D72

Keywords: youth voting, education, college, STEM, large-scale administrative data

Corresponding author:
Jonathan Smith
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992
USA

E-mail: jsmith500@gsu.edu

* This research represents the views of the authors and not of any organization.

Sparked by the concern that the United States is falling behind other countries in core science and mathematics competencies and the implications for the American economy, policymakers have placed heavy weight on increasing high school and college students' exposure to science, technology, engineering, and mathematics (i.e. STEM) coursework (Mayo 2009; Leshner 2018; Manduca et al. 2017; Foster et al. 2010; Chirikov et al. 2020; Ferrini-Mundy 2013). In the last ten Congresses, for example, over 2,000 pieces of individual legislation relating to STEM education have been introduced, with even more policies being proposed and enacted at the state and local level (*STEM & Vital Signs - Education Commission of the States* N.d.). Indeed, there has been a corresponding increase in STEM coursework. Over the past two decades, the number of students majoring in STEM related fields has increased by more than 50 percent (*More Students Are Earning STEM Degrees* N.d.; *STEM Education Data and Trends* N.d.*a*,N).

While there are likely many direct and indirect benefits to individuals and society from studying STEM (Beede et al. 2011; Xie, Fang and Shauman 2015; Fry, Kennedy and Funk 2021; Dahl, Rooth and Stenberg N.d., e.g.,), some critics are concerned that STEM coursework comes at the expense of developing other core competencies (Teitelbaum 2014; Ravitch and Viteritti 2001; Board 2015; Freeman 2006; Xie and Killewald 2012; Kelly et al. 2004; Lowell and Salzman 2007; Casselman 2014). At the heart of this debate are concerns about whether a STEM-focused pedagogy allows room for what many reformers argue is *the* foundational mission of the American education system—the development of young peoples' capacity to engage with and preserve democracy (Plato N.d.; Dewey 1923, 1903; Holbein and Hillygus 2020; Campbell 2006; Ravitch and Viteritti 2001; Chomsky 2003; Lipset 1959; Converse 1972; De Tocqueville 2003). After all, young people are engaged in civic and democratic processes at shockingly low rates and, by some accounts, declining levels (Holbein and Hillygus 2020). And, according to some studies, STEM majors are voting at some of the lowest rates of all young people (Hillygus 2005; Ro and Bergom N.d.; Niemi and Hanmer 2010) and are much less likely to contribute their time, money, and energy to political and social causes (Ravitch and Viteritti 2001; Allgood et al. 2012; Sax 2004; Astin et al. 1997). This has led many critics ranging from policymakers and educators to thought-leaders and scholars to argue that STEM education needs a "civic engagement makeover" that incorporate civic and social issues (Garlick and Levine 2017; Teitelbaum 2014; Ro and Bergom N.d.; Hillygus 2005; Ravitch and Viteritti 2001; Feuer 2021; Condon and Wichowsky 2018; Higgins, Wallace and Bazzul 2018).

Critical to debate on STEM coursework and civic engagement are two questions that we address in this paper—whether STEM students are in fact less civically engaged than non-STEM students and what role does STEM coursework contribute to that relationship? There are at least two reasons these are difficult questions to answer. First, data that contain the necessary variables and sample are difficult to come by. Most work in the area use survey-based measures of civic engagement (Teitelbaum 2014; Ravitch and Viteritti 2001; Board 2015; Freeman 2006; Xie and Killewald 2012; Kelly et al. 2004; Lowell and Salzman 2007; Casselman 2014), which may be contaminated by social desirability bias (Karp and Brockington 2005; Holbrook and Krosnick 2010). Additionally, samples are often small in size or for a single group, cohort, or point in time (Fraga and Holbein 2020), each of which are valuable but can yield idiosyncratic results. Second, it is difficult to disentangle whether any relationship between STEM coursework and civic engagement is *caused* by the coursework or simply reflects pre-existing levels of interest in civic or political engagement. Many studies include a limited set of control variables to address selection bias (Teitelbaum 2014; Ravitch and Viteritti 2001; Board 2015; Freeman 2006; Xie and Killewald 2012; Kelly et al. 2004; Lowell and Salzman 2007; Casselman 2014), but data and surveys are limited in their ability to allay endogeneity concerns.

In this paper, we begin to answer these questions by creating the largest and most comprehensive dataset in the U.S. that links high school and college students and their studies to validated voting records. Our data consists of over 23 million students who graduated high school between 2004 and 2013 and took either the SAT, PSAT/NMSQT (PSAT), and/or Advanced Placement (AP) exams. The College Board's exam data are linked to the National Student Clearinghouse data, which includes college enrollment spells and college majors. We therefore have two distinct measures of participation in STEM coursework - STEM AP exam taking and college STEM majors. These administrative educational records are then linked to validated voter files. The voting records of (initially) over 200 million Americans, provided by Data Trust, LLC, include whether (but not for whom) individuals vote in the 2004-2016 federal elections. These merged data are not only an order of magnitude (if not more) larger than previous work in this space, which allows us to get precise estimates, but the educational records have a rich set of demographic and academic variables that allow us to account for many sources of selection into STEM coursework and fields.

On average, we find relatively small differences in voting rates among STEM and non-STEM students. Unconditionally, students who take one additional AP STEM exam are 0.35 percentage points *more* likely to vote in their first eligible election. College STEM majors are only 0.9 percentage points *more* likely to vote after graduating college. After we add a rich set of controls, such as high school (or college) attended, parental income and education, and PSAT scores, the estimates become negative, but remain small in magnitude. Specifically, students who take one additional AP STEM exam are 0.25 percentage points *less* likely to vote in their first eligible election and college STEM majors are only 0.6 percentage points *less* likely to vote after graduating college. These results are robust and qualitatively similar across a variety of specifications, different election types, demographic groups, and time periods. (However, they are quite modest in size.)

Additionally, the reversal of the signs on the coefficient estimates suggest that selection into STEM coursework is, in general, a valid concern in the literature that attempts to isolate the causal effect of STEM coursework on voting. In our analyses, we cannot fully rule out all sources of selection, despite our rich set of controls and sibling fixed effects. We suspect, however, that any remaining selection issues bias the estimates away from zero (i.e., more negative relationship between STEM and voting) and our estimates provide an upper bound. Fortunately, our large sample yields very precise estimates and such that the bounds of the 95 percent confidence intervals are not noticeably distinct from modest coefficient estimates. Our results suggest that, contrary to popular wisdom, STEM students, *on average*, do not vote at appreciably lower rates than non-STEM students and STEM coursework, *on average*, does not meaningfully decrease the chances that young people will be civically engaged. To put the magnitude of the effect into context, the differences in voting rates between STEM and non-STEM students is about the size (in absolute terms) of the impact of a single get-out-the-vote mailer (Green, McGrath and Aronow 2013). At their surface, these modest magnitudes are not strong endorsements for civic-based interventions to target STEM-oriented students. But this may not be a universal truth. Some individual estimates on AP exams (e.g., U.S. Government and Politics) and college majors (e.g., History) are much more likely to vote than someone who studies typical STEM fields. Additionally, our estimates are relative to an unobserved counterfactual. STEM majors are likely required to take some social science or history courses at most colleges. Therefore, our estimates are about the difference in coursework between a STEM and non-STEM major, not the difference between taking all STEM and all non-STEM courses.

Overall, we take our results to show that there is reason to be concerned about low and (by some accounts) declining youth turnout, but targeting the typical STEM student or intervening with typical non-STEM coursework or material, may not produce the desired results—at least not in high school or college.

1 Data

This project makes use of administrative educational data from the College Board (CB) and National Student Clearinghouse (NSC), as well as national voting records. Combining them generates one of the largest and richest datasets connecting individual-level education records from high school and college with voting outcomes.

1.1 Educational Records

Our education data come from two main sources: the CB and NSC. CB owns and operates three major exams taken by high school students – the SAT, PSAT/NMSQT (PSAT), and Advanced Placement (AP). The individual-level data includes all students who took at least one of the exams in the high school graduating cohorts of 2004 to 2013, which is approximately 3 million students per cohort and well over 20 million unique exam takers.

The SAT is one of two college entrance exams considered in admissions and program placement by thousands of colleges across the U.S. Approximately 1.5 million students per cohort take the exam. The SAT is scored between 400 and 1600 - 200 to 800 for each the math and verbal sections. Upon registration, students complete a questionnaire that providing with their names and date of birth, demographics, such as gender, race/ethnicity, and parental income and education. It also includes which high school they attend and SAT scores on all attempts. The PSAT is similar to the SAT but more broadly taken, typically prior to the SAT.

Advanced Placement is a program that offers high school course content in over 30 different subjects, ranging from English and history to calculus and chemistry. Students typically take a year-long course in their high school and have the option to take the corresponding subject's AP exam. High schools determine which AP courses are offered, if any, and the courses are offered at different years in a high school career but most are taken in junior and senior years. AP exams are scored as integers between 1 and 5 and those scoring high enough are eligible for college credit. We have the full set of AP exams taken (not courses taken)[1] and each score for the over one million students who take at least one exam per high school cohort.

We also make use of siblings in the College Board data. The sibling identifiers are borrowed from previous work using similar data.[2]

NSC data are a near census of college enrollment spells in the U.S. It includes enrollment dates for each student in each college for approximately 98 percent of all enrollees in the U.S.[3] We focus on the first college enrolled and first college major, the latter of which is only available for students who earn a degree.[4] These majors are then categorized by their two-digit Classification of Instructional Program (CIP) code, which we further be categorize into STEM and non-STEM majors.[5]

---

[1] We do not observe the fraction of students take the AP course but not the AP exam. A recent study of four public school districts find that 85 percent of course-enrollees take the exam (Fazlul, Jones and Smith 2021).

[2] Previous work identifies siblings as having the same last name, street address, and high school attended. This process misses some siblings but has a very low chance of falsely identifying siblings. See (Altmejd et al. 2021) and (Smith 2013) for a further discussion and examples.

[3] The biggest deficiency is for-profit colleges, despite including some of the largest ones.

[4] We only make use of the first major listed in the data but only 6 percent of graduates have more than one major.

[5] CIP codes come from the U.S. Department of Education. Details can be found here: https://nces.ed.gov/pubs2002/cip2000/index.asp.

1.2 Voting Records

In the United States, each state collects and reports its own voting data, but they all publicly detail whether (but not for whom) each individual votes. We obtained nationwide voting data from the Data Trust, LLC—one of the many vendors in this space. The Data Trust combines and standardizes data from each state's election governing body on the tens of millions of people who vote in elections from 2004 to 2016. The key outcome variables in the data are for each biannual national election (i.e., 2004, 2006,..., 2016), did the person vote in the primary election and separately the general election. Therefore, for someone who was 18 year of age by November 2004, we observe their voting history in all elections.[6]

We transform the voting records into a single observation per person, as opposed to a single observation per person-state, based on name and date of birth (DOB). The process is detailed in the Appendix A.

1.3 Analytic Dataset

We match the educational records to the voting records using name and DOB for those who live in the 50 states or D.C.[7] We start by matching unique name and DOB combinations from each dataset and then move to a series of fuzzy matching methods, as described in Appendix A. We also create a series of first name, last name, and first and last name identifiers for potential false matches (and non-matches) that we later use in robustness tests, including common names, partially missing DOB in the voting records, and females who are more likely to change their last name. In a series of validation exercises in Appendix A (section B), we show that the patterns of voting rates in our matched dataset follow those of nationally available statistics by age, race/ethnicity, and over time.

With a fully matched dataset, we define two distinct analytic samples based on two measures

[6]Voters can be removed from a state's records. This varies over time and by state but it does not impact most voters. We perform several analyses with recent elections where voters have not had a chance to be removed.

[7]CB and NSC data had previously been matched to one another and used in dozens of studies.

of STEM - AP subject and college major.[8] Our AP sample includes all 23.5 million CB exam takers, including PSAT and SAT takers who may not have taken any AP exams. We count the number of AP STEM exams and number of AP Non-STEM exams for each student,[9] which average XX and YY over the sample period. Our college major sample includes the over 8 million college graduates to whom we know college major.[10] We categorize majors into STEM and non-STEM, which are mutually exclusive, based on their two-digit CIP code.

For most of our analyses, we focus on whether the student voted in their first general election, which take place every two years, around the time the students complete high school or college. In the AP sample, this is the first eligible election that the student is 18 years old. In the college sample, this is the first election after graduating college. 16.7 percent and 27.9 percent of the AP sample and college graduate sample, respectively, vote in their first election. Additional summary statistics on voting rates, different voting outcomes,[11] and background variables are in the Appendix Table B1.

2 Methodology

To assess the relationship between STEM coursework or majors and voting, we use regression analyses. When focusing on AP courses, we run variants of the following regression:

$$y_{icse} = \alpha_0 + \alpha_1 AP\_STEM\_CT_i + \alpha_2 AP\_NONSTEM\_CT_i + X'_{ics}\delta + \epsilon_{icse} \tag{1}$$

The outcome is whether individual i in cohort c at school s voted in election e. We are primarily interested in the coefficients α_1 and α_2, which shows how one additional AP exam (STEM or non-

[8]Appendix Table B1 shows the summary statistics of the two samples, which include information on sex, race/ethnicity, and self-reported parental income.

[9]AP STEM and non-STEM are listed in Appendix Table B2.

[10]Including non-majors provides another reference group but does not change results.

[11]We also consider alternative voting outcomes, including total number of elections voted (through 2016), whether voted in a primary or general election, whether voted in a presidential or non-presidential election, and counts of these variables.

STEM) relates to the probability of voting, holding the other exam count constant. We then run several regressions, which add different sets of controls X. Specifically, we run six regressions where X includes:

1. No controls.
2. Cohort fixed effects.
3. Demographics and cohort fixed effects.
4. PSAT scores, demographics, and cohort fixed effects.
5. School-by-cohort fixed effects, PSAT scores, and demographics (preferred).
6. Sibling fixed effects, PSAT scores, demographics, and cohort fixed effects.

By-running the series of regressions, we examine how adding increasingly more controls impacts the coefficients. The first equation are just the unconditional differences in voting rates. Adding variables like PSAT scores and demographics is intended to (partially) capture selection into these courses based on underlying academic ability and differential preferences based on demographic differences. Our preferred specification is the fifth one, which adds in high school-by-cohort fixed effects. This means we are comparing students in the same high school cohort to one another and so they all had the same set of AP courses to choose from, unlike the previous set of regressions. Finally, the sixth regression adds sibling fixed effects, which compares the AP exam taking of siblings. This has some advantages relative to the previous regressions, such as accounting for time-invariant differences in the family, but it does not control for the differences in AP courses offered. Additionally, there is the potential for siblings to influence one another's voting habits, which should attenuate the estimates on exam taking towards zero.

We perform a similar analysis in the college graduate sample but our equation is slightly modified as follows:

$$y_{icse} = \beta_0 + \beta_1 STEM\_MAJOR_i + X'_{ics}\gamma + \omega_{icse} \tag{2}$$

The primary difference with the previous equation is *STEM_MAJOR* is a binary variable, such that β_1 describes the difference in the voting outcome between STEM and non-STEM majors. The other difference of note is that we use college or college-by-cohort fixed effects in place of high school. This means we are comparing students in the same college, sometimes the same college and cohort, one of whom is a STEM major and the other is not, also accounting for PSAT score and demographics. Finally, in a robustness test, the X_{ics} includes a control for whether the student voted in their first eligible election after turning 18, so as to account for propensity to vote prior to most college and major coursework.

We stress that these methods cannot entirely rule out other sources of selection into STEM and non-STEM course-work. However, the rich set of controls accounts for many plausible explanations students select into STEM courses (e.g., availability, academic ability, and demographics). Any remaining selection likely biases our estimates away from zero and also can likely be considered an upper bound on the causal effects of the coursework.

3 Results

3.1 AP Courses

Table 1 shows the results for the AP sample. Column 1 shows that unconditionally, each additional AP STEM course is associated with a 0.35 percentage point increase in the probability of voting in the first eligible election, controlling for the number of non-STEM AP courses. That is a 2 percent increase relative to the sample's mean voting rate of 16.7 percent. Alternatively stated, taking three additional AP STEM courses is associated with a one percentage point increase in voting in the first election. The corresponding estimate for an additional AP non-STEM course is 1.88 percentage points (11 percent), much larger in magnitude but also positive,

Next, we show the unconditional relationship between the number of AP courses and voting in Figure 1 but disaggregated by course counts.[12] The bottom left corner of the heat map shows that students who take no AP courses, STEM or non-STEM, are relatively unlikely to vote in their first

[12]Students who take five or more AP courses in STEM or non-STEM are grouped together.

eligible election (15 percent). There are two ways to think about STEM-oriented students. First, moving from left to right, students are taking more AP STEM courses. But for the bottom two rows, the other rows show a decrease in voting rates as the number of AP STEM courses increase. For example, 22 percent of students who take 3 AP non-STEM courses and zero AP STEM courses voted in their first election, compared to 18 percent of students who take 3 AP non-STEM courses and 5 or more AP STEM courses. Second, conditional on the number of STEM courses taken (i.e., a column), the more AP non-STEM courses taken are associated with higher voting rates. For example, among students who take 3 AP STEM courses, those who take no AP non-STEM courses versus 5 AP non-STEM courses vote at a rate of 17 and 23 percent, respectively. This second way of looking at the figure demonstrates that defining a STEM student may not be just about how many AP STEM courses someone takes but also about how few AP non-STEM courses someone takes (and the combination of the two).

Moving across the columns of Table 1, we see that adding in cohort fixed effects (column 2) does not change the estimates but controlling for demographics (column 3), attenuates the estimates. Adding in PSAT scores (column 4), an exam often used to sort students into AP courses, flips the sign of the coefficient on AP STEM courses. Adding in high school-by-cohort fixed effects (column 5) does not appreciably change the estimates, which is an indication that additional control variables to account for selection may have limited impact on the magnitude of the coefficient. Overall, column 5 shows that students in the same high school and cohort, with the same demographics and PSAT scores, who take an additional AP STEM course are slightly less likely to vote in their first eligible election (0.25 percentage points). Conversely, those students who take an additional AP non-STEM course are 1.3 percentage points more likely to vote. Adding siblings fixed effects (column 6) attenuates the estimates, consistent with both smaller relationships between courses and voting and separately sibling spillovers, and there is no longer an estimate on AP STEM courses that is statistically different than zero.

Collectively, Table 1 shows that students who take relatively many AP non-STEM courses are more likely to vote and students who take relatively many AP STEM courses are only slightly less likely to vote, after accounting for underlying differences between students. In the Appendix

C, we describe a series of robustness tests on our primary specification (5), largely focused on potential issues related to the matching process. The results are qualitatively unchanged.

In Appendix B, we show and discuss four additional findings. First, conditional on our covariates, we do not find non-linearities in the number of courses taken, providing support for our linear specification and suggesting that the relationship between a marginal AP course and voting is similar throughout the distribution of course count (Appendix Table B3). Second, we show that the broad conclusions are the same across demographics and subgroups, but to slightly varying degrees of magnitude (Appendix Table B4). Third, we show that the results are similar when considering alternative outcomes (Appendix Table B5). In particular, when considering voting in off-cycle elections or count of elections voted, the results hold but are different absolute magnitudes (similar percents). Fourth and finally, we see relatively persistent results for several election cycles but the relationships in the long-term are less consistent, perhaps because of degradation of the voting records over time (Appendix Table B6).

Next, we explore the relationship between each distinct AP exam and voting in the first eligible election. We estimate coefficients using the preferred fifth regression with controls for PSAT, demographics, and high school-by-cohort fixed effects but instead of using STEM and non-STEM counts, we use a set of binary variables for taking each AP course. Figure 2 shows the coefficient estimates and demonstrates that the earlier results are not driven by a single course.[13] Rather, almost all coefficient estimate of AP STEM courses are below zero but small in magnitude, and almost all coefficient estimates for non-STEM courses are above zero and larger in magnitude. AP Microeconomics and Macroeconomics are exceptions on the non-STEM side. And AP Environmental Science is an exception on the STEM side. And noticeably, the largest relationships are with AP Government and AP U.S. History, both of which have content related to governments, elections, and voting. The coefficient on AP U.S. Government is 2.6 percentage points, which corresponds to a 15.6 percent increase in the probability of voting compared to students who do not take the course.

[13]The corresponding regression results are in Appendix Table C7.

3.2 College Major

Table 2 shows the results for the college graduate sample. Column 1 shows that unconditionally, majoring in STEM is associated with a 0.88 percentage point increase in the probability of voting in the first after college graduation. That is a 3.2 percent increase relative to the sample's mean voting rate of 27.9 percent.

Moving across the columns of Table 2, we see that adding in cohort fixed effects (column 2) and controlling for demographics (column 3) both attenuates the previous estimates. Adding in PSAT scores (column 4), again flips the sign of the coefficient, suggesting a negative relationship between STEM majors and voting. Adding in college-by-cohort fixed effects (column 5) does not appreciably change the estimates. Overall, column 5 shows that students in the same college and cohort, with the same demographics and PSAT scores, who major in a STEM field are slightly less likely to vote in their first election after college graduation (0.64 percentage points, 2.3 percent). Adding siblings fixed effects (column 6) reverses the sign to be positive again but small in magnitude.

Collectively, Table 2 shows that students who major in STEM are slightly less likely to vote than their non-STEM peers, after accounting for underlying differences between students. Similar to the AP course-taking analyses, we present additional findings in the Appendices B and C that show the results are robust to the specification, potential matching issues, and alternative voting outcomes. We also show that when we control for whether the student voted in the first eligible election after high school, results are attenuated but qualitatively similar (Appendix Table C6). This last result suggest that not all the relationship between college major and voting can be explained by previous voting habits.

Next, we explore the relationship between each college major and voting after college graduation. We estimate coefficients using the preferred fifth regression with controls for PSAT, demographics, and college-by-cohort fixed effects but instead of using a single STEM major variable, we use a set of binary variables for taking each college major. Figure 3 shows the coefficient estimates of each major relative to the omitted category - other (uncommon) non-STEM majors.[14] Four of the five STEM majors show negative coefficients, including engineering, which is a very popular major. Computers and information sciences is the only coefficient that has a positive relationship with voting relative to the omitted other majors. Notably, history majors are 7 percentage points (25 percent) more likely to vote than the omitted other majors, even after accounting for the background factors and college enrolled. Some non-STEM majors have coefficients below zero, including business, but the magnitude of the coefficients are smaller than the humanities and social sciences on the rightmost of the figure.

4 Discussion

This paper has documented the relationship between STEM coursework in high school and college with the propensity to vote. Overall, we find two consistent results. First, without adjusting for covariates, STEM students, as measured by AP STEM courses and college major, are slightly more likely to vote than non-STEM students. This is somewhat at odds with the prevailing wisdom and previous research

What should we make of the magnitudes of the relationship between STEM and voting being quite modest (with the exception of a few AP courses and majors)? This finding leads to two points of discussion. First, whether the noted differences are worthy of an intervention, say to increase the voting rates of a typical STEM student, relative to a typical non-STEM student, is not immediately clear. Many state and local governments, agencies, and educational institutions promote and require civics and non-STEM coursework to increase civic engagement. These may indeed be worthy measures, which could even explain such modest differences, but using these measures to further close any STEM versus non-STEM voting gaps may be only have modest impacts.

A second related point is that though we cannot rule out additional sources of student selection, our rich dataset has ruled out several plausible sources of selection. As such, our estimates

[14] The corresponding regression results are in Appendix Table C8. Other majors are the least common majors that are not easily categorized but they are all non-STEM.

are likely to be upper bounds on the causal impacts of AP coursework or college majors on these voting outcomes. And some individual courses and majors, such as social science, have as much as a 4 percentage point (15 percent) higher probability of voting than some STEM majors. This suggests that the upper bound of the causal impact of certain types of non-STEM coursework may in fact be large. Future research or policymakers can consider the magnitudes of our estimates when considering what to expect from any policy or intervention targeted at STEM (or non-STEM) students or coursework.

In total, our work provides evidence that the role of STEM in shaping the next generation of citizen leaders is much more nuanced than previous studies have suggested.

References

Allgood, Sam, William Bosshardt, Wilbert Van der Klaauw and Michael Watts. 2012. "Is economics coursework, or majoring in economics, associated with different civic behaviors?" *The Journal of Economic Education* 43(3):248–268.

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson and Jonathan Smith. 2021. "O brother, where start thou? Sibling spillovers on college and major choice in four countries." *The Quarterly Journal of Economics* 136(3):1831–1886.

Astin, Alexander W et al. 1997. *What matters in college?* Josey Bass.

Beede, David, Tiffany Julian, David Langdon, George McKittrick, Beethika Khan and Mark Doms. 2011. "Women in STEM: A Gender Gap to Innovation. ESA Issue Brief# 04-11." *US Department of Commerce* .

Board, National Science. 2015. "Revisiting the STEM workforce: A companion to science and engineering indicators 2014.".

Campbell, David E. 2006. *Why we vote: How schools and communities shape our civic life*. Vol. 100 Princeton University Press.

Casselman, Ben. 2014. "The Economic Guide to Picking A College Major." *FiveThirtyEight* .

Chirikov, Igor, Tatiana Semenova, Natalia Maloshonok, Eric Bettinger and René F Kizilcec. 2020. "Online education platforms scale college STEM instruction with equivalent learning outcomes at lower cost." *Science Advances* 6(15):eaay5324.

Chomsky, Noam. 2003. *Chomsky on democracy & education*. Psychology Press.

Condon, Meghan and Amber Wichowsky. 2018. "Developing citizen-scientists: Effects of an inquiry-based science curriculum on STEM and civic engagement." *The Elementary School Journal* 119(2):196–222.

Converse, Philip E. 1972. "Change in the American electorate." *The human meaning of social change* pp. 263–337.

Dahl, Gordon B, Dan-Olof Rooth and Anders Stenberg. N.d. "High School Majors and Future Earnings." *American Economic Journal: Applied Economics*. Forthcoming.

De Tocqueville, Alexis. 2003. *Democracy in America, Vo1. 10*. Regnery Publishing.

Dewey, John. 1903. "Democracy in education." *The elementary school teacher* 4(4):193–204.

Dewey, John. 1923. *Democracy and education: An introduction to the philosophy of education*. macmillan.

Fazlul, Ishtiaque, Todd Jones and Jonathan Smith. 2021. "College Credit on the Table? Advanced Placement Course and Exam Taking." *Economics of Education Review* 84:102155.

Ferrini-Mundy, Joan. 2013. "Driven by diversity." *Science* 340(6130):278–278.

Feuer, Michael J. 2021. "Science, civics, and democracy.".

Foster, Kevin M, Kathleen B Bergin, Anna F McKenna, Don L Millard, Lance C Perez, Joan T Prival, Daphne Y Rainey, M Sevian, Hannah, Elizabeth A VanderPutten and James E Hamos. 2010. "Partnerships for STEM education." *Science* 329(5994):906–907.

Fraga, Bernard and John Holbein. 2020. "Measuring youth and college student voter turnout." *Electoral Studies* 65:102086.

Freeman, Richard B. 2006. "Does globalization of the scientific/engineering workforce threaten US economic leadership?" *Innovation policy and the economy* 6:123–157.

Fry, Rick, Brian Kennedy and Cary Funk. 2021. "STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity: Higher Education Pipeline Suggests Long Path Ahead for Increasing Diversity, Especially in Fields Like Computing and Engineering." *Pew Research Center* .

Garlick, JA and P Levine. 2017. "Where civics meets science: Building science for the public good through civic science." *Oral Diseases* 23(6):692–696.

Green, Donald P, Mary C McGrath and Peter M Aronow. 2013. "Field experiments and the study of voter turnout." *Journal of Elections, Public Opinion and Parties* 23(1):27–48.

Higgins, Marc, Maria FG Wallace and Jesse Bazzul. 2018. "Disrupting and displacing methodologies in STEM education: From engineering to tinkering with theory for eco-social justice." *Canadian Journal of Science, Mathematics and Technology Education* 18(3):187–192.

Hillygus, D Sunshine. 2005. "The missing link: Exploring the relationship between higher education and political engagement." *Political behavior* 27(1):25–47.

Holbein, John B and D Sunshine Hillygus. 2020. *Making young voters: Converting civic attitudes into civic action*. Cambridge University Press.

Holbrook, Allyson L and Jon A Krosnick. 2010. "Social desirability bias in voter turnout reports: Tests using the item count technique." *Public Opinion Quarterly* 74(1):37–67.

Karp, Jeffrey A and David Brockington. 2005. "Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries." *The Journal of Politics* 67(3):825–840.

Kelly, Terrence K, William P Butz, Stephen Carroll, David M Adamson and Gabrielle Bloom. 2004. The US scientific and technical workforce: improving data for decisionmaking. Technical report RAND CORP SANTA MONICA CA.

Leshner, Alan I. 2018. "Student-centered, modernized graduate STEM education." *Science* 360(6392):969–970.

Lipset, Seymour Martin. 1959. "Some social requisites of democracy: Economic development and political legitimacy1." *American political science review* 53(1):69–105.

Lowell, B Lindsay and Hal Salzman. 2007. "Into the Eye of the Storm: Assessing the Evidence on Science and Engineering Education, Quality, and Workforce Demand." *Urban Institute (NJ1)* .

Manduca, Cathryn A, Ellen R Iverson, Michael Luxenberg, R Heather Macdonald, David A McConnell, David W Mogk and Barbara J Tewksbury. 2017. "Improving undergraduate STEM education: The efficacy of discipline-based professional development." *Science Advances* 3(2):e1600193.

Mayo, Merrilea J. 2009. "Video games: A route to large-scale STEM education?" *Science* 323(5910):79–82.

More Students Are Earning STEM Degrees. N.d.
URL: *https://www.ppic.org/blog/more-students-are-earning-stem-degrees/*

Niemi, Richard G and Michael J Hanmer. 2010. "Voter turnout among college students: New data and a rethinking of traditional theories." *Social Science Quarterly* 91(2):301–323.

Plato. N.d. *The Laws*. 641b7–10.

Ravitch, Diane and Joseph P Viteritti. 2001. *Making good citizens: Education and civil society*. Yale University Press.

Ro, Hyun Kyoung and Inger Bergom. N.d. "Why don't STEM majors vote as much as others?".
URL: *http://theconversation.com/why-dont-stem-majors-vote-as-much-as-others-89015*

Sax, Linda J. 2004. "Citizenship development and the American college student." *New Directions for Institutional Research* 2004(122):65–80.

Smith, Jonathan. 2013. "Ova and out: Using twins to estimate the educational returns to attending a selective college." *Economics of Education Review* 36:166–180.

STEM Education Data and Trends. N.d.*a*.
URL: *https://www.nsf.gov/nsb/sei/edTool/data/college-02.html*

STEM Education Data and Trends. N.d.*b*.
URL: *https://www.nsf.gov/nsb/sei/edTool/data/college-14.html*

STEM & Vital Signs - Education Commission of the States. N.d.
URL: *https://www.ecs.org/terms-of-use/stem-vital-signs/*

Teitelbaum, Michael. 2014. "The Myth of the Science and Engineering Shortage.".
URL: *https://www.theatlantic.com/education/archive/2014/03/the-myth-of-the-science-and-engineering-shortage/284359/*

Xie, Yu and Alexandra A Killewald. 2012. *Is American science in decline?* Harvard University Press.

Xie, Yu, Michael Fang and Kimberlee Shauman. 2015. "STEM education." *Annual review of sociology* 41:331–357.

5 Data and Materials Availability

The College Board data is proprietary as it contains sensitive individual student records. Likewise, the Data Trust voter file data are also proprietary. We have signed data use agreements that legally prohibit us from sharing either of these. We can, however, share the code that produced the results contained herein.

Figure 1: The Relationship Between AP Exam Taking and Voting in First Eligible Election

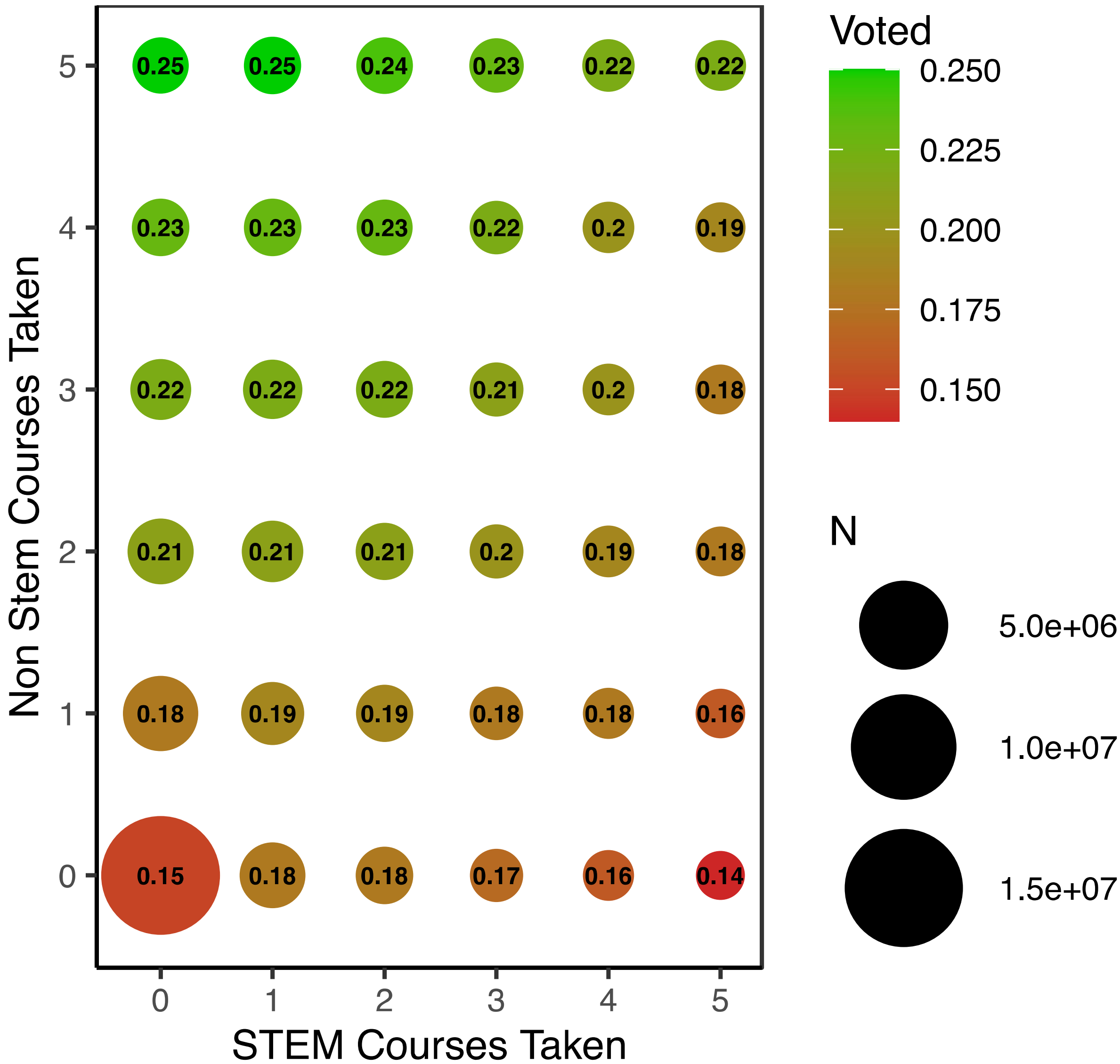

Figure 2: The Relationship Between AP Exam Taking and Voting in first Election, by AP Exam Subject

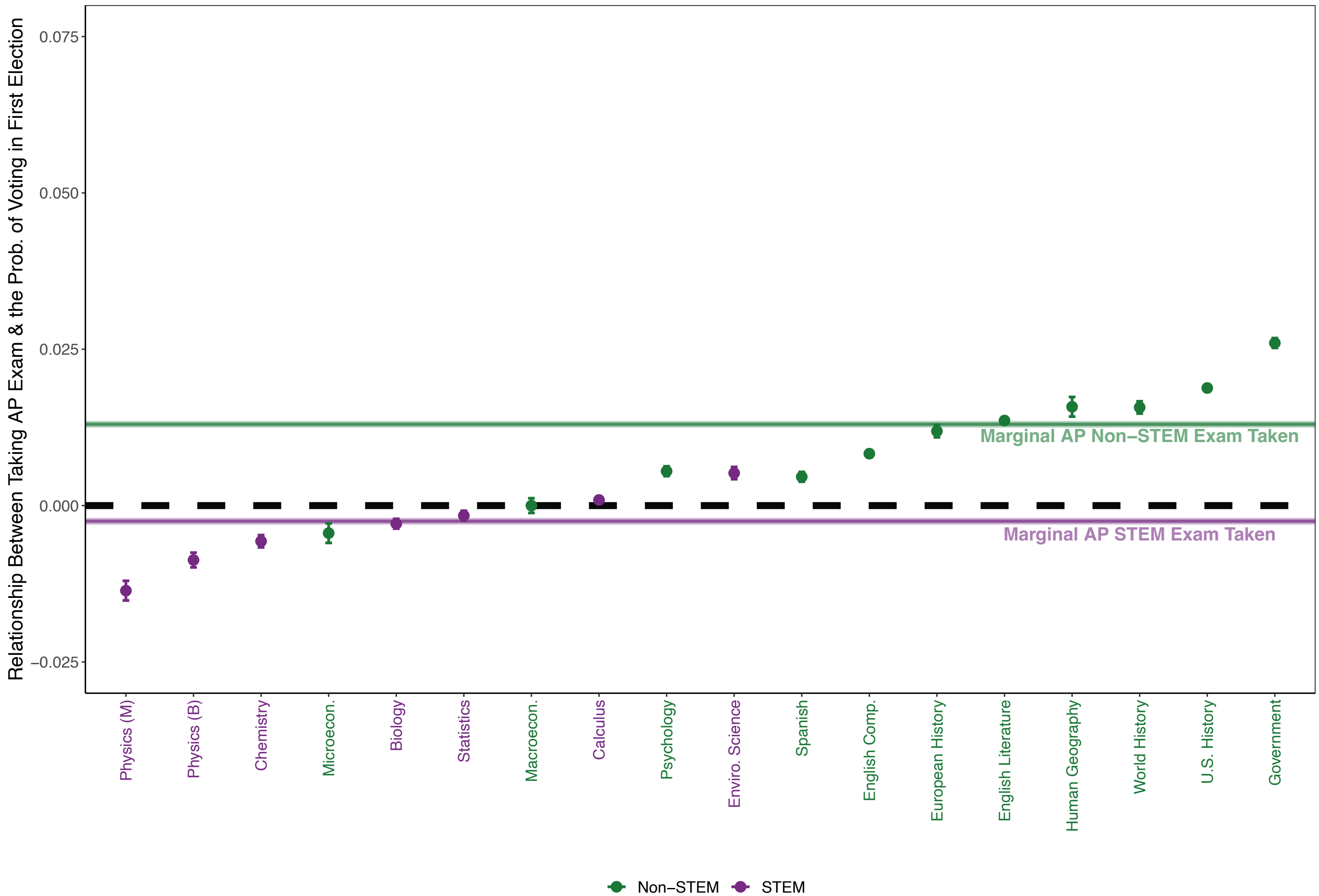

Figure 3: The Relationship Between College Major and Voting, by College Major

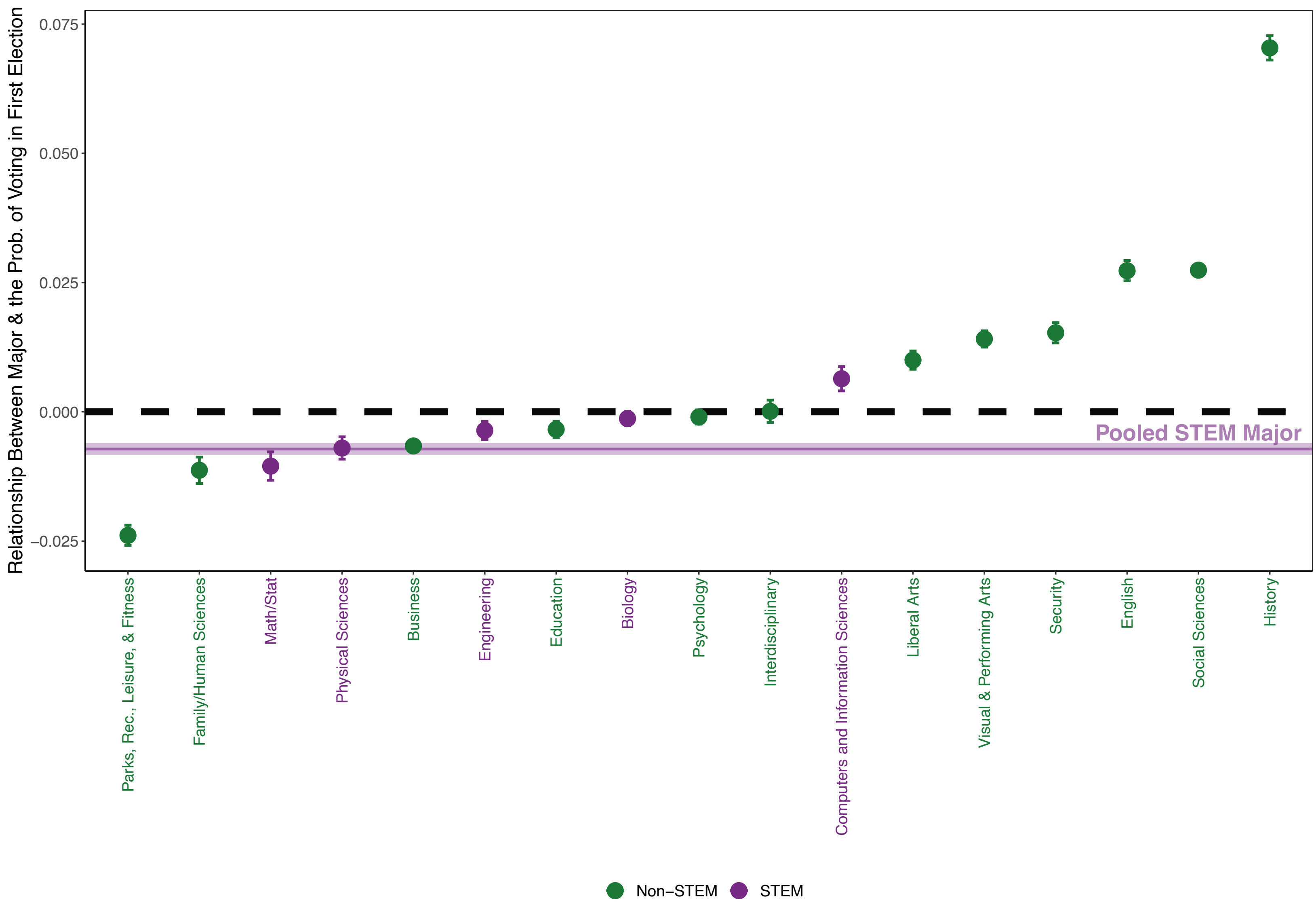

Table 1: High School STEM Taking and Voting

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES			Voted in First Eligible Election			
Took STEM AP	0.0035***	0.0032***	0.0018***	-0.0030***	-0.0025***	-0.0005
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0004)
Took Non-STEM AP	0.0188***	0.0183***	0.0162***	0.0131***	0.0130***	0.0071***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Parent Less than BA			0.0165***	0.0150***	0.0156***	-0.0031***
			(0.0003)	(0.0003)	(0.0003)	(0.0010)
Parent BA+			0.0312***	0.0258***	0.0248***	-0.0075***
			(0.0003)	(0.0003)	(0.0003)	(0.0011)
Parent Education Missing			-0.0290***	-0.0247***	-0.0310***	-0.0217***
			(0.0003)	(0.0003)	(0.0003)	(0.0010)
Parent Income Less than 100,000			0.0178***	0.0165***	0.0164***	-0.0071***
			(0.0003)	(0.0003)	(0.0003)	(0.0009)
Parent Income Greater than 100,000			0.0099***	0.0070***	0.0078***	-0.0088***
			(0.0004)	(0.0004)	(0.0004)	(0.0010)
Parent Income Missing			0.0140***	0.0114***	0.0121***	-0.0025***
			(0.0003)	(0.0003)	(0.0003)	(0.0008)
Female			-0.0083***	-0.0086***	-0.0112***	-0.0189***
			(0.0002)	(0.0002)	(0.0002)	(0.0004)
Asian			-0.0634***	-0.0761***	-0.0767***	-0.0085***
			(0.0005)	(0.0005)	(0.0005)	(0.0026)
Black			0.0535***	0.0500***	0.0419***	-0.0009
			(0.0005)	(0.0005)	(0.0005)	(0.0025)
4.p_ethnic			-0.0252***	-0.0311***	-0.0182***	-0.0011
			(0.0005)	(0.0005)	(0.0005)	(0.0023)
5.p_ethnic			0.0076***	0.0016**	-0.0115***	-0.0041
			(0.0007)	(0.0007)	(0.0007)	(0.0040)
6.p_ethnic			-0.0118***	-0.0174***	-0.0189***	-0.0039*

			(0.0005)	(0.0005)	(0.0005)	(0.0023)
White			0.0281***	0.0123***	0.0085***	-0.0019
			(0.0004)	(0.0004)	(0.0005)	(0.0017)
Other Unknown			-0.0004	-0.0105***	-0.0109***	-0.0051**
			(0.0005)	(0.0005)	(0.0006)	(0.0020)
PSAT Score				0.0029***	0.0030***	0.0022***
				(0.0001)	(0.0001)	(0.0004)
PSAT Squared				-0.0000***	-0.0000***	-0.0000***
				(0.0000)	(0.0000)	(0.0000)
psat_cb				0.0000***	0.0000***	0.0000***
				(0.0000)	(0.0000)	(0.0000)
Took PSAT Junior Year				0.0205***	0.0230***	0.0094***
				(0.0002)	(0.0002)	(0.0007)
PSAT Missing				0.1404***	0.1182***	0.0701***
				(0.0034)	(0.0035)	(0.0115)
Cohort Fixed Effects		X	X	X		X
High School-by-Cohort Fixed Effects					X	
Sibling Fixed Effects						X
Constant	0.1530***	0.1755***	0.1545***	0.0101***	0.0165***	0.1687***
	(0.0001)	(0.0003)	(0.0006)	(0.0035)	(0.0035)	(0.0116)
Observations	23,565,652	23,565,652	23,565,652	23,565,652	23,565,652	6,543,857
R-squared	0.005	0.037	0.049	0.051	0.092	0.553

Notes: The comparison group here are students not taking AP exams. The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP courses are defined as AP Biology, AP Chemistry, AP Calculus AB, Physics 1: Algebra-Based, Physics C: Mechanics, and Statistics. Voting information comes from Data Trust, LLC. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2: College STEM Major and Voting

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Voted in First Election After College Graduation					
STEM Major	0.0088***	0.0049***	0.0009**	-0.0062***	-0.0064***	0.0027*
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0015)
Parent Less than BA			0.0133***	0.0121***	0.0151***	-0.0010
			(0.0006)	(0.0006)	(0.0006)	(0.0038)
Parent BA+			0.0366***	0.0285***	0.0302***	-0.0040
			(0.0006)	(0.0006)	(0.0006)	(0.0039)
Parent Education Missing			-0.0256***	-0.0253***	0.0019***	-0.0120***
			(0.0006)	(0.0006)	(0.0007)	(0.0038)
Parent Income Less than 100,00			0.0168***	0.0151***	0.0155***	0.0005
			(0.0006)	(0.0006)	(0.0006)	(0.0029)
Parent Income Greater than 100,00			0.0200***	0.0154***	0.0131***	-0.0044
			(0.0006)	(0.0006)	(0.0006)	(0.0031)
Parent Income Missing			0.0288***	0.0240***	0.0180***	0.0051*
			(0.0006)	(0.0006)	(0.0006)	(0.0028)
Female			-0.0366***	-0.0364***	-0.0376***	-0.0412***
			(0.0003)	(0.0003)	(0.0003)	(0.0012)
PSAT Score				-0.0043***	-0.0041***	-0.0046***
				(0.0004)	(0.0004)	(0.0016)
PSAT Squared				0.0000***	0.0000***	0.0000***
				(0.0000)	(0.0000)	(0.0000)
psat_cb				-0.0000***	-0.0000***	-0.0000**
				(0.0000)	(0.0000)	(0.0000)
Took PSAT Junior Year				0.0182***	0.0172***	0.0087***
				(0.0005)	(0.0005)	(0.0021)
PSAT Missing				-0.1373***	-0.1564***	-0.1576***

				(0.0122)	(0.0122)	(0.0555)
Asian			-0.0704***	-0.0913***	-0.0946***	0.0008
			(0.0015)	(0.0015)	(0.0015)	(0.0095)
Black			0.0723***	0.0608***	0.0696***	-0.0135
			(0.0015)	(0.0015)	(0.0015)	(0.0127)
4.p_ethnic			0.0061***	-0.0076***	0.0106***	-0.0042
			(0.0016)	(0.0016)	(0.0016)	(0.0109)
5.p_ethnic			0.0124***	-0.0014	-0.0120***	-0.0132
			(0.0022)	(0.0022)	(0.0021)	(0.0174)
6.p_ethnic			0.0106***	-0.0030*	-0.0055***	-0.0178*
			(0.0016)	(0.0016)	(0.0016)	(0.0098)
White			-0.0029**	-0.0214***	-0.0065***	-0.0044
			(0.0014)	(0.0014)	(0.0014)	(0.0068)
Other Unknown			-0.0075***	-0.0240***	-0.0193***	-0.0061
			(0.0016)	(0.0016)	(0.0016)	(0.0077)
Cohort Fixed Effects		X	X	X		X
College-by-Cohort Fixed Effects					X	
Sibling Fixed Effects						X
Constant	0.2770***	0.2241***	0.2220***	0.3524***	0.4076***	0.4401***
	(0.0002)	(0.0005)	(0.0016)	(0.0122)	(0.0123)	(0.0560)
Observations	8,083,710	8,083,710	8,083,710	8,083,710	8,083,710	3,003,746
R-squared	0.000	0.043	0.052	0.053	0.088	0.732

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. Major information comes from the National Student ClearinghouseSTEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix A - Matching Process

In this appendix, we discuss the process of matching the educational records with the voting records and then we assess the match quality to a nationally representative dataset on voting rates.

A Matching Process

A.1 Preparing the Education Data

The College Board data initially consists of 39.6 million observations across cohorts graduating high school between 2004 and 2018. We then restrict the sample to students who attended high school in one of the 50 states or Washington, DC, and those who have complete information on birth date (i.e., month, day, and year). This leaves over 37.6 million observations that serve as the base sample, to which we later focus on the over 23 million observations in the 2004 to 2013 cohorts in our analyses.

A.2 Preparing the Voting Data

We begin with a dataset that keeps voters with a birth year after 1984 to be consistent with the education data. We also keep voters where their year of birth is missing, leaving us with 52,921,628 observations.

With this initial voting dataset, the goal is to get one observation per person. Getting one observation per person presents two challenges relative to the education data. First, the DOB may be partially incomplete and in rare cases, entirely missing.[15] In the education data, we removed those observations from the base sample. We cannot drop those voting records because someone with missing DOB may be a match to the education data. The supposed match in the education data will look as if they never voted if we remove those voting observations. Second, the same

[15]4.5 percent had a missing birth year, 14.6 percent had a missing birth month, and 21 percent had a missing birth day.

person can show up more than once across states if they move (and potentially within the same state if a state's records are inexact and/or missing a piece of the birth date).

To address the above issues, we first go get to one observation per person per state to one observation per person across states. This involves finding exact matches and then fuzzy matches using name and DOB, all while making sure there is no overlap in the same election, which is indicative of this being more than one person. We also create a series of flags, which we use in robustness tests, when there are common names or missing information on the DOB.

A.2.1 Identifying Voters That Are Not Duplicates

We first identify all voter observations across all states that do not need to be deduplicated and set them aside to what will be the final dataset of deduplicated voters. To do so, We start with all observations with a unique first and last name combination across all states. At the end of this step, 34.4 percent of the voter data are considered unique observations.

With the set of voters with non-unique first and last name combinations, we identify all observations that were unique by first name, last name, and birth year. After this step an additional 21.8 percent of the voter data are considered unique observations. We then identify all observations that were unique by first name, last name, birth year and birth month. At the end of this step, an additional 20.4 percent of the voter data are considered unique observations. Then we do the same by adding birth day, producing an additional 10.9 percent of the voter data considered unique. We do the same thing again with the addition of middle initial. After this step an additional 3.8 percent of the voter data are considered unique. Along each step, we remove the "unique" voter but generate a flag that describes how the uniqueness was determined.

At the end of this process, 91.3 percent of the voter data are considered unique observations.

A.2.2 Identifying Duplicate Voters

With the remaining 8.7 percent of voting data, we identify the observations with the same DOB information, the same exact first and last name, but differing voting history (e.g., one observation voted in an election in 2010 but did not vote in 2012 and the observation with the same name and

DOB information had the opposite voting pattern). The vast majority of these observations were collapsed into a single observation and we retained whether we collapsed two or three (Less than 1.1 percent of these observations were unable to be collapsed into a single observation). Almost all were initially two observations that we collapsed into one.

This deduplication process accounted for roughly 7.3 percent of the original voting observations, leaving just 1.5 percent of original voting observations. We take these observations and place them with the uniquely identified and deduplicated observations and treat them as unique.

From here, we created our final dataset to use in the matching to the College Board data. It contains each dataset that from each round of deduplication to identify unique observations and the remaining observations that were not uniquely identified. In all, we created a dataset with 50,975,728 observations of voting history.

After each de-duplication round, in which observations were uniquely identified, we created a flag to identify all the observations that we determined were uniquely identified for that round to use in a series of robustness tests.

A.3 Exact Matching

We start by using exact matching of the educational data to the voting data. An exact match required perfect agreement between the first name, last name, middle initial, and date of birth, none of which could be incomplete in either dataset for this step. We identified 14,175,803 unique matches.[16] This implies that 35.5 percent of the education sample was found in the voting records through exact matching.

A.4 Fuzzy Matching

After removing any observation that was exactly matched from both the education and voting datasets, we implement a fuzzy matching algorithm, in three broad steps.

[16] 83,652 students were matched to multiple voting records (usually two), creating 193,986 observations. In our main analyses, we randomly choose one of these matches but our results are entirely insensitive to alternatives analyses, largely because this impacted such a small fraction of the sample.

First, we use College Board's 23 step fuzzy matching algorithm - a process they use in other applications. The algorithm starts by very slowly loosening the exact matching criteria. Specifically, it starts with exact matches on first name, last name, and DOB but one or both of the datasets are missing a middle initial, everything matches exactly but for one edit to one name, and everything matches but the first and last name are swapped. The most relaxed criteria - the 23rd step - matches exactly on first name, date of birth, and gender but one of the two last names from the two datasets is a suffix of the other and only one of the middle initials is missing. After each step, the matched observations are not replaced for additional matches. We also retain each step in which the observations are matched for robustness tests. This process generates 6,547,716 additional matches, which is an additional 16.4 percent of the education observations.

Second, we used the data linkage method (also known as the editing distance method) employed by Dusetzina, Tyree, Meyer, Meyer, Green, & Carpenter (2014). The method calculates a probability that two strings are a match with the following formula:

$$\sum_{i=1}^{2}[1-(length(name_i)*spedis(name_i,name_{-i}))/2400)]/2 \tag{3}$$

The two names from each dataset, indexed by i, are compared in both character length and also "spelling distance" (i.e., *spedis*). Spelling distance is a common function in statistical software that compares the letters in the name.[17]

We consider any value greater than 0.95 a high enough probability to be a match. We first do this for first names, maintaining an exact match on last name and date of birth, and then again for last name. These generate an additional 27,557 matches (one percent of the education data).

Third and finally, we slightly loosen the criteria on birth date by using exact matches on everything previously described but the voting data has a missing birth day or birth month, but not both and not missing year, so there is no conflicting information. This generates 635,344 additional matches (nearly two percent of the education data).

In total, we matched approximately 27.7 million observations, accounting for almost 53.6 per-

[17]Documentation for *spedis* in SAS, the statistical software we used, can be found here: https://support.sas.com/resources/papers/proceedings/proceedings/sugi25/25/cc/25p086.pdf.

cent of the education sample.

B Assessing Match Quality

Next, we assess the quality of our match. To do so, we compare voting rates in our matched data for different groups and different elections to reported voting rates in the Voting and Registration dataset from the U.S. Census Bureau.[18] These data come from a supplement to the Current Population Survey. We focus on the voting rates among 18-24 year olds in the 2012 and 2016 elections, a time period that our matched data covers well.

In Appendix Tables A1, A2, and A3, we compare voting rates in the two datasets by election, age, race, sex, state, and sometimes combinations of those variables. Generally speaking, we find lower voting rates in our matched data than in the Census data, but the patterns across subgroups follow one another. The lower voting rates is expected, because our matching process is imperfect and somewhat conservative. And it is reassuring that the relative voting rates across subgroups generally match.

Appendix Table A1 shows voting rates in the two datasets by age and sex. For simplicity, we use the high school graduation year as an indicator for when students turn 18. This means that someone who graduates high school in 2012 is assumed to be 18 in the 2012 election and someone who graduates in 2011 is assumed to be 19. The table shows that the voting increases with age (in both elections), according to the Census. Our matched data show a similar pattern, although at lower rates.

Appendix Table A1 also highlights the differences between the two samples by sex. Similar to the Census data, older males vote more than younger males in the matched data. However, this is not true for females, especially in the 2012 election. This is likely because we had difficulty matching to women who change their last name. This also explains why females vote at higher rates than males in both datasets for the youngest cohorts, before women typically get married and change their last name, but not so for older cohorts. For older cohorts, Census data suggest

[18] https://www.census.gov/topics/public-sector/voting.html

females are more likely to vote than males, but the matched data does not. These facts motivate some robustness tests that focus on males and recent cohorts.

Appendix Table A2 shows voting rates in the two datasets by race and sex. The table shows that Black people are the most likely to vote in 2012 in both elections and Asian people are least likely. Similarly, White people are most likely to vote in 2016 in both datasets and Asian people are the least likely. The similar patterns between the datasets across races and elections is comforting.

Finally, voting rates in the two datasets are compared by state. It is immediately clear that our matched sample reflects the voting rates of the Census in some states better than others. This is partially because College Board has relatively low coverage in some states, like Mississippi. However, in states where College Board has substantial coverage, such as Virginia, the two voting rates are well aligned. This motivates a few additional robustness tests, including only using states where College Board has substantial coverage and only using states where the voting rates between these two datasets are well aligned.

Appendix Table A1. Comparing voting rates in 2012 and 2016 elections between Census data and Matched data, by age and by sex.

		Voted in 2012 (Percent)			Voted in 2016 (Percent)		
	Census Age (in years)	*High School Cohort*	*Census*	*Matched Data*	*High School Cohort*	*Census*	*Matched Data*
Overall	18	2012	31.3	25.2	2016	32.3	30.0
	19	2011	33.9	25.2	2015	36.9	30.9
	20	2010	39.7	24.7	2014	40.1	31.0
	21	2009	37.5	24.8	2013	38.8	31.5
	22	2008	41.0	26.9	2012	41.4	33.0
	23	2007	40.0	26.7	2011	41.6	33.1
	24	2006	41.5	26.6	2010	43.8	33.5
Male	18	2012	27.8	23.2	2016	30.5	27.1
	19	2011	30.4	22.8	2015	34.4	27.2
	20	2010	34.4	22.7	2014	37.4	27.3
	21	2009	34.3	23.3	2013	35.8	27.9
	22	2008	38.3	26.3	2012	36.3	29.7
	23	2007	39.3	27.1	2011	38.5	30.4
	24	2006	37.1	28.2	2010	41.9	31.6
Female	18	2012	34.8	27.2	2016	34.3	32.8
	19	2011	37.5	27.5	2015	39.5	34.4
	20	2010	44.7	26.5	2014	42.7	34.5
	21	2009	40.9	26.2	2013	42.0	35.0
	22	2008	43.7	27.5	2012	46.1	36.3
	23	2007	40.8	26.4	2011	44.8	35.8
	24	2006	46.0	25.1	2010	45.8	35.2

Notes: The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

Appendix Table A2. Comparing voting rates of 18-24 year olds in 2012 and 2016 elections between Census data and Matched data, by race and sex

	Voted in 2012 (Percent)		**Voted in 2016 (Percent)**	
	Census	*Matched Data (High School Cohorts 2006-2012)*	*Census*	*Matched Data (High School Cohorts 2010-2016)*
Total	38.0	25.7	39.4	31.8
Male	34.7	24.7	36.5	28.7
Female	41.3	26.7	42.4	34.9
Asian	20.1	We	25.2	25.9
Black	45.9	33.5	40.2	29.1
Hispanic	26.7	20.1	27.2	26.8
White	37.8	26.8	41.1	36.4
Male + Asian	18.1	15.3	21.2	21.8
Male + Black	41.4	26.5	36.5	22.1
Male + Hispanic	24.0	17.7	21.8	22.5
Male + White	34.6	28.0	38.3	34.9
Female + Asian	22.1	19.1	29.5	30.0
Female + Black	50.2	39.6	43.7	35.4
Female + Hispanic	29.7	22.3	32.7	30.7
Female + White	41.0	25.7	43.9	37.8

Notes: The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

Appendix Table A3. Comparing voting rates of 18-24 year olds in 2012 and 2016 elections between Census data and Matched data, by state.

	Voted in 2012 (Percent)		**Voted in 2016 (Percent)**	
State	*Census*	*Matched Data*	*Census*	*Matched Data*
Mississippi	62.4	17.3	46.1	18.1
Minnesota	57.0	21.4	49.6	22.5
Wisconsin	53.3	24.7	45.6	21.2
Colorado	52.5	23	43.1	25.8
South Carolina	51.3	28.4	42.7	33.9
New Hampshire	50.0	17.8	*	21.1
Oregon	47.6	25.7	45.2	26.4
Iowa	46.6	41.1	35.5	52.9
Massachusetts	45.6	33.5	39.9	42.8
Rhode Island	45.2	27.2	*	35.9
North Carolina	45.1	16.8	44.8	18.5
Ohio	44.7	38.1	39.6	43.1
Michigan	43.5	17.1	36.1	22.3
Maine	42.8	13.7	48.8	14.3
Missouri	42.5	33.2	45.9	43.5
Maryland	42.1	33.3	48.0	40.1
Virginia	42.0	37.1	54.6	42.9
Montana	40.7	33.3	*	40.2
Delaware	40.5	21	*	22.0
Louisiana	40.4	11.6	49.2	14.9
Arizona	40.0	12.3	35.8	17.8
Pennsylvania	39.9	27.1	48.7	40.5
Nevada	38.2	27.3	37.6	32.3
Washington	38.2	33.7	42.2	38.7
Georgia	37.9	17.1	40.5	15.6
Kentucky	37.1	28.7	51.1	40.3
New Mexico	37.1	11.3	37.8	12.3
Connecticut	36.9	27.9	37.0	37.7
Florida	36.8	32	33.1	36.9
California	36.5	23	37.5	33.9
Nebraska	36.2	30.7	50.1	45.2
Indiana	35.7	24.8	40.8	31.9
Alabama	35.3	34	41.4	40.0
New York	35.0	27.1	34.6	33.2
New Jersey	34.9	32.4	35.9	38.4
Utah	34.8	28.2	41.6	38.3
Tennessee	34.0	26.9	29.9	32.7
South Dakota	32.7	12.3	*	16.6
Illinois	32.2	29.1	45.3	41.6
Kansas	30.1	26.5	33.8	36.1
Idaho	29.8	9.2	40.7	13.3
Oklahoma	27.2	21.9	32.4	34.2
Arkansas	24.3	26	33.1	30.9
West Virginia	22.6	28.1	32.2	38.0
Texas	22.5	19.1	27.3	24.1
Hawaii	22.1	8	20.4	10.1
District of Columbia	*	19.9	*	20.9
North Dakota	*	15	*	19.2
Alaska	*	12.4	*	15.1
Vermont	*	11.3	*	14.9
Wyoming	*	9.2	*	16.5

Notes: The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age. * indicates that Census data are not available.

Appendix B - Additional Results

Appendix B contains several tables of results we briefly mention in the main text. It begins with the sample summary statistics at the student level and at the AP exam and college major level. It follows with non-linearities in the number of AP exams taken, heterogeneous effects by student characteristics, alternative outcomes, and persistence of effects.

A Tables

Appendix Table B1

Appendix Table B2

Appendix Table B3

Appendix Table B4

Appendix Table B5

Appendix Table B6

Table B1: Summary Statistics

	AP Sample	College Graduate Sample
Gender		
Male	0.47	0.42
Female	0.53	0.58
Race		
White	0.57	0.70
Black	0.14	0.08
Asian	0.07	0.08
Hispanic	0.16	0.10
Other	0.04	0.04
Race Missing	0.03	0.01
Parental Education		
Parent No College	0.11	0.10
College No BA	0.13	0.14
College BA Plus	0.29	0.43
College Missing	0.47	0.32
Parental Income		
Income Less Than $50K	0.13	0.13
Income $50K - $100K	0.13	0.18
Income Greater than $100K	0.09	0.14
Income Missing	0.65	0.55
AP Exams		
AP STEM Count	0.29	--
AP non-STEM Count	0.68	--
Voting Outcomes		
Mean Vote in First Election	0.167	0.279
Total Number of Votes	0.91	0.761
N	23,565,652	8,083,710

	Mean Vote in First Election	Total Number of Votes
AP Sample		
Took Only STEM AP	0.176	0.989
Took Only Non-STEM AP	0.196	1.032
Took Both STEM AP	0.213	1.126
Took Non-STEM AP	0.204	1.075
Took STEM AP	0.203	1.091
Took No AP	0.149	0.827
College Graduate Sample		
STEM Majors	0.286	0.774
Non-STEM Majors	0.277	0.758

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC.

Appendix Table B2: AP Exam and Major Summary Statistics

AP Exams

AP Non-STEM Exams	Test Takers	Voting Rate
U.S. History	3,150,633	0.219
World History	928,715	0.226
U.S. Government	1,706,188	0.235
Psychology	1,303,444	0.213
English Language	2,814,409	0.213
Macroeconomics	638,433	0.217
Microeconomics	383,044	0.213
English Literature	3,000,803	0.217
European History	859,517	0.227
Human Geography	289,313	0.242
Spanish Language	949,971	0.178
AP STEM Exams		
Biology	1,401,147	0.201
Chemistry	908,777	0.202
Calculus AB	2,062,830	0.203
Environmental Science	628,454	0.214
Physics B	550,878	0.203
Physics C: Mechanics	272,337	0.202
Statistics	1,049,516	0.205

College Majors

Non-STEM Majors	Count of Majors	Voting Rate
Education	389,310	0.243
Family and Human Development	115,788	0.242
English	228,196	0.303
Liberal Arts	486,675	0.281
Interdisciplinary	176,550	0.266
Psychology	472,303	0.279
Security and Protective Services	224,266	0.299
Social Sciences	738,997	0.315
Visual and Performing Arts	419,627	0.289
Business	1,260,125	0.272
History	152,531	0.349
Kinesiology	186,447	0.253
Other Major	1,932,515	0.263
STEM Majors		
Computer Science	166,924	0.306
Engineering	355,149	0.291
Biology	499,116	0.279
Mathematics	96,921	0.279
Physical Sciences	182,270	0.284

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. Major information comes from the National Student Clearinghouse. Other Majors consist of all majors that were non-STEM and smaller totals. Voting information comes from Data Trust, LLC.

Appendix Table B3: Non-Linear AP

	Voted in First Eligible Election	
AP STEM Count = 1	0.0038***	0.0013*
	(0.0003)	(0.0007)
AP STEM Count = 2	-0.0025***	0.0006
	(0.0004)	(0.0010)
AP STEM Count = 3	-0.0117***	-0.0011
	(0.0007)	(0.0015)
AP STEM Count = 4	-0.0221***	-0.0070***
	(0.0012)	(0.0026)
AP STEM Count = 5 or More	-0.0350***	-0.0198***
	(0.0023)	(0.0053)
AP Non-STEM Count = 1	0.0222***	0.0106***
	(0.0003)	(0.0007)
AP Non-STEM Count = 2	0.0346***	0.0161***
	(0.0003)	(0.0008)
AP Non-STEM Count = 3	0.0434***	0.0214***
	(0.0004)	(0.0010)
AP Non-STEM Count = 4	0.0518***	0.0271***
	(0.0006)	(0.0013)
AP Non-STEM Count = 5 or More	0.0636***	0.0380***
	(0.0006)	(0.0014)
High School-by-Cohort Fixed Effects	X	
Sibling Fixed Effects		X
Observations	23,565,652	6,543,857
R-squared	0.092	0.553

Notes: The comparsion group here are students taking 0 AP exams. The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table B4: Heterogeneous Effects

AP Sample: Outcome = Voted in First Eligible Election

Subgroup	Male	Female	White	Black	Asian	Hispanic	Other Race	Race Missing	Parental Income < $50K	Parental Income $50K - $100K	Parental Income > $100K	Parental Income Missing	Parental Education - No College	Parental Education - College, no BA	Parental Education - BA or Higher	Parental Education - Missing	PSAT Below Median	PSAT Above Median
AP STEM Exam	-0.0036***	-0.0019***	-0.0001	-0.0002	-0.0087***	-0.0014***	-0.0071***	-0.0030	-0.0068***	-0.0065***	-0.0030***	-0.0005**	-0.0054***	-0.0054***	-0.0051***	0.0050***	-0.0032***	0.0011*
	(0.0002)	(0.0002)	(0.0002)	(0.0006)	(0.0003)	(0.0004)	(0.0007)	(0.0021)	(0.0004)	(0.0004)	(0.0004)	(0.0002)	(0.0005)	(0.0004)	(0.0002)	(0.0003)	(0.0002)	(0.0007)
AP Non-STEM Exam	0.0141***	0.0122***	0.0146***	0.0084***	0.0109***	0.0104***	0.0118***	0.0232***	0.0087***	0.0107***	0.0107***	0.0149***	0.0100***	0.0108***	0.0112***	0.0168***	0.0132***	0.0142***
	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0004)	(0.0011)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0003)
Subgroup Mean	0.169	0.164	0.182	0.193	0.102	0.127	0.149	0.127	0.159	0.195	0.198	0.158	0.156	0.183	0.208	0.139	0.138	0.191
Observations	11,162,809	12,402,843	13,374,303	3,270,928	1,600,859	3,705,363	971,570	642,629	3,044,728	3,056,543	2,107,909	15,356,472	2,650,274	3,047,652	6,763,446	11,104,280	12,075,567	7,062,125
R-squared	0.102	0.108	0.093	0.148	0.142	0.116	0.216	0.247	0.134	0.135	0.142	0.096	0.132	0.136	0.112	0.093	0.101	0.110

College Major Sample: Outcome = Voted in First Election After College Graduation

Subgroup	Male	Female	White	Black	Asian	Hispanic	Other Race	Race Missing	Parental Income < $50K	Parental Income $50K - $100K	Parental Income > $100K	Parental Income Missing	Parental Education - No College	Parental Education - College, no BA	Parental Education - BA or Higher	Parental Education - Missing	PSAT Below Median	PSAT Above Median
STEM Major	-0.0097***	-0.0058***	-0.0058***	-0.0039**	-0.0109***	-0.0133***	-0.0215***	0.0039	-0.0131***	-0.0074***	-0.0031***	-0.0056***	-0.0141***	-0.0054***	-0.0067***	-0.0049***	-0.0069***	-0.0052***
	(0.0006)	(0.0006)	(0.0005)	(0.0018)	(0.0011)	(0.0015)	(0.0023)	(0.0045)	(0.0012)	(0.0010)	(0.0011)	(0.0006)	(0.0014)	(0.0012)	(0.0006)	(0.0008)	(0.0005)	(0.0016)
Subgroup Mean	0.298	0.264	0.277	0.345	0.215	0.289	0.270	0.282	0.254	0.276	0.296	0.280	0.262	0.277	0.305	0.248	0.287	0.272
Observations	3,370,887	4,712,823	5,622,451	629,310	667,586	774,385	286,089	103,889	1,027,571	1,424,667	1,146,834	4,484,638	821,985	1,132,913	3,514,215	2,614,597	5,710,226	1,115,285
R-squared	0.076	0.111	0.093	0.106	0.107	0.104	0.131	0.209	0.089	0.090	0.113	0.092	0.103	0.096	0.100	0.077	0.096	0.092

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared. The top panel uses high school-by-cohort fixed effects and the bottom uses college-by-cohort fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table B5: Alternative Outcomes

	Panel A				Panel B					
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)
	Voted in First Eligible Presidential Election	Voted in First Eligible Non-Presidential Election	Voted in First Eligible Presidential Election	Voted in First Eligible Non-Presidential Election	Count of Elections Voted					
AP STEM Course Count	-0.0012***	-0.0048***	0.0009**	-0.0033***	0.0332***	0.0387***	0.0246***	-0.0080***	-0.0085***	-0.0020*
	(0.0002)	(0.0001)	(0.0004)	(0.0003)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0012)
AP Non-STEM Course Count	0.0179***	0.0079***	0.0092***	0.0038***	0.0752***	0.0905***	0.0782***	0.0586***	0.0603***	0.0340***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0007)
Demographics	X	X	X	X			X	X	X	X
PSAT Scores	X	X	X	X				X	X	X
Cohort Fixed Effects			X	X		X	X	X		X
High School-by-Cohort Fixed Effects	X	X							X	
Sibling Fixed Effects			X	X						X
Observations	23,565,652	23,565,652	6,543,857	6,543,857	23,565,652	23,565,652	23,565,652	23,565,652	23,565,652	6,543,857
R-squared	0.082	0.042	0.563	0.548	0.007	0.051	0.072	0.077	0.125	0.604

	Panel C				Panel D					
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)
	Voted in First Presidential Election After College Graduation	Voted in First Non-Presidential Election After College Graduation	Voted in First Presidential Election After College Graduation	Voted in First Non-Presidential Election After College Graduation	Count of Election Voted After College Graduation					
STEM Major	-0.0047***	-0.0099***	0.0075***	-0.0069***	0.0164***	0.0313***	-0.0022**	-0.0365***	-0.0258***	-0.0094***
	(0.0005)	(0.0004)	(0.0013)	(0.0012)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0033)
Demographics	X	X	X	X			X	X	X	X
PSAT Scores	X	X	X	X				X	X	X
Cohort Fixed Effects			X	X		X	X	X		X
College-by-Cohort Fixed Effects	X	X							X	
Sibling Fixed Effects			X	X						X
Observations	8,083,710	5,978,240	3,003,746	2,261,972	8,083,710	8,083,710	8,083,710	8,083,710	8,083,710	3,003,746
R-squared	0.066	0.027	0.730	0.781	0.000	0.058	0.075	0.079	0.110	0.736

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental+A17 Science, Physics B, Physics C: Mechanics, and Statistics. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared. Panel A uses high school-by-cohort fixed effects. Panel C uses college-by-cohort fixed effects. Panel B and D use sibling fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table B6: Voting Rates Over Time, 2004 and 2011 Cohorts

	(1) Voted in 2004 General Election	(2) Voted in 2006 General Election	(3) Voted in 2008 General Election	(4) Voted in 2010 General Election	(5) Voted in 2012 General Election	(6) Voted in 2014 General Election	(7) Voted in 2016 General Election
				Panel A: 2004 Cohort			
AP STEM Exam	-0.0028***	-0.0058***	-0.0022***	-0.0020***	0.0010	-0.0009	0.0092***
	(0.0006)	(0.0004)	(0.0007)	(0.0005)	(0.0007)	(0.0005)	(0.0007)
AP Non-STEM Exam	0.0101***	0.0062***	0.0128***	0.0100***	0.0162***	0.0131***	0.0221***
	(0.0003)	(0.0002)	(0.0004)	(0.0003)	(0.0004)	(0.0003)	(0.0004)
Observations	2,003,470	2,003,470	2,003,470	2,003,470	2,003,470	2,003,470	2,003,470
R-squared	0.054	0.036	0.062	0.042	0.063	0.050	0.081
				Panel B: 2011 Cohort			
AP STEM Exam	0.0001	0.0001	0.0002	-0.0008***	-0.0033***	-0.0058***	0.0044***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0005)	(0.0003)	(0.0005)
AP Non-STEM Exam	0.0004***	0.0003***	0.0004***	0.0026***	0.0191***	0.0083***	0.0243***
	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0003)
Observations	2,567,701	2,567,701	2,567,701	2,567,701	2,567,701	2,567,701	2,567,701
R-squared	0.014	0.014	0.015	0.039	0.087	0.044	0.108

Notes: The comparsion group here are students not taking AP exams. The data are based on the universe of College Board test-takers in the 2004 and 2011 graduating high school cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Voting information comes from Data Trust, LLC. Robust standard errors in parentheses. Regressions control for demographics, PSAT Scores, and PSAT Scores squared. Panel A uses high school-by-cohort fixed effects. Panel B uses college-by-cohort fixed effects. *Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1* p<0.01, ** p<0.05, * p<0.1

Appendix C - Robustness Tests

In Appendix C, we perform several different analyses to test the robustness of the results from our main specifications. The results from the different analyses remain similar, overall, and provide support for the findings from our main tables that individuals taking AP STEM courses and majoring in a STEM field were less likely to participate in voting and those taking AP non-STEM courses and majoring in a non-STEM field are more likely to vote.

First, we re-estimate our main results on the set of states to which our voting rates are similar to the voting rate in the external U.S. Census data. By doing so, we are removing states where the matching process may have been relatively difficult. Results are in Appendix Table C1. We show the results both for the set of states where the voting rate for our sample differed by less than 5 percentage points compared to data from the U.S. Census, and similarly, 10 percentage points. The results are mostly consistent with previous results. Noticeably, some coefficients on AP STEM exams are not statistically distinguishable from zero.

Appendix Tables C2 and C3 shows the sensitivity of our results to the matching process. While matching, we produced a series of flags indicating potentially problematic names and we also maintained information on how the match was generated (e.g., exact, fuzzy, using missing DOB). The birth date flag identifies observations that had similar name and/ birth date information as at least one other observation in the data, but at least one of those observations was missing an element of the birth date variables. For instance, a year flag indicates that the observation had the same first and last name of at least one other observation in the data, but at least one of those observations had missing birth year information. A month flag indicates that the observation had the same first name, last name, and birth year of at least one other observation but at least one of those observations had missing birth month information. A day flag indicates that the observations had the same first name, last name, birth year, and birth month of at least one other observation, but at least one of those observations had missing birth day information. All of these flags were created before the process of identifying the uniqueness of an observation described in Appendix A. The five unique rounds described in Appendix A indicate when in the identifying unique observation process an observation was determined to be unique. Round one means the observation was determined to be unique by their first and last name. Round two indicates that the observation was determined to be unique by their first name, last name, and birth year. Round three identifies observations that were deemed unique by their first name, last name, birth year, and birth month. Round four indicates that the observation was determined to be unique by their first name, last name, birth year, birth month, and birth day. Round five means that the observations were identified as unique by their first name, last name, birth year, birth month, birth day, and middle initial.

Using all of the various flag and uniqueness round identifiers does not change the story of our findings from the main tables. Results from Appendix Table C2 shows that even when looking at the entire dataset (beyond the analytical sample), observations that were flagged for birth date data, and observations by their uniqueness identifying round we that that the students taking an AP STEM course being less likely to vote in the first election they are eligible for and students taking an AP non-STEM course being more likely to vote in the that same election compared to students not taking an AP exam. A similar analysis was completed for the college sample (Appendix Table C3), and the results were similar to those from the main results showing that STEM majors were less likely to vote in the first election after graduation college.

To determine whether our results are being driven by observations that were poorly matched our matching process. To do so, we assign all fuzzy matches as having not voted and re-estimate our main models. Appendix Table C4 shows that the results are consistent with the main results. For example, Panel B results looking at the relationship between having a STEM major and likelihood to vote in the first election after graduating college, tell a similar story to the results from Table 2. However, the results from Panel A Column 2 show that those taking an AP STEM course were more likely to vote in the first presidential election compared to those taking no AP courses. Despite this difference, the remaining results in Panel A are consistent with previous results. Panel C contains the results when completing the analysis, but only comparing siblings. These results are mostly consistent with results from Table 1, Table 2, and Appendix Table C3. The lone exception is in Column 1, where the results for those taking an AP STEM course flip to slightly positive and significant. The results from this table tells a very similar story about the relationship between STEM education and voting, providing assurance that our estimates from Tables 1 and 2 are robust to various specifications.

Appendix Table C5 examines the relationship between taking an AP STEM and AP non-STEM course on voting in a student's first eligible election but splits the sample based on issues related to names. First, we show results for increasingly less common names. Less common names reduce the change of a false positive match. Next, we split the sample by sex, since females may have changed their last name in the voting data relative to the education data. Finally, we split the sample by high school graduating cohort, knowing that voting records may be purged as time goes on. This implies that the most recent cohorts (e.g., 2013) are likely to perform the best. All results are consistent with Table 1 and provide confidence that our results are not being driven by issues related to matching on difficult names. We complete similar analyses in Appendix Table C6 for the college major sample. These results are consistent with those of Table 2 and show that our results for both samples hold up when completing the analyses on different subgroups.

Finally, Appendix Table C7 shows the college major results but adds in a control for whether the student voted in their first eligible election after high school. This control is extremely predictive of whether someone votes in the future. It also allows us to account for an additional level of selection. Even after adding the control, the main specification shows a negative relationship between STEM majors and voting, albeit small in magnitude. The relationship turns positive for the sibling fixed effect, which also happened in the main tables.

A Tables

Appendix Table C1

Appendix Table C2

Appendix Table C3

Appendix Table C4

Appendix Table C5

Appendix Table C6

Appendix Table C7

Appendix Table C1: Comparison with Census Data

	Panel A	
	(1)	(2)
	Voted in First Eligible Election Diff of < 5pp	Voted in First Eligible Election Diff of < 10 pp
AP STEM Exam	0.0001	-0.0008
	(0.0009)	(0.0007)
AP Non-STEM Exam	0.0210***	0.0224***
	(0.0005)	(0.0004)
Observations	837,798	1,351,881
R-squared	0.106	0.098

	Panel B	
	(1)	(2)
	Voted in First Election After College Graduation Diff of < 5pp	Voted in First Election After College Graduation Diff of < 10 pp
STEM Major	-0.0081***	-0.0095***
	(0.0030)	(0.0023)
Observations	210,046	367,680
R-squared	0.072	0.064

Notes: This table includes students where the difference in voting rates, for 18-24 year olds, between this paper's data and Census data was less than 5 and 10 percentage points, respectively. The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Regressions control for demographics,

Appendix Table C2: Likelihood of Voting in First Eligible Election by AP Course conditional on Flags and Uniqueness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Voted in First Eligible Election					
AP STEM Exam	-0.0025***	-0.0026***	-0.0056***	-0.0051***	-0.0041***	-0.0084***	-0.0062***	-0.0061***	-0.0066***	-0.0056***
	(0.0001)	(0.0001)	(0.0004)	(0.0005)	(0.0007)	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0016)
AP Non-STEM Exam	0.0130***	0.0129***	0.0112***	0.0101***	0.0086***	0.0102***	0.0114***	0.0099***	0.0101***	0.0073***
	(0.0001)	(0.0001)	(0.0002)	(0.0003)	(0.0004)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0009)
Analytical Dataset	X									
Entire Dataset		X								
Year Flag			X							
Month Flag				X						
Day Flag					X					
Unique Round 1						X				
Unique Round 2							X			
Unique Round 3								X		
Unique Round 4									X	
Unique Round 5										X
Observations	23,565,652	23,809,668	4,045,955	2,332,555	1,131,876	4,766,777	3,104,951	3,108,963	1,881,817	251,284
R-squared	0.092	0.091	0.141	0.166	0.229	0.146	0.161	0.154	0.181	0.435

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared and uses high school-by-cohort fixed effects. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table C3: Likelihood of Voting in First Election After College Graduation by Major conditional on Flags and Uniqueness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Voted in First Election After College Graduation									
STEM Major	-0.0064***	-0.0066***	-0.0095***	-0.0099***	-0.0055***	-0.0148***	-0.0074***	-0.0109***	-0.0138***	-0.0121***
	(0.0004)	(0.0004)	(0.0010)	(0.0014)	(0.0018)	(0.0010)	(0.0012)	(0.0012)	(0.0015)	(0.0033)
Analytical Dataset	X									
Entire Dataset		X								
Year Flag			X							
Month Flag				X						
Day Flag					X					
Unique Round 1						X				
Unique Round 2							X			
Unique Round 3								X		
Unique Round 4									X	
Unique Round 5										X
Observations	8,083,710	8,190,661	1,556,306	881,457	514,369	1,811,747	1,181,427	1,168,415	709,336	126,746
R-squared	0.088	0.087	0.096	0.100	0.109	0.106	0.111	0.102	0.100	0.156

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared and uses college-by-cohort fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C4: Robustness to Fuzzy Matches

	Voted in Eligible First Election		Voted in First Election After College Graduation
AP STEM Exam	-0.0011***	STEM Major	-0.0057***
	(0.0001)		(0.0004)
AP Non-STEM Exam	0.0118***		
	(0.0001)		
Observations	23,565,652	Observations	8,083,710
R-squared	0.090	R-squared	0.080

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Fuzzy matches are matches between the education and voting data that were not exact on first and last name and exact date of birth. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Regressions control for demographics, PSAT Scores, and PSAT Scores squared. AP results use high school-by-cohort fixed effects. STEM Major results use college-by-cohort fixed effects. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table C5: Robustness Test of AP STEM and Non-STEM Course Taking and Voting - Voted in First Eligble Election

VARIABLES	(1) Full Sample	(2) Name Commonness <= 100	(3) Name Commonness <= 50	(4) Name Commonness <= 25	(5) Name Commonness <= 10	(6) Name Commonness <= 5	(7) Name Commonness <= 3	(8) Name Commonness <= 1	(9) Males	(10) Females	(11) Sibling Subsample	(12) Cohort = 2004	(13) Cohort = 2005	(14) Cohort = 2006	(15) Cohort = 2007	(16) Cohort = 2008	(17) Cohort = 2009	(18) Cohort = 2010	(19) Cohort = 2011	(20) Cohort = 2012	(21) Cohort = 2013
AP STEM Exam	-0.0025***	-0.0026***	-0.0026***	-0.0027***	-0.0027***	-0.0028***	-0.0028***	-0.0031***	-0.0036***	-0.0019***	-0.0032***	-0.0028***	-0.0052***	-0.0040***	-0.0016***	0.0009*	-0.0055***	-0.0024***	-0.0013***	-0.0009*	-0.0053***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0006)	(0.0004)	(0.0004)	(0.0006)	(0.0006)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0004)
AP Non-STEM Exam	0.0130***	0.0129***	0.0130***	0.0130***	0.0131***	0.0130***	0.0130***	0.0128***	0.0141***	0.0122***	0.0116***	0.0103***	0.0065***	0.0079***	0.0149***	0.0157***	0.0093***	0.0078***	0.0197***	0.0197***	0.0089***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0002)
Observations	23,565,652	22,374,589	21,448,782	20,251,707	18,314,628	16,628,428	15,269,877	11,835,732	11,162,809	12,402,843	6,543,857	2,003,470	2,075,878	2,139,340	2,277,767	2,396,612	2,458,330	2,479,964	2,567,701	2,573,508	2,593,082
R-squared	0.092	0.092	0.092	0.093	0.094	0.095	0.096	0.100	0.102	0.108	0.113	0.054	0.043	0.036	0.069	0.073	0.047	0.040	0.087	0.091	0.060

Notes: The comparsion group here are students not taking AP exams. The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. STEM AP exams are defined as Biology, Chemistry, Calculus AB, Environmental Science, Physics B, Physics C: Mechanics, and Statistics. Voting information comes from Data Trust, LLC. Cohorts are by high school graduation. Regressions control for demographics, PSAT Scores, PSAT Scores squared, and use high school-by-cohort fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C6: Robustness Test of STEM Major and Voting - First Election After Graduating College

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
VARIABLES	Full Sample	Name Commonness <= 100	Name Commonness <= 50	Name Commonness <= 25	Name Commonness <= 10	Name Commonness <= 5	Name Commonness <= 3	Name Commonness <= 1	Males	Females	Sibling Subsample	Cohort = 2004	Cohort = 2005	Cohort = 2006	Cohort = 2007	Cohort = 2008	Cohort = 2009	Cohort = 2010	Cohort = 2011	Cohort = 2012	Cohort = 2013
STEM Major	-0.0064***	-0.0063***	-0.0063***	-0.0062***	-0.0063***	-0.0065***	-0.0067***	-0.0077***	-0.0097***	-0.0058***	-0.0053***	-0.0148***	-0.0098***	-0.0032**	-0.0115***	-0.0152***	-0.0059***	0.0057***	-0.0022	-0.0049***	-0.0186***
	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0014)	(0.0012)	(0.0013)	(0.0014)	(0.0012)	(0.0011)	(0.0012)	(0.0014)	(0.0016)	(0.0045)
Observations	8,083,710	7,703,530	7,390,924	6,981,369	6,312,639	5,725,042	5,247,245	4,035,799	3,370,887	4,712,823	3,003,746	735,802	895,589	827,383	885,189	953,707	962,631	1,051,158	950,941	679,093	142,217
R-squared	0.088	0.089	0.090	0.090	0.091	0.092	0.092	0.093	0.076	0.111	0.090	0.042	0.039	0.043	0.047	0.048	0.031	0.037	0.073	0.069	0.090

Notes: The comparsion group here are students not taking AP exams. The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Cohorts are by high school graduation. Regressions control for demographics, PSAT Scores, PSAT Scores squared, and use college-by-cohort fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C7: Likelihood of Voting Conditional on Previously Voting

	(1)	(2)	(3)	(4)	(5)	(6)
	Voted in First Election After College Graduation					
STEM Major	0.0088***	0.0093***	0.0054***	0.0002	-0.0025***	0.0044***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0015)
Voted before Graduating College	-	0.4078***	0.4033***	0.4025***	0.3980***	0.3418***
		(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0013)
Demographics			X	X	X	X
PSAT Scores				X	X	X
Cohort Fixed Effects		X	X	X		X
College-by-Cohort Fixed Effects					X	
Sibling Fixed Effects						X
Observations	8,083,710	8,083,710	8,083,710	8,083,710	8,083,710	3,003,746
R-squared	0.088	0.224	0.226	0.227	0.251	0.765

Notes: The data are based on the universe of College Board test-takers in the 2004-2013 cohorts. Demographic information is all self-reported. Major information comes from the National Student Clearinghouse. STEM Majors were defined as majors where the first two digits of the CIP codes were 11, 14, 15, 26, 27, 40, and 41. Voting information comes from Data Trust, LLC. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1