Does more math in high school increase the share of female STEM workers? Evidence from a curriculum reform*

Martin Biewen†, Jakob Schwerter‡

March 12, 2019

Abstract

This paper studies the consequences of a curriculum reform of the last two years of high school in one of the German federal states on the share of male and female students who complete degrees in STEM subjects and who later work in STEM occupations. The reform had two important aspects: (i) it equalized the exposure to math for all students by making advanced math compulsory in the last two years of high school, (ii) it roughly doubled the instruction time and increased the level of instruction in math and natural sciences for some 80 percent of students, more so for women than for men. Our results provide some evidence that the reform had positive effects on the share of men completing STEM degrees and later working in STEM occupations but no such effects for women. The positive effect for men appears to be driven by a positive effect for engineering and computer science which was partly counteracted by a negative effect for math and physics.

JEL-Classification: I23, J16, J24

Keywords: academic degrees, occupational choice, gender differences

---

*We thank various seminar audiences, participants at the EADE 2018, and Eric Bettinger, Alvaro Choi, Nicolas Hübner, Moshe Justman, Benjamin Nagengast, Pia Schober, Perrihan Saygin, Elisabeth Tipton, Ulrich Trautwein and Hermann van de Werfhorst for many helpful comments and discussions. This paper uses data from the Centre for Higher Education Research and Science Studies (DZHW), Hannover.

†University of Tübingen, LEAD Tübingen, IZA Bonn, martin.biewen@uni-tuebingen.de

‡University of Tübingen, LEAD Tübingen, jakob.schwerter@uni-tuebingen.de
1 Introduction

Past and recent developments in technological change strongly suggest that future economic growth is primarily expected in fields related to science, technology, engineering, and mathematics (STEM) (OECD, 2010). One possibility of promoting this growth is to foster female participation in STEM subjects with the goal of increasing the number of female STEM graduates and female STEM workers. In addition to this macro perspective, and due to their relation to the high productivity sectors of the economy, STEM-related jobs are usually well-paid. Attracting more females into STEM subjects may therefore be seen as a way to improve female career prospects, reduce the gender wage gap and earnings inequality between genders over the life-cycle (OECD, 2007).

In this paper, we exploit an exogenous shock in the form of a curriculum change at the high school level in order to investigate whether and to what extent it is possible to draw more females into STEM subjects. The reform intended to improve students’ general preparation for university studies and the labor market in one of the German federal states (Schavan, 1999). However, its strongest ingredient was an increase in math and additional natural science classes during the last two years of high school.

The literature cites several possible reasons for the low share of females in STEM subjects over the stages of the education system: (i) ability (Berlingieri and Zierahn, 2014; Friedman-Sokuler and Justman, 2016), (ii) tastes and preferences (Ardicianono, 2004; Ceci and Williams, 2010), (iii) stereotypes (Cheryan, 2012; Franceschini et al., 2014), (iv) path decisions in school (Broecke, 2010; Justman and Méndez, 2018), (v) dropping out of a STEM study programme (Ehrenberg, 2010; Kokkelenberg and Sinha, 2010), (vi) underrepresentation in faculties (Carrell et al., 2010; Griffith, 2010), and (vii) failure to transform a STEM degree into a STEM occupation in the labor market (Danbold and Huo, 2017; Sassler et al., 2017a). This paper contributes to aspects (iv) and (vii). First, using an exogenous shock in path decisions, we evaluate if a reform at the end of high school changes study decisions. We do this by evaluating the number of resulting STEM degrees of high school graduates before and after the reform using the other federal states as a control group. Due to data limitations, we are not able to determine whether the reform affected tastes, preferences or stereotypes. Furthermore, we observe the transition of graduates into the labor market. We are thus able to test whether the reform led to changes in the eventual share of females entering STEM occupations. We use difference-in-difference methods to investigate the effects of the reform in which we include interactions with gender in order to determine differential effects for women compared to men.
For our analysis, we exploit a relatively underused data source for Germany, the graduate surveys from the German Center for Higher Education and Science Studies (DZHW).\footnote{One of the few studies using these data known to us is Parey and Francesconi (2018).} The DZHW surveys provide representative samples of the population of graduates from German tertiary education institutions. We focus on the first wave of cohorts 2005/2006, 2008/2009 and 2013/2014 who were surveyed one year after their graduation. The surveys include secondary education background information, tertiary education decisions as well as information on the transition into the labor market after graduation.

Our analysis follows the recent literature, e.g., Justman and Méndez (2018), by further distinguishing between the case of all STEM fields combined, and the case of STEM subfields that are especially math-intensive (here labeled PTEM). In particular, the latter include all subjects in technology, engineering, math and physics. Recent research suggests different effects for PTEM as opposed to life sciences (biology and chemistry). Among other things, these groups involve different requirements of mathematics. In addition, we look into even more specific combinations of subjects to identify the primary drivers of our results, such as the already mentioned life sciences group, a group consisting of mathematics and physics and, finally, a group combining engineering and computer sciences. The latter group differs from the previous ones in that German high schools do not offer courses in these subjects.

Our study makes the following contributions. First, to our best knowledge, our paper is one of the first ones to use quasi-experimental variation to study the effects of curricula changes on the inflow into STEM occupations, and one of the few ones to use quasi-experimental information to evaluate the effects of curricula changes on college major choices (for the latter, see Joensen and Nielsen (2016), Jia (2016), De Philippis (2017), and the more detailed literature review below). Second, in contrast to many of the quasi-experimental interventions studied in the literature, the reform we are considering is unique and comprehensive in that it affected all students of the last two years in high school by making advanced math courses compulsory which were chosen by only 20 percent of students before the reform. This led to the situation that a very large proportion of the population were compliers of the reform, apart from the fact that it ‘ leveled the playing field’ between the genders and students of different ability (Domina and Saldana, 2012). This is in contrast to other interventions considered in the literature which often reached considerably smaller fractions of the population or only subgroups of certain ability levels (e.g., high-ability students as in Joensen and Nielsen (2016) and De Philippis (2017), or low-ability students as in Jia (2016)).

A final important aspect of our study is that we take great care in computing correct standard errors for our difference-in-difference estimates. We demonstrate that this may substantially change the interpretation of results. Since Bertrand et al. (2004), it is well
known that standard errors in difference-in-difference frameworks are underestimated because of intra-correlation within clusters. As a common practice, difference-in-difference methods use robust standard errors that are clustered at the broadest level. The resulting test-distribution is the student-t-distribution $t(G - 1)$, where $G$ is the number of clusters (Cameron and Miller, 2015). This correction, however, performs poorly if only a few (in our application only one) cluster are treated, the number of clusters is low (in our application fourteen), and the cluster sizes differ to a large extent (Mackinnon and Webb, 2018). Following Roodman et al. (2018), we, therefore, calculate wild cluster bootstrap, wild subcluster bootstrap, and ordinary wild bootstrap p-values and compare them with the p-values from the $t(G - 1)$ standard errors. This comparison reveals a considerable increase in the p-values compared to the conventional practice, even after including state/year fixed effects and state-specific variables aimed at taking out intra-cluster correlation.

Our final results provide some evidence for a positive male treatment effect of the reform on completed degrees driven by the subfields of engineering and computer science. For women, we do not find such effects. We also find a smaller but significant negative effect on the number of math and physics degrees for both men and women. We obtain similar patterns for the entrance into STEM occupations after graduation, driven mostly by the group of engineering and computer sciences, as well as a small negative effect for math and physics occupations.

The structure of this paper is as follows. Section 2 discusses the related literature. In Section 3, we describe the institutional background and the reform in more detail. Sections 4 and 5 provide details about our data and econometric approach. In section 6, we present and discuss our empirical results. Section 7 concludes.

2 Related literature

In the following literature review, we focus on two types of contributions. First, we present studies that already evaluated the reform in order to be able to compare our results to theirs. Second, we connect our analysis to a broader literature that considers the effects of school curricula on educational and economic outcomes. We do not review articles that specifically deal with STEM occupations in the labor market (see, e.g., Spitz-Oener and Priesack, 2018) or the more general topic of women in STEM (Kahn and Ginther, 2018).

Using a different data set, Hübner et al. (2018) considered the same reform in Baden-Württemberg with a particular focus on high school effects and university entry decisions. In particular, they analyzed gender differences in math achievement, math self-concept, and interests in realistic and investigative areas. They further considered which field of
study the individuals chose two years after the completion of high school. Using only data for the reformed state Baden Württemberg and a before-after comparison, the authors find that gender differences in math achievement decreased but increased for math self-concept and realistic as well as investigative vocational interest. They did not find a significant effect of the reform on the choice of study subjects. An important difference to our study is that they only considered initial study choices (not the successful completion of degrees) and no labor market outcomes.

Görllitz and Gravert (2016) and Görllitz and Gravert (2018) also analyzed the same reform using aggregate administrative data obtained from the Federal Statistical Office. Their first paper finds evidence for an increase of high school dropouts which vanishes over time for males but stays persistent for females. The individuals not dropping out of high school, however, appeared to be better prepared for and more likely to enroll in university studies after the reform. The second paper explored the positive effects on higher education enrollment due to the reform in more detail. According to Görllitz and Gravert (2018), this higher enrolment did not go along with an increase of females entering STEM subjects. Only males show a robust positive effect for STEM. Our study differs from Görllitz and Gravert (2016) and Görllitz and Gravert (2018) in that we use microdata, that we not only consider college degrees but also occupational choices after graduation, and that we take into account a rich set of individual and aggregate covariates in our analysis.

Our study connects to a wider literature that analyzes the effects of differences in school curricula, especially with regard to math and natural sciences, to later educational and economic outcomes. A number of papers have studied the effects of differential exposure to math curricula on later decisions in high school and on college attendance. For example, using observational data for the U.S. and controlling for selection on observables, Aughinbaugh (2012) found that a more rigorous high school math curriculum is associated with a higher probability of attending college. Justman and Méndez (2018) showed for Australia, that choices of STEM subjects in later high school years are not driven by prior differences in mathematical achievement but that women required stronger signals for mathematical ability to choose male-dominated subjects. A few studies have used quasi-experimental variation to study the effects of math curricula on later school outcomes. Broecke (2010) exploited the introduction of a ‘triple science’ option in British high schools and showed that those choosing this option were more likely to choose science courses at later grades. However, this effect was restricted to men, and it was stronger for pupils from lower backgrounds. Domina and Saldana (2012) examined the intensification of mathematics curricula in American high schools over the period 1982 to 2004. Their results suggest that intensification generally reduced social stratification in course credit completion but left inequality in some more advanced subareas very pronounced. Based on a regression discontinuity design, Cortes et al. (2012) studied an intensive math
instruction policy that doubled instruction time for low-skilled 9th graders. They show that this policy had substantial positive effects on test scores, high school graduation, and college enrollment.

A considerable literature has looked at college major choice and its determinants (for an overview, see Altonji et al., 2012). Here, we only focus on articles that address the question of STEM vs. non-STEM majors. In an early contribution based on controlling for observables, Levine and Zimmerman (1995) considered the effects of taking more high-school math on wages, college majors and gender-traditional occupations. They found that more math was associated with a higher likelihood of completing a technical degree, working in a technical job or a job traditional for one’s sex, but only for women. Ardicianono (2004) estimated a sophisticated structural choice model of college major choices incorporating aspects such as learning about one’s abilities and uncertainty of educational outcomes. He found that math ability (but not verbal ability) is important for selection into certain majors (especially STEM), but that ability differences are far from enough to account for observed choices. Rather, differences in job and school preferences dominate college major choices. Based on data for Ontario, Card and Payne (2017) study STEM major choices in relation to an index of STEM readiness at the end of high school. They show that men and women do not differ in STEM readiness but that males not interested in STEM subjects are less likely to start university studies. Only few studies have used quasi-experimental variation in order to study the effect of prior exposure to math on college major choices (as we do in our study). Jia (2016) exploits state-specific increases in high school math curriculum requirements in the U.S. in order to measure the effect of stricter math requirements on college STEM attainment. She finds that stricter requirements increase STEM attainment to a certain extent, but only for white males. De Philippis (2017) also used quasi-experimental variation in the form of a reform that allowed secondary schools in the U.K. to offer more science to high-ability 14-year olds. Again, her results suggest that introducing this option increased men’s willingness to enroll in STEM degrees but not women’s.

A much smaller literature has focussed on the effects of math and science curricula on STEM choices and outcomes outside the education system. One strand of the literature started by Altonji (1995) considers the effects of math curricula on later wages, see e.g., Rose and Betts (2004) (using observational data) and Joensen and Nielsen (2009), Joensen and Nielsen (2016), Goodman (2019) (using quasi-experimental data). However, these contributions do typically not address the question of STEM vs. non-STEM occupations. For example, Goodman (2019) shows that state changes in minimum high school math requirements substantially increased underprivileged students’ completed math coursework and later earnings. Joensen and Nielsen (2009) and Joensen and Nielsen (2016) exploited a pilot scheme in Denmark that reduced the cost of choosing advanced math at high school.
They show that this pilot scheme drew girls from the top of the ability distribution and boys from the middle of the ability distribution into choosing more advanced mathematics. Their results suggest that only the female but not the male compliers benefited from the pilot scheme in the form of higher later earnings, a higher rate of completed STEM degrees, and higher career outcomes. Joensen and Nielsen (2009) and Joensen and Nielsen (2016) did not address whether more math at high school increased the inflow into STEM occupations as we do in this paper. In fact, we are not aware of any other study that uses quasi-experimental variation in school curricula to study its effect on STEM occupations. Based on a selection on observables strategy, Morgan et al. (2013) examined college major selection and occupational plans (i.e. not actual outcomes). Already conditioning on having completed a STEM degree, Sassler et al. (2017a) study transitions of STEM graduates into STEM occupations. They find that the highest proportion of transitions into STEM vs. non-STEM jobs can be accounted for by the type of STEM degree (i.e. engineering and computer science vs. other degrees), while attitudes and expectations account for a much smaller part.

Finally, we want to point out a number of studies that focussed on certain aspects of college major choice that will be important for the interpretation of our results. Ardicianono (2004), Zafar (2013) and Wiswall and Zafar (2015) suggest that preferences and not factors like expectations or perceived abilities explain gender differences in college major choices and later wages. For a similar result, see Daymont and Andrisani (1984). Shi (2018) finds that differences in other-regarding and in professional preferences are more important for the intentions of females to enroll or not to enroll in engineering programs than prior achievement or lack of academic self-confidence. Ceci and Williams (2010) summarize that “...among a combination of interrelated factors, preferences and choices – both freely made and constrained – are the most significant cause of women’s underrepresentation [in math-intensive fields].”

What are possible sources of gender differences in preferences for different fields? Buser et al. (2017) and Gneezy et al. (2003) point out that an important factor behind the STEM gap may be gender differences in willingness to compete. According to these results, men seek competitive fields while women try to avoid them. One reason why mathematics and STEM subjects can be considered particularly competitive is because they allow for a clear distinction between ‘right’ and ‘wrong’. Another possible source of female underrepresentation may be biased self-assessment (Corell, 2001), stereotypes and identity issues (e.g. Cech et al., 2011; Cheryan, 2012; Franceschini et al., 2014; DelCarpio and Guadalupe, 2018). For example, Franceschini et al. (2014) suggest that women are more easily intimidated by ‘stereotype threat’, i.e., by pieces of information that make STEM subjects appear inappropriate for them. Finally, there is evidence that gender differences in underlying preferences and resulting choices may be determined by cultural differences,
see, e.g., Guiso et al. (2008), who find that gender math differences are less pronounced in more gender-equal cultures. Friedman-Sokuler and Justman (2018) find the opposite result for Israel where the STEM gender gap in education is smaller in the Arab than in the Hebrew part of the population. Sassler et al. (2017b) also point out that negative STEM gender gaps are far from universal, and that countries such as Iran, Saudi Arabia, and Malaysia exhibit positive STEM gender gaps.

3 Institutional background

In Germany, educational policies are largely determined at the federal state level allowing states some degree of freedom to deviate from the general structure of the school system that is shared by all states. Using this freedom, the federal state of Baden-Württemberg (the third largest of all federal states), introduced a significant reform of the high school curriculum in 2002 that provides us with a natural experiment. Apart from Baden-Württemberg, only the federal states Mecklenburg-Vorpommern and Sachsen-Anhalt carried out other school reforms at the high school level during the period of interest, which is why we drop these states from our analysis.

In general, the German school system is characterized by a pronounced early tracking school structure. After the fourth or sixth grade, students sort into different secondary school types. Most common is the differentiation between Hauptschule (lower secondary school track), Realschule (middle secondary school track) and Gymnasium (highest secondary school track). Hauptschule ends after nine years, Realschule after ten years. Both school types leave students without a higher education entrance qualification (HEEQ). Following graduation from lower or middle secondary school, students may continue with Gymnasium to obtain a HEEQ. In addition to Gymnasium, there are specialized/vocational high schools and a number of more indirect ways to obtain a HEEQ, which are, however, chosen only by a minority of students. The HEEQ from Gymnasium generally allows individuals to enroll at universities in all types of studies. Specialized and vocational high schools HEEQ usually allow students to study only a subset of study programmes for which the secondary school was specialized. For tertiary education in Germany, two options are available: universities and universities of applied science (including different variants). Studies at universities of applied science are more practical and less academically oriented.

The pre-reform high school curriculum was similar for all German states. In this curriculum, all students attended similar courses during the first ten or eleven school years. At grade ten or eleven, however, students were asked to choose a specific combination of subjects for the last two years of high school, with mild restrictions on which combinations
and exam levels were possible. From the chosen classes, two had to be of an advanced level and several others of a basic level. In Baden-Württemberg, for example, at least a basic math and a basic German class had to be taken, in addition to at least one natural science class. If math, German or a science class was chosen at an advanced level, then students could fill their basic courses with other subjects. If they chose math, German and a science class as basic courses, however, they were free to choose other subjects such as languages, arts or even sports as their advanced classes. Advanced classes were held five hours a week, basic classes only two hours per week. Given the nature of specialized/vocational high schools, the choices at these schools were less flexible. For the HEEQ in both types of school, three written high school exams and one oral exam in two advanced and two basic courses had to be passed.

In 1999, Baden-Württemberg announced a reform affecting students starting their second last school year at the Gymnasium from 2002 onwards. Specialized/vocational high schools introduced a modified version of these reforms one year later. As a consequence, academic high school graduates from 2004 and specialized/vocational high school graduates from 2005 onwards were affected by the reform. The reformed high school curriculum forced all students to attend a mandatory advanced class in mathematics, in German and in one foreign language. In addition, two more advanced courses in one natural science and/or another foreign language had to be taken. This implied that the total number of mandatory natural science courses increased from one to two (one of them potentially at the basic level). Because of the higher number of different classes, advanced classes were reduced from 5 to 4 hours per week. The minimum instruction time increased from 26 hours per week to 30 hours per week. Although there were some additional aspects to the reform, it is fair to say that its essential ingredient was a significant shift towards more instruction time in math and natural sciences for a large number of students who previously would not or would only have chosen these subjects at the basic level (with only half the instruction time).

— Figures 1 and 2 here —

In order to illustrate the drastic nature and the comprehensive reach of the reform, figures 1 and 2 present raw administrative data showing the impact of the reform as compared to the situation in other federal states. The figures refer to the second qualification phase of the German Abitur at the Gymnasium. Administrative data is available from 2002, but a number of missings and differences in coding make some values before 2003 unusable. As figure 1 shows, advanced math participation varied from around .08 for Niedersachsen in 2004 up to 1 for Baden-Württemberg. The graph demonstrates that the reform in Baden-Württemberg had a substantial impact. Without mandatory advanced math
classes, the highest share was around .5 in Saarland. All other states range between .1 and .35. Only Thüringen was constant above .4. Figure 2 shows that the proportion of females taking advanced math classes was generally lower than that of males. Again some values are missing, e.g., because there is no gender-specific administrative data available. Unfortunately, the value for Baden-Württemberg in 2003 is missing as well. However, we have no reason to believe that the gender difference in Baden-Württemberg before the reform was much different from that in other states. For females, the highest percentage without mandatory advanced math classes was .4, but mostly somewhere between .10 and .25. This difference shows that, in general, females were more affected by the reform than males.

Taken together, these numbers illustrate the dramatic impact the curriculum reform had on the level and the instruction time for the subject math during the last two years of high school. For over 80 percent of students, instruction time doubled as students who would have attended a basic course before the reform (2 hours per week), were forced into the advanced course after the reform (4 hours a week). For women, the percentage of students switching to more instruction time was even higher as the share of female students taking advanced math courses before the reform was below that of men. For natural sciences, we observe a similarly high impact of the reform on instruction time through the introduction of an additional advanced level course in one of the natural sciences (details not shown here).

4 Econometric methods

4.1 Difference-in-difference estimation

For our estimations, we employ a difference-in-difference setup with gender interactions in order to obtain gender-specific difference-in-difference effects. In this setup, the situation before and after the reform is compared where the non-treated federal states serve as a counterfactual for the treated state. Our regression model is

\[
y_{ist} = \alpha + \gamma_1 \cdot \text{After}_{it} + \gamma_2 \cdot \text{BaWu}_s + \rho \cdot \text{Treatment}_{ist} \\
+ \gamma_3 \cdot \text{Female}_{ist} + \gamma_4 \cdot (\text{After}_{it} \cdot \text{Female}_{ist}) + \gamma_5 \cdot (\text{BaWu}_s \cdot \text{Female}_{ist}) \\
+ \lambda \cdot (\text{Treatment}_{ist} \cdot \text{Female}_{ist}) \\
+ X'_{ist} \cdot \beta + \eta_s + \nu_t + \epsilon_{ist}
\]

where the index \(s\) indicates in which federal state individual \(i\) obtained the higher edu-
cation entrance qualification (HEEQ) and \( t \) denotes the year the individual obtained her HEEQ. The dependent variable \( y_{ist} \) represents a binary outcome, e.g., whether or not the individual \( i \) from state \( s \) and HEEQ-year \( t \) later completed a STEM degree or worked in a STEM occupation. The dummies \( After_t \) and \( BaWu_s \) indicate whether the individual obtained her HEEQ after the reform year 2004 (rather than before), and whether the individual obtained her HEEQ in the state of Baden Württemberg (rather than elsewhere). The treatment variable \( Treatment_{ist} \) is the product of \( After_t \) and \( BaWu_s \) indicating whether the individual’s high school curriculum was changed by the reform. The vector \( X_{ist} \) contains a number of individual and federal state covariates explained in more detail below, while \( \eta_s \) and \( \nu_t \) are HEEQ-state and -year fixed effects.

In order to differentiate the difference-in-difference effects between genders, we include interactions of the difference-in-difference terms with a female dummy \( Female_{ist} \) indicating whether individual \( i \) is a woman. As a consequence, \( \rho \) represents the treatment effect for men (i.e. individuals with \( Female_{ist} = 0 \)), while \( \rho + \lambda \) represents the treatment effects for women (i.e. individuals with \( Female_{ist} = 1 \)). The parameter \( \lambda \) represents the gender difference of the reform effect. Overall, this setup identifies the reform effects \( \rho \) and \( \rho + \lambda \) by comparing individuals before and after the reform in Baden Württemberg where the situation before and after the reform year in other federal states is taken as a counterfactual scenario. There may be general time-constant differences between treatment and control states \( \eta_s \) as well as common time trends in STEM participation \( \nu_t \) (common across all states). Moreover, we include into \( X_{ist} \) a large number of time-varying covariates at the state level (such as income per capita, unemployment rate, density of tertiary institutions, see below) that aim to pick up potentially differential developments in STEM participation across states.

The reform effects \( \rho \) and \( \rho + \lambda \) represent the total effects of the reform, i.e. those for the larger group of individuals who would have had much lower exposure to math and natural sciences without the reform, and those for the smaller group of individuals who would have participated in advanced math and natural science courses anyway, but whose instruction time would have been slightly higher without the reform. Despite its mixed nature, the treatment effect estimated here represents an interesting and relevant policy parameter corresponding to a well-defined real-world intervention which could in principle be implemented in other federal states as well.

### 4.2 Standard errors with few treated clusters

It is well-known that in difference-in-difference setups, it is crucial to control for potential intra-cluster dependence. In our application, this clustering refers to both the state and the time dimension. Ignoring intra-cluster dependence will downward bias standard errors
and lead to over-rejection rates (Bertrand et al., 2004). Even when correcting for the group dependency using cluster-robust standard errors, there may be remaining bias because of finite clusters sizes, unbalanced treated and untreated numbers of clusters, and different cluster sizes (Cameron and Miller, 2015). The usual clustered variance estimation formula requires that not only the number of observations but also the number of clusters to go to infinity (Cameron and Miller, 2015). This assumption is clearly violated in setups such as ours, with just a few clusters and the number of treated clusters very small (one in our case).

Mackinnon and Webb (2017), Mackinnon and Webb (2018), and Roodman et al. (2018) have analyzed the problems of cluster inference in cases with few clusters, few treated clusters and (possibly ‘wildly’) differing cluster sizes. Mackinnon and Webb (2017) introduced the wild subcluster bootstrap and showed that it is superior to other inference procedures in the case with only few treated clusters. The wild subcluster bootstrap is an intermediate case between the wild cluster bootstrap (clustering at the highest level of clusters observed in the data) and the ordinary individual wild bootstrap (treating each individual as a cluster). For example, if states are the highest level clusters, the wild subcluster bootstrap may cluster at the state-year level. Mackinnon and Webb (2017) also advocate the comparison of restricted and unrestricted standard errors (i.e., with or without imposing the null hypothesis) as a diagnostic test for the validity of standard errors. If both do not coincide, one can be sure about their invalidity. If they are close, there is probably no problem.

Mackinnon and Webb (2018) present simulation results for a difference-in-difference setup with just one treated cluster and varying cluster sizes similar to ours. Their results suggest that in this scenario, the individual wild bootstrap performs best, while conventional clustered standard errors dramatically over-reject and wild cluster bootstrap tests either over-reject (unrestricted version) or under-reject (restricted version). Moreover, for the individual wild bootstrap, the restricted and the unrestricted versions coincide well. In our empirical application below, we follow this interpretation and Roodman et al. (2018) by computing conventional \( t(G - 1) \) clustered standard errors, cluster, wild subcluster, and individual wild bootstrap p-values. Our results are very much in line with Mackinnon and Webb (2017), Mackinnon and Webb (2018) and Roodman et al. (2018). We obtain very small p-values for conventional clustered standard errors, and discrepancies between restricted and unrestricted (sub)cluster p-values which become smaller as we approach the case of the individual wild bootstrap. Taking the limit case of the individual wild bootstrap as the most credible one, we obtain significant treatment only in a number of cases, whereas especially the conventional clustered p-values would suggest an extremely high level of statistical significance throughout our results.
As explained in the previous section, we include a large set of time-varying covariates at the state level into our difference-in-difference regressions to pick up potentially differential time trends across states. One might hope that this additionally takes out intra-cluster correlation, mitigating problems of cluster inference. It is, therefore, interesting to see how the inclusion of such variables changes cluster inference. We found that including such variables generally did not change conclusions from cluster inference very much, or if it did, then in unsystematic ways.

5 Data

The data for our study are provided by the Centre for Higher Education Research and Science Studies (DZHW), Hannover. The DZHW starts a new survey of university graduates every four years. For our analysis, we use the cohorts 2005, 2009 and 2013. The survey includes rich information on parental background, on the individual’s higher education entrance qualification, on choices during the study programmes, and on labor market entry. For our analysis, we eventually exclude individuals with a HEEQ obtained before 1995 for cohort 2005, before 1999 for cohort 2009 and before 2003 for cohort 2013 in order to drop unrepresentative long-term students. For the same reason, we exclude individuals born before 1970 in cohort 2005, before 1974 in cohort 2009 and before 1978 in cohort 2013.

Table 1 shows some basic sample information. All summary statistics and estimates reported in the following use the survey weights provided by the DZHW. The three cohorts have approximately the same size. The table also lists the individual-level covariates that we include in our difference-in-difference regressions. These include gender, cohort, age, four categories of parental education and two categories of parental occupation.

Table 2 presents summary statistics for the degree and occupational outcome variables used in the regressions. The degree variables are dummies indicating whether or not a particular individual obtained a degree in a particular (sub)field. Labels such as ‘at least one STEM degree’ mean that we have a small number of individuals with more than one degree.

---

2We use only the first wave of each cohort because subsequent waves suffer from considerable attrition and because for cohort 2013, only the first wave is available at this point.
but count them as STEM if at least one of them is in STEM. Following common practice, we include into STEM all fields from science, technology, engineering and mathematics. More precisely, our STEM category includes sciences (biology, chemistry, pharmacy, geosciences, physics), technology (computer science), engineering (all subfields of engineering) and mathematics. As indicated above, we also consider smaller subsets of STEM fields. In the category PTEM, we only include STEM subfields with a particularly pronounced mathematical or technical orientation (i.e. physics, computer science, engineering and mathematics). We also consider the smaller STEM subsets life sciences (biology and chemistry), math and physics, as well as engineering and computer science.

For the occupational outcomes, the data provides the KldB occupation code (German classification of occupations). For 2005, this is the KldB 1992, whereas for the other cohorts it is the KldB 2010. The German Federal Employment Agency provides a translation for STEM and non-STEM occupations, but only for the KldB 2010 (Bundesagentur für Arbeit, 2016). For the KldB 1992 codes, we used a translation form KldB 1992 to KldB 2010 which left us with a small number of cases for which it was not possible to assign a clear STEM or non-STEM status based on the 2010 STEM classification (because occupations were more or less specific in the KldB 1992 classification than in the KldB 2010 classification). For these cases, we used a specific procedure, details of which are available on request. As table 2 shows, around 34 percent of our sample members obtained a STEM degree, and for 27 percent this was in PTEM. Note that PTEM includes only physics from the sciences fields in STEM and thus must be smaller. Around 5.3 percent of the observations in the sample have at least one degree in life sciences and 3.97 percent in math or physics, the latter being the smallest group we investigate. Engineering and computer science is the biggest group within STEM with a share of 22.6 percent. For the occupational outcomes, we obtain a qualitatively similar picture as for the degrees, except that the numbers are generally lower (because not all individuals work in jobs that directly match their fields of study).

In order to address potential issues with the parallel trend assumptions and in order to minimize remaining intra-cluster correlation, we include a set of state and time specific variables as shown in table 3. We include variables from four different groups: economics, education, politics, and other variables. All variables are measured at the state level.
6 Empirical Results

6.1 Effects on STEM degrees

Our discussion of the empirical results focuses on estimated treatment effects, the gender difference in treatment effects, and the appropriate p-values for the hypothesis that treatment and gender differences are zero. Apart from the conventional clustered p-value (here labeled $t(G - 1)$), we compute six different bootstrap p-values: the restricted wild cluster bootstrap (WCR), the unrestricted wild cluster bootstrap (WCU), the restricted wild subcluster bootstrap (WSR), the unrestricted wild subcluster bootstrap (WSU), the restricted ordinary wild bootstrap (WOR), and the unrestricted ordinary wild bootstrap (WOU), as suggested by Mackinnon and Webb (2017) and Roodman et al. (2018). As in Mackinnon and Webb (2017) and Roodman et al. (2018), we chose as subclusters the state-year level.

The first set of results is shown in table 4. Depending on the specification, we obtain a positive reform effect for male graduation in STEM fields of 6 to 11 percentage points. The question is of course what degree of statistical significance we can attribute to this result. The p-values obtained from conventional clustered standard errors signal high degrees of significance, but we know that this impression should not be trusted. On the other hand, the wild cluster bootstrap and the wild subcluster bootstrap generally offer diverging results for the restricted and unrestricted versions, which, according to Mackinnon and Webb (2017) and Roodman et al. (2018), signals their invalidity. However, we also observe that the difference between the restricted and unrestricted values tends to shrink as we move from the cluster to the wild subcluster bootstrap. This is illustrated in figure 3 in which we plot the distribution of bootstrapped p-values for the different bootstrap variants (the picture looks similar to the ones in Roodman et al., 2018). The difference between the restricted and the unrestricted values is very small for the case in which each individual is treated as its own cluster (i.e., ordinary individual wild bootstrap) lending credibility to the p-values for this case. Taking these p-values, we conclude that the male treatment effect for STEM degrees is statistically significant at the 10 percent level if we take specification (3) which includes state and year effects. For specification (4), which includes in addition the state level variables, the p-values jump above the 10 percent level suggesting no statistical significance at conventional levels. Note, however, that adding the rather large number of extra regressors may have opposing effects on statistical significance. On the one hand, one may gain precision by reducing the error variance and intra-cluster correlation.
On the other hand, however, one consumes degrees of freedom potentially leading to less significant results. To summarize, we obtain a positive treatment effect on STEM degrees for men, which is not or only weakly statistically significant. This is in line with the results in Hübner et al. (2018) who also did not find statistically significant reform effects on study decisions.

The second half of table 4 reports the results for the gender difference of the reform effect. The results imply that, if there is a positive treatment effect for men, it is undone for women. Taking the non-diverging ordinary wild bootstrap p-values of our preferred specifications (3) and (4), this negative gender difference is significant at the 10 percent level and marginally significant at the 5 percent level. We, therefore, obtain the result that the reform effects on STEM degrees significantly differed between men and women in the sense that women’s propensity to complete a STEM degree was not (or even negatively) affected by the reform.

### 6.2 Heterogeneity within STEM degrees: PTEM, life sciences, math and physics, engineering and computer science

As the combined group of STEM subjects is a rather broad category, we now look into subgroups within STEM. For this analysis, we only report the results for the most comprehensive specification including the full set of covariates as well as year and state fixed effects (this was specification (4) in table 4; more detailed results are available in the appendix). The first category we consider is PTEM, which includes all the fields in STEM except biology, chemistry, geosciences, and pharmacy. PTEM is a policy-relevant category as it includes all the STEM subfields with a particularly pronounced mathematical or technical orientation. For this category, we find a positive but insignificant effect in column (2) of table 5.

--- Table 5 here ---

Next, we consider life sciences (biology and chemistry). For this category, we obtain a male treatment effect close to zero which is also not statistically significant (column (3) of table 5). By contrast, our estimates suggest a statistically significant negative effect of the reform on the successful completion of math or physics degree for men as shown in column (4) table 5. Finally, our results indicate a large reform effect on the number of male engineering and computer science degrees which is statistically significant at the 10 percent level based on the ordinary wild bootstrap p-values which appear to be most reliable according to column (5) of table 5. Note that the counteracting effects in math and physics on the one hand, and in engineering and computer science on the other, are
the likely explanation that we obtain no significant effect for the PTEM category which combines all of these subjects (column (2) of 5). This shows that it is important to consider heterogeneity within STEM subjects.

The lower half of table 5 shows the gender difference of the above effects. The estimated coefficient for the gender difference for PTEM is negative (undoing the positive but insignificant male treatment effect), but the statistical significance of this gender difference is questionable. All other gender differences are small and statistically insignificant. An exception is the negative gender difference in the engineering and computer science category which is marginally significant and largely neutralizes for females the very positive reform effect found for males (column (5) of table 5).

On balance, our results suggest that the positive but statistically very weak male treatment effect found for STEM degrees (table 4) is driven by a positive effect for engineering and computer science, while there appears to be a smaller negative effect on math and physics degree. For women, the positive effect on engineering and computer science degrees is either weaker or non-existent, while the negative effect on math and physics appears to be similar to that of men.

### 6.3 Effects on STEM occupations

For male graduates entering STEM occupations after graduation, our preferred specification (4) in table 6 suggests positive and statistically significant effects of the reform. Again, the negative and marginally significant gender difference in the lower half of table 6 points to a non-effect of the reform women’s occupational choices after graduation. The picture for occupations is clearer in terms of statistical significance than that for the degrees, possibly because STEM occupations represent a smaller percentage of all occupations than STEM degrees represent among all degrees.

In table 7, we differentiate between different subfields of STEM occupations. We point out that some of these subfields represent only very small percentages of jobs in the labor market. The reason is that occupations are harder to map in a clear way to different subjects than degrees. Interestingly, the results present a similar picture as the one for degrees. There is a positive (but this time slightly more significant) effect for males on PTEM occupations, and a corresponding positive effect on engineering and computer science occupations. These effects are non-existent for women. Apart from this, there are statistically significant effects on life sciences and math/physics for men, which are very
small in magnitude, however.

— Table 7 here —

6.4 Placebo test

Given that our sample design is cohort-based (with new cohorts every four years), we do not observe long enough and continuous pre-treatment periods in order to graphically check the common trend assumption. We, therefore, report the following placebo test defining an artificial (i.e., non-existent) treatment for Baden Württemberg for the years before the actual treatment year 2004. For this test, we have to exclude the HEEQ-years from the treatment year onwards. We then define the year 2000 as the artificial treatment year so that the years 2000, 2001, 2002 and 2003 represent the placebo treatment period. In order to have enough observations for the comparison group, we include in the placebo analysis also individuals born before 1995 (not included in our main analysis).

— Table 8 here —

The placebo results for STEM degrees shown in table 8 give the desired result of no significant treatment and gender difference effects for STEM and all its subcategories, with one exception. The exception is a significant positive effect for math and physics for men which points in the opposite direction of the actual treatment effect for this category as shown in table 5. This calls into question the result for the original treatment effect, although it is comforting that the placebo effect goes into the opposite direction. A possible reason for potentially spurious results for the math and physics category is that this is a very small group of students for which spurious changes may easily look large.

— Table 9 here —

For the results on STEM occupations and its subcategories, the placebo results are shown in table 9. Here, we find no significant placebo effects whatsoever, increasing our confidence in the difference-in-difference results presented in tables 6 and 7.


7 Conclusion

This paper has analyzed the consequences of a substantial curriculum reform of the last two years of high school in one of the German federal states on the share of male and female students who complete university degrees in STEM subjects and who work in STEM occupations after graduation. Among some other aspects, the curriculum reform doubled the instruction time and increased the level of instruction in math and natural sciences for some 80 percent of the students and an even higher proportion female students.

Our results based on difference-in-difference regressions provide weak evidence for a positive effect on the share of male STEM graduates which appears to be driven by a significantly positive effect for engineering and computer science combined with a significant, but counteracting negative effect on the completion of math and physics degrees. Despite the fact that women were affected to a larger extent by the reform than men, we find no positive reform effects on the female completion of STEM degrees, but the same relatively small negative effect on the number of math and physics degrees.

A possible interpretation of these results is that the reform increased the level of preparation and motivation in math and natural sciences for male students who were interested in pursuing a technical degree in engineering or computer science, while the increased exposure to math and natural science may also have deterred some students from pursuing these subjects directly in a study programme. For women, we observe no significant reform effects (except for a small negative effect on the completion of math and physics degrees) suggesting that, although they would have been better prepared for engineering and computer science degrees, and although more women than men were affected by the reform, it was not the case that more women pursued and eventually completed such degrees.

Our results for occupations are very much in line with the results for completed degrees but slightly clearer in terms of statistical significance. In particular, we find a positive reform effect on the share of men working in STEM occupations which is also statistically significant. We find no such effect for women. Again, the positive effect on the share of individuals ending up in STEM occupations appears to be driven by increased shares of individuals working in engineering and computer science occupations, while there is a small counteracting effect on the share of men working in math and physics occupations.

Our results are consistent with the hypothesis that gender differences in STEM attainment and STEM employment are rooted in preference differences or issues of role identity that make it hard to increase female STEM participation without deeper cultural or institutional changes (Ceci and Williams, 2010; Cech et al., 2011; Wiswall and Zafar, 2015; DelCarpio and Guadalupe, 2018). Given that the reform forced all students into an
advanced math treatment, it may also have induced negative effects in terms of stereotype threat (Franceschini et al., 2014) or underscored the competitive dimensions of the subject (Buser et al., 2017). Our results for math and physics indicate that there may have been deterrence effects also for men who received a clearer signal about the nature of the subject after the reform, potentially leading some individuals away from pursuing the math or physics in study programs. Another possibility is that these effects arise because for some individuals, math exposure was reduced by the reform (for those who would have attended the advanced math course for five hours per week before the reform instead of four hours after the reform).

Finally, following contributions such as Bertrand et al. (2004), Cameron and Miller (2015) and Mackinnon and Webb (2018), our results provide a further empirical example of how inappropriate standard errors in clustered difference-in-difference may potentially lead to erroneous conclusions about statistical significance. Compared to the conventional use of clustered standard errors, using inference procedures proposed by Mackinnon and Webb (2018) and Roodman et al. (2018) that have been shown to be the most appropriate so far, drastically reduces the statistical significance of our estimates and call for a more cautious interpretation.

References


Broecke, S. (2010). Does Offering More Science at School Increase the Supply of Scientists? The Impact of Offering Triple Science at GCSE on Subsequent Educational Choices and


Sources: Statistical Offices of the Federal States. Notes: Panel A includes the states Brandenburg (BB), Baden-Württemberg (BW), Bayern (BA), Mecklenburg-Vorpommern (MV), Schleswig-Holstein (SH), Sachsen-Anhalt (ST) and Thüringen (TH). Panel B shows the development for Saarland (SL) and Sachsen (SN), Berlin (BE), Bremen (HB), Hessen (HE), Hamburg (HH), Nordrhein-Westfalen (NW), Rheinland-Pfalz (RP) and Niedersachsen (NS). The data is provided by each state on a voluntary basis, with missing years.

Source: Statistical Offices of the Federal States. Notes: Some states did not provide gender specific information for all years, which is why there are some missing years.
Figure 3: Bootstrap p-value distribution for STEM degrees

Notes: The plots show the bootstrapped $t$-statistics based on the particular method. The vertical line displays the $t$-statistic of the regression. See text for more explanations.
Figure 4: Bootstrap p-value distribution for subfield degrees

Notes: The plots show the bootstrapped t-statistics based on the particular method. The vertical line displays the t-statistic of the regression. See text for more explanations.
Table 1: General sample information

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment period</td>
<td>21,633</td>
<td>.3475</td>
<td>0.4762</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HEEQ from BaWu</td>
<td>21,633</td>
<td>.1664</td>
<td>0.3725</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Treated individuals</td>
<td>21,633</td>
<td>.0587</td>
<td>0.2351</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Female</td>
<td>21,633</td>
<td>.5366</td>
<td>0.4987</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Treated females</td>
<td>21,633</td>
<td>.0349</td>
<td>0.1834</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: HEEQ from BaWu is a dummy, equal to one if the higher education entrance qualification (HEEQ) is from Baden-Württemberg and zero otherwise. Not included are individuals with a HEEQ from foreign countries or with missing information.

Table 2: Descriptive statistics for outcomes

<table>
<thead>
<tr>
<th>Degree outcomes</th>
<th>Obs</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one degree in STEM</td>
<td>21,633</td>
<td>.3455</td>
<td>0.4755</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>At least one degree in PTEM</td>
<td>21,633</td>
<td>.267</td>
<td>0.4424</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>At least one degree in life sciences</td>
<td>21,633</td>
<td>.0557</td>
<td>0.2294</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>At least one degree in math or physics</td>
<td>21,633</td>
<td>.0397</td>
<td>0.1953</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>At least one degree in engineering or computer science</td>
<td>21,633</td>
<td>.2257</td>
<td>0.4180</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: The drop in the number of observations is due to the fact that within the first year after graduation not all individuals have entered the labor market. STEM = biology, chemistry, pharmacy, geosciences, physics, computer science, engineering, mathematics. PTEM = physics, computer science, engineering, mathematics. Life sciences = biology and chemistry.
| Table 3: Summary statistics for federal state variables |
|-----------------------------|----------------|-----------|-----------|-----------|-----------|
| **Economics**               |                |           |           |           |           |
| Unemployment rate (OECD)    | 21,633         | 0.0461    | 0.0222    | 0.0220    | 0.0387    | 0.1129    |
| Labor market participation rate | 21,633         | .4979     | 0.0250    | 0.4174    | .5011     | 0.5515    |
| GDP per capita              | 21,633         | 27.7301   | 5.9514    | 13.7080   | 27.5260   | 55.9290   |
| Income per capita           | 21,633         | 17.1211   | 2.0551    | 11.0530   | 17.0180   | 22.3950   |
| CPI                         | 21,633         | 8.9117    | 0.5140    | 7.9300    | 8.8633    | 9.9692    |
| **Education**               |                |           |           |           |           |           |
| Density of tertiary institutions | 21,633         | 5.5914    | 1.3241    | 3.4975    | 5.1922    | 12.3998   |
| BAföG expenditure per capita | 21,633         | 0.0211    | 0.0091    | 0.0104    | 0.0191    | 0.0592    |
| Funded students per capita  | 21,633         | 8.1646    | 3.0702    | 4.5418    | 7.3832    | 19.0762   |
| **Elections/Politics**      |                |           |           |           |           |           |
| Voter turnout               | 21,633         | 0.6320    | 0.0580    | 0.5220    | 0.6300    | 0.8350    |
| Votes for SPD in percent    | 21,633         | 32.0057   | 10.6544   | 9.8000    | 33.3000   | 54.1000   |
| Votes for CDU in percent    | 21,633         | 43.3251   | 9.0963    | 18.7000   | 43.0000   | 60.7000   |
| Votes for DieLinke in percent | 21,633         | 4.1655    | 7.8700    | 0         | 0         | 28.0000   |
| Votes for DVU in percent    | 21,633         | 0.2590    | 1.0759    | 0         | 0         | 6.3000    |
| Votes for NPD in percent    | 21,633         | .5716     | 1.4894    | 0         | 0         | 9.2000    |
| Votes for REP in percent    | 21,633         | 2.2908    | 2.1317    | 0         | 1.8000    | 10.9000   |
| Votes for Piraten in percent | 21,633         | 0.0068    | 0.1068    | 0         | 0         | 1.9000    |
| Votes for FW in percent     | 21,633         | 1.9922    | 2.6047    | 0         | 0         | 10.2000   |
| Votes for Grüne in percent  | 21,633         | 7.5969    | 2.6047    | 0         | 7.5000    | 16.5000   |
| **Other**                   |                |           |           |           |           |           |
| Asylum applications per capita | 21,633         | 1.0866    | 0.5998    | .2638     | 1.0301    | 2.5944    |

Table 4: Regression results for STEM degree and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: At least one degree in STEM</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.0601</td>
<td>0.0618</td>
<td>0.0739</td>
<td>0.1069</td>
</tr>
<tr>
<td><strong>P-values for male treatment effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-0.11</td>
<td>-0.09</td>
<td>-0.01</td>
<td>0.30</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>25.13</td>
<td>0.00</td>
<td>21.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>7.41</td>
<td>5.71</td>
<td>6.81</td>
<td>5.91</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>16.02</td>
<td>16.42</td>
<td>14.71</td>
<td>12.51</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.0819</td>
<td>-0.0844</td>
<td>-0.0881</td>
<td>-0.0891</td>
</tr>
<tr>
<td><strong>P-values for gender difference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>25.13</td>
<td>0.00</td>
<td>24.92</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>4.00</td>
<td>1.80</td>
<td>1.60</td>
<td>1.10</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>11.71</td>
<td>9.31</td>
<td>7.11</td>
<td>7.81</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th>Age, Cohort, Parents</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0776</td>
<td>0.0865</td>
<td>0.0955</td>
<td>0.099</td>
</tr>
</tbody>
</table>

STEM = biology, chemistry, pharmacy, geosciences, physics, computer science, engineering, mathematics. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. Age stands for the birth year and the squared birth year. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command *bootees*, see Roodman et al. (2018).
Table 5: Regression results for subfield degrees and bootstrapped p-values

<table>
<thead>
<tr>
<th>Subfields</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTEM</td>
<td>Life Sciences</td>
<td>Math &amp; Physics</td>
<td>Engineering &amp; Computer Sciences</td>
</tr>
<tr>
<td>Male treatment effect</td>
<td>0.0951</td>
<td>0.0054</td>
<td>-0.0600</td>
<td>0.1580</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>31.90</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>31.73</td>
<td>3.50</td>
<td>64.86</td>
<td>61.26</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>25.93</td>
<td>16.32</td>
<td>75.58</td>
<td>77.68</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>23.12</td>
<td>24.12</td>
<td>71.27</td>
<td>74.07</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.0723</td>
<td>-0.0073</td>
<td>0.0114</td>
<td>-0.0845</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.05</td>
<td>-</td>
<td>22.97</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>19.82</td>
<td>0.00</td>
<td>53.75</td>
<td>41.34</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>4.50</td>
<td>3.50</td>
<td>66.77</td>
<td>68.27</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>10.21</td>
<td>14.41</td>
<td>66.37</td>
<td>64.96</td>
</tr>
</tbody>
</table>

Set of covariates

| All variables included in specification (4) | Yes | Yes | Yes | Yes |

Additional regression information

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.1259</td>
<td>0.0237</td>
<td>0.014</td>
<td>0.1345</td>
</tr>
</tbody>
</table>

PTEM = physics, computer science, engineering, mathematics. Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. The bootstrap p-values are calculated using the Stata command bootees, see Roodman et al. (2018).
Table 6: Regression results for STEM occupation and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: Occupation in STEM</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.1022</td>
<td>0.1032</td>
<td>0.1131</td>
<td>0.1459</td>
</tr>
<tr>
<td>( t(G-1) )</td>
<td>-</td>
<td>0.24</td>
<td>-</td>
<td>0.09</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>36.94</td>
<td>4.60</td>
<td>38.24</td>
<td>1.30</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>33.73</td>
<td>26.13</td>
<td>26.93</td>
<td>23.02</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>27.23</td>
<td>27.53</td>
<td>19.42</td>
<td>18.12</td>
</tr>
</tbody>
</table>

| Gender Difference           | -0.1131 | -0.1116 | -0.1181 | -0.1151 |
| \( t(G-1) \)                | -       | 0.01    | -      | 0.01    | -      | 0.00    | -      | 0.01    |
| Wild cluster bootstrap      | 29.23   | 0.30    | 27.53  | 0.00    | 27.23  | 0.00    | 31.43  | 0.00    |
| Ordinary wild bootstrap     | 10.91   | 12.11   | 9.11   | 9.81    | 10.41  | 8.01    | 9.71   | 9.61    |

Set of covariates

- Age, Cohort, Parents
  - No
  - Yes
  - Yes

- State and Year Fixed Effects
  - No
  - No
  - Yes

- State Variables
  - No
  - No
  - Yes

Additional regression information

- Number of observations
  - 16009
  - 16009
  - 16009
  - 16009

- \( R^2 \)
  - 0.0903
  - 0.0932
  - 0.1097
  - 0.1127

STEM = biology, chemistry, pharmacy, geosciences, physics, computer science, engineering, mathematics. P-values are presented in percent. \( t(G-1) \) refers to the p-value from the student t-distribution with \( G \) the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
Table 7: Regression results for subfield occupations and bootstrapped p-values

<table>
<thead>
<tr>
<th>Subfields</th>
<th>(2) PTEM</th>
<th>(3) Life Sciences</th>
<th>(4) Math &amp; Engineering</th>
<th>(5) Computer Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.1088</td>
<td>0.0371</td>
<td>-0.0178</td>
<td>0.1266</td>
</tr>
<tr>
<td>$P$-values for male treatment effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.05</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>24.22</td>
<td>0.20</td>
<td>23.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>37.64</td>
<td>22.22</td>
<td>11.61</td>
<td>5.31</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>13.31</td>
<td>12.61</td>
<td>1.40</td>
<td>0.90</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.1003</td>
<td>-0.0148</td>
<td>0.0091</td>
<td>-0.1094</td>
</tr>
<tr>
<td>$P$-values for gender difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td>0.95</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>32.63</td>
<td>0.00</td>
<td>36.04</td>
<td>8.61</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>9.41</td>
<td>8.71</td>
<td>24.42</td>
<td>23.82</td>
</tr>
</tbody>
</table>

Set of covariates

| All variables included in specification 4 | Yes | Yes | Yes | Yes |

Additional regression information

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.1182</td>
<td>0.0106</td>
<td>0.0087</td>
<td>0.116</td>
</tr>
</tbody>
</table>

PTEM = physics, computer science, engineering, mathematics. Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STEM</td>
<td>PTEM</td>
<td>Life Sciences</td>
<td>Math &amp; Physics</td>
<td>Engineering &amp; Computer Sciences</td>
</tr>
<tr>
<td>Male treatment effect</td>
<td>-0.0367</td>
<td>-0.0298</td>
<td>0.0151</td>
<td>0.0725</td>
<td>-0.1052</td>
</tr>
<tr>
<td><strong>P-values for male treatment effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-16.32</td>
<td>-18.83</td>
<td>-13.41</td>
<td>-0.00</td>
<td>-0.37</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>54.85</td>
<td>25.93</td>
<td>49.75</td>
<td>29.03</td>
<td>36.94</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>52.45</td>
<td>45.05</td>
<td>55.86</td>
<td>53.45</td>
<td>41.94</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>57.86</td>
<td>61.36</td>
<td>64.06</td>
<td>62.06</td>
<td>46.65</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>0.024</td>
<td>0.0543</td>
<td>-0.0208</td>
<td>-0.0084</td>
<td>0.0691</td>
</tr>
<tr>
<td><strong>P-values for gender difference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>27.52</td>
<td>-</td>
<td>5.95</td>
<td>-</td>
<td>2.62</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>51.95</td>
<td>43.54</td>
<td>39.24</td>
<td>6.01</td>
<td>33.83</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>74.07</td>
<td>74.47</td>
<td>42.94</td>
<td>42.04</td>
<td>23.42</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>74.57</td>
<td>76.28</td>
<td>45.75</td>
<td>45.05</td>
<td>29.23</td>
</tr>
</tbody>
</table>

**Set of covariates**

All variables included in specification 4: Yes, Yes, Yes, Yes, Yes

**Additional regression information**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>13285</td>
<td>13285</td>
<td>13285</td>
<td>13285</td>
<td>13285</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.1</td>
<td>0.1275</td>
<td>0.0262</td>
<td>0.023</td>
<td>0.1306</td>
</tr>
</tbody>
</table>

STEM = biology, chemistry, pharmacy, geosciences, physics, computer science, engineering, mathematics. PTEM = physics, computer science, engineering, mathematics. Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. The bootstrap p-values are calculated using the Stata command *bootees*, see Roodman et al. (2018).
Table 9: Placebo test for STEM occupations and subfields

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>-0.0654</td>
<td>-0.0556</td>
<td>-0.0098</td>
<td>0.0068</td>
<td>-0.0624</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-4.32</td>
<td>-7.80</td>
<td>-5.19</td>
<td>-13.19</td>
<td>-7.47</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>55.26</td>
<td>12.61</td>
<td>53.85</td>
<td>19.72</td>
<td>27.23</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>35.54</td>
<td>25.43</td>
<td>41.64</td>
<td>33.63</td>
<td>20.02</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>37.34</td>
<td>39.24</td>
<td>48.45</td>
<td>50.85</td>
<td>34.33</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>0.0368</td>
<td>0.0166</td>
<td>0.0202</td>
<td>-0.0167</td>
<td>0.3344</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-12.98</td>
<td>-32.72</td>
<td>-1.53</td>
<td>-0.00</td>
<td>20.94</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>44.74</td>
<td>20.26</td>
<td>73.07</td>
<td>71.27</td>
<td>32.53</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>55.86</td>
<td>56.96</td>
<td>83.18</td>
<td>83.68</td>
<td>19.62</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>60.86</td>
<td>58.96</td>
<td>86.29</td>
<td>87.39</td>
<td>19.72</td>
</tr>
</tbody>
</table>

Set of covariates

| All variables included in specification 4 | Yes | Yes | Yes | Yes | Yes |

Additional regression information

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>11089</th>
<th>11089</th>
<th>11089</th>
<th>11089</th>
<th>11089</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.097</td>
<td>0.101</td>
<td>0.0152</td>
<td>0.0106</td>
<td>0.0983</td>
</tr>
</tbody>
</table>

STEM = biology, chemistry, pharmacy, geosciences, physics, computer science, engineering, mathematics.
PTEM = physics, computer science, engineering, mathematics. Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
### Table A1: Regression results for PTEM degree and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: At least one degree in PTEM</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.0717</td>
<td>0.0733</td>
<td>0.08</td>
<td>0.0951</td>
</tr>
</tbody>
</table>

P-values for male treatment effect

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild cluster bootstrap</td>
<td>26.83</td>
<td>0.90</td>
<td>24.92</td>
<td>0.60</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>10.91</td>
<td>6.41</td>
<td>11.71</td>
<td>7.01</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>16.82</td>
<td>17.42</td>
<td>18.92</td>
<td>17.42</td>
</tr>
</tbody>
</table>

Gender Difference

| Gender Difference                    | -0.0685    | -0.0699    | -0.0731    | -0.0723    |

P-values for gender difference

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild cluster bootstrap</td>
<td>23.02</td>
<td>0.00</td>
<td>22.22</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>7.61</td>
<td>4.60</td>
<td>4.60</td>
<td>3.80</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>15.32</td>
<td>13.51</td>
<td>12.01</td>
<td>15.02</td>
</tr>
</tbody>
</table>

Set of covariates

<table>
<thead>
<tr>
<th>Age, Cohort, Parents</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Additional regression information

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.1084</td>
<td>0.1126</td>
<td>0.123</td>
<td>0.1259</td>
</tr>
</tbody>
</table>

PTEM = physics, computer science, engineering, mathematics. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
Table A2: Regression results for life sciences degree and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: At least one degree in Life Sciences</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>-0.0101</td>
<td>-0.011</td>
<td>-0.0076</td>
<td>0.0054</td>
</tr>
<tr>
<td><strong>P-values for male treatment effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>2.68</td>
<td>-</td>
<td>4.27</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>32.63</td>
<td>1.80</td>
<td>38.74</td>
<td>6.01</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>29.73</td>
<td>30.13</td>
<td>35.84</td>
<td>41.74</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>26.83</td>
<td>26.03</td>
<td>35.64</td>
<td>34.33</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.0042</td>
<td>-0.005</td>
<td>-0.0054</td>
<td>-0.0073</td>
</tr>
<tr>
<td><strong>P-values for gender difference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>34.93</td>
<td>-</td>
<td>32.12</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>69.87</td>
<td>62.66</td>
<td>63.86</td>
<td>54.55</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>83.28</td>
<td>84.28</td>
<td>77.68</td>
<td>78.28</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>83.88</td>
<td>85.09</td>
<td>78.28</td>
<td>77.88</td>
</tr>
</tbody>
</table>

### Set of covariates

<table>
<thead>
<tr>
<th>Age, Cohort, Parents</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Additional regression information

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
<th>21633</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0038</td>
<td>0.013</td>
<td>0.0205</td>
<td>0.0237</td>
</tr>
</tbody>
</table>

Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with $G$ the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
Table A3: Regression results for math or physics degree and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: At least one degree in Math or Physics</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>-0.0405</td>
<td>-0.0397</td>
<td>-0.0409</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

**P-values for male treatment effect**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>23.52</td>
<td>0.00</td>
<td>26.13</td>
<td>0.00</td>
<td>26.23</td>
<td>0.00</td>
<td>8.21</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>16.52</td>
<td>12.71</td>
<td>17.52</td>
<td>14.71</td>
<td>16.52</td>
<td>14.51</td>
<td>5.41</td>
<td>5.71</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>7.81</td>
<td>7.81</td>
<td>9.01</td>
<td>8.91</td>
<td>6.31</td>
<td>6.41</td>
<td>2.40</td>
<td>1.10</td>
</tr>
</tbody>
</table>

| Gender Difference       | 0.0113 | 0.0107 | 0.0112 | 0.0114 |

**P-values for gender difference**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>25.02</td>
<td>-</td>
<td>26.66</td>
<td>-</td>
<td>26.06</td>
<td>-</td>
<td>26.78</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>52.25</td>
<td>48.15</td>
<td>57.56</td>
<td>49.55</td>
<td>55.06</td>
<td>47.95</td>
<td>56.86</td>
<td>49.35</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>80.38</td>
<td>82.48</td>
<td>84.68</td>
<td>83.98</td>
<td>82.28</td>
<td>83.68</td>
<td>83.68</td>
<td>84.58</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>74.27</td>
<td>72.87</td>
<td>75.18</td>
<td>76.98</td>
<td>74.27</td>
<td>74.47</td>
<td>75.88</td>
<td>75.18</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th></th>
<th>Age, Cohort, Parents</th>
<th>State and Year Fixed Effects</th>
<th>State Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Cohort, Parents</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th></th>
<th>Number of observations</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21633</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with $G$ the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command bootees, see Roodman et al. (2018).
### Table A4: Regression results for engineering and computer science degree and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: At least one degree in Engineering or Computer Science</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.115</td>
<td>0.1157</td>
<td>0.1237</td>
<td>0.158</td>
</tr>
<tr>
<td><strong>P-values for male treatment effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>24.02</td>
<td>0.40</td>
<td>21.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>7.51</td>
<td>3.90</td>
<td>6.61</td>
<td>4.00</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>7.41</td>
<td>7.41</td>
<td>8.81</td>
<td>9.11</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.0812</td>
<td>-0.0819</td>
<td>-0.0851</td>
<td>-0.0845</td>
</tr>
<tr>
<td><strong>P-values for gender difference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>30.73</td>
<td>0.40</td>
<td>28.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>5.11</td>
<td>4.20</td>
<td>5.51</td>
<td>3.90</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>9.81</td>
<td>11.11</td>
<td>6.81</td>
<td>9.11</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, Cohort, Parents</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>State and Year Fixed Effects</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>State Variables</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of observations</strong></td>
<td>21633</td>
<td>21633</td>
<td>21633</td>
<td>21633</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.1165</td>
<td>0.1183</td>
<td>0.1315</td>
<td>0.1345</td>
</tr>
</tbody>
</table>

P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
Table A5: Regression results for PTEM occupation and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: Occupation in PTEM</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.0911</td>
<td>0.0907</td>
<td>0.1004</td>
<td>0.1088</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.19</td>
<td>-</td>
<td>0.09</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
<td>0.05</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>37.04</td>
<td>1.00</td>
<td>32.13</td>
<td>0.30</td>
<td>29.73</td>
<td>0.00</td>
<td>24.22</td>
<td>0.20</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>31.53</td>
<td>31.23</td>
<td>26.43</td>
<td>23.62</td>
<td>29.63</td>
<td>28.23</td>
<td>37.64</td>
<td>22.22</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>23.82</td>
<td>23.52</td>
<td>18.62</td>
<td>19.32</td>
<td>12.41</td>
<td>11.21</td>
<td>13.31</td>
<td>12.61</td>
</tr>
</tbody>
</table>

| Gender Difference                  | -0.0989| -0.0977| -0.1047| -0.1003 |

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>25.63</td>
<td>0.00</td>
<td>23.92</td>
<td>0.00</td>
<td>24.32</td>
<td>0.00</td>
<td>32.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>11.21</td>
<td>11.91</td>
<td>10.91</td>
<td>11.81</td>
<td>8.61</td>
<td>8.81</td>
<td>9.41</td>
<td>8.71</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th>Age, Cohort, Parents</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0959</td>
<td>0.0986</td>
<td>0.1148</td>
<td>0.1182</td>
</tr>
</tbody>
</table>

PTEM = physics, computer science, engineering, mathematics. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with G the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command *bootees*, see Roodman et al. (2018).
Table A6: Regression results for life sciences occupation and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: Occupation in Life Sciences</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.0111</td>
<td>0.0124</td>
<td>0.0127</td>
<td>0.0371</td>
</tr>
</tbody>
</table>

**P-values for male treatment effect**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td></td>
<td>7.62</td>
<td></td>
<td>5.05</td>
<td></td>
<td>4.63</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>41.24</td>
<td>24.22</td>
<td>38.84</td>
<td>20.62</td>
<td>35.64</td>
<td>22.32</td>
<td>23.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>54.05</td>
<td>48.85</td>
<td>53.35</td>
<td>45.75</td>
<td>55.46</td>
<td>51.95</td>
<td>11.61</td>
<td>5.31</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>43.24</td>
<td>44.34</td>
<td>36.74</td>
<td>38.24</td>
<td>37.04</td>
<td>34.63</td>
<td>1.40</td>
<td>0.90</td>
</tr>
</tbody>
</table>

| Gender Difference          | -0.0142 | -0.0138 | -0.0134 | -0.0148 |

**P-values for gender difference**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td></td>
<td>0.64</td>
<td></td>
<td>0.87</td>
<td></td>
<td>1.03</td>
<td></td>
<td>0.95</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>34.83</td>
<td>3.80</td>
<td>35.74</td>
<td>4.90</td>
<td>38.64</td>
<td>7.31</td>
<td>36.04</td>
<td>8.61</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Cohort, Parents</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0028</td>
<td>0.0071</td>
<td>0.0106</td>
<td></td>
</tr>
</tbody>
</table>

Life sciences = biology and chemistry. P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with $G$ the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).
Table A7: Regression results for math and physics occupation and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: Occupation in Math or Physics</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>-0.0132</td>
<td>-0.0132</td>
<td>-0.0129</td>
<td>-0.0178</td>
</tr>
</tbody>
</table>

**P-values for male treatment effect**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.01</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
<td>0.02</td>
<td>-</td>
<td>0.02</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>33.03</td>
<td>0.40</td>
<td>31.63</td>
<td>1.30</td>
<td>32.63</td>
<td>0.60</td>
<td>26.23</td>
<td>0.20</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>13.91</td>
<td>10.31</td>
<td>11.71</td>
<td>8.41</td>
<td>10.11</td>
<td>5.71</td>
<td>5.01</td>
<td>2.10</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>9.81</td>
<td>10.51</td>
<td>10.41</td>
<td>12.41</td>
<td>9.31</td>
<td>9.11</td>
<td>5.31</td>
<td>4.50</td>
</tr>
</tbody>
</table>

| Gender Difference | 0.0082 | 0.0083 | 0.0084 | 0.0091 |

**P-values for gender difference**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.12</td>
<td>-</td>
<td>0.11</td>
<td>-</td>
<td>0.14</td>
<td>-</td>
<td>0.08</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>28.33</td>
<td>0.60</td>
<td>27.53</td>
<td>0.70</td>
<td>27.13</td>
<td>0.20</td>
<td>29.73</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>22.32</td>
<td>20.32</td>
<td>20.32</td>
<td>18.92</td>
<td>18.82</td>
<td>19.42</td>
<td>15.62</td>
<td>13.51</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>18.52</td>
<td>19.62</td>
<td>16.02</td>
<td>17.52</td>
<td>17.72</td>
<td>18.32</td>
<td>14.71</td>
<td>14.91</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Cohort, Parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0036</td>
<td>0.0052</td>
<td>0.0074</td>
<td>0.0087</td>
</tr>
</tbody>
</table>

P-values are presented in percent. $t(G-1)$ refers to the p-value from the student t-distribution with $G$ the number of clusters. *Age* stands for the birth year and the squared birthyear. *Cohort* is a dummy for cohort 2009 and one for 2013. *Parents* refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command *bootees*, see Roodman et al. (2018).
Table A8: Regression results for engineering and computer science occupations and bootstrapped p-values

<table>
<thead>
<tr>
<th>Outcome: Occupation in Engineering or Computer Sciences</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male treatment effect</td>
<td>0.1043</td>
<td>0.1039</td>
<td>0.1134</td>
<td>0.1266</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.03</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>33.73</td>
<td>0.30</td>
<td>30.53</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild subcluster bootstrap</td>
<td>23.52</td>
<td>21.02</td>
<td>17.62</td>
<td>13.01</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>14.41</td>
<td>14.11</td>
<td>10.91</td>
<td>9.01</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>-0.1071</td>
<td>-0.106</td>
<td>-0.1131</td>
<td>-0.1094</td>
</tr>
<tr>
<td>t(G-1)</td>
<td>-</td>
<td>0.00</td>
<td>-</td>
<td>0.00</td>
</tr>
<tr>
<td>Wild cluster bootstrap</td>
<td>23.02</td>
<td>0.00</td>
<td>21.62</td>
<td>0.00</td>
</tr>
<tr>
<td>Ordinary wild bootstrap</td>
<td>8.31</td>
<td>6.61</td>
<td>6.61</td>
<td>6.81</td>
</tr>
</tbody>
</table>

**Set of covariates**

<table>
<thead>
<tr>
<th>Age, Cohort, Parents</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>State and Year Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Variables</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Additional regression information**

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
<th>16009</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.0925</td>
<td>0.0953</td>
<td>0.1125</td>
<td>0.116</td>
</tr>
</tbody>
</table>

P-values are presented in percent. t(G-1) refers to the p-value from the student t-distribution with $G$ the number of clusters. Age stands for the birth year and the squared birthyear. Cohort is a dummy for cohort 2009 and one for 2013. Parents refers to the highest educational background of the parents and the highest occupational information. State and year fixed effects are set for the state and year of the higher education entrance qualification. State variables refer to 21 variables, which vary only at the state level and over time. The bootstrap p-values are calculated using the Stata command `bootees`, see Roodman et al. (2018).