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

DISCUSSION PAPER SERIES

IZA DP No. 12236

Does More Math in High School Increase the Share of Female STEM Workers? Evidence from a Curriculum Reform

Martin Biewen
Jakob Schwerter

MARCH 2019

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

DISCUSSION PAPER SERIES

IZA DP No. 12236

Does More Math in High School Increase the Share of Female STEM Workers? Evidence from a Curriculum Reform

Martin Biewen
University of Tübingen, LEAD and IZA

Jakob Schwerter
University of Tübingen and LEAD

MARCH 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Does More Math in High School Increase the Share of Female STEM Workers? Evidence from a Curriculum Reform*

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 all students' exposure to math by making advanced math compulsory in the last two years of high school; and (ii) it roughly doubled the instruction time and increased the level of instruction in math and the natural sciences for some 80 percent of students, more so for females than for males. 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 effects for men appear 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

Corresponding author:
Martin Biewen
School of Business and Economics
University of Tübingen
Mohlstraße 36
72074 Tübingen
Germany
E-mail: martin.biewen@uni-tuebingen.de

* We thank various seminar audiences, participants at the EADE 2018, and Eric Bettinger, Alvaro Choi, Nicolas Hübner, Moshe Justman, Benjamin Nagengast, Aderonke Osikominu, Mirjam Reutter, Pia Schober, Martin Salm, Perrihan Saygin, Elisabeth Tiptom, 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.

1 Introduction

Recent technological changes strongly suggest that future economic growth can primarily be expected in fields related to science, technology, engineering, and mathematics (STEM) (OECD, 2010). One way to promote 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, STEM-related jobs are usually well-paid due to their relation to high productivity sectors of the economy. Attracting more females into STEM subjects can therefore be seen as a way to improve women's career prospects and reduce the gender wage gap and gender-related earnings inequality 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 to investigate whether and to what extent it is possible to draw more females into STEM subjects. The reform was intended to improve students' general preparation for university studies and the labor market in one of the German federal states (Schavan, 1999). However, its largest component was an increase in math and 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 in various stages of the educational 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 university 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 addresses aspects (iv) and (vii). First, by taking advantage of an exogenous shock in path decisions, we evaluate whether a reform at the end of high school changes university study decisions. We do this by evaluating the number of STEM degrees obtained by 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 graduates' transition 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, including interactions with gender in order to determine differential effects for women compared to men.

In our analysis, we exploit a relatively underutilized data source for Germany, the graduate surveys from the German Center for Higher Education and Science Studies (DZHW).[1] The DZHW surveys provide representative samples of the population of graduates from German tertiary education institutions. We focus on the first wave of cohorts in 2005/2006, 2008/2009 and 2013/2014, who were surveyed one year after graduation. The surveys include background information on secondary education, tertiary education decisions as well as information on the transition to the labor market after graduation.

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

Our study makes the following contributions. First, to the best of our knowledge, our paper is one of the first to use quasi-experimental variation to study the effects of curricula changes on the inflow into STEM occupations, and one of the few 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* in the last two years of high school by making advanced math courses, which were chosen by only 20 percent of students before the reform, compulsory. This meant that a very large proportion of the population were compliers to the reform, and that it 'leveled the playing field' between the genders and students of different ability levels (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 with 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 to compute 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

[1] One of the few studies using these data known to us is Parey and Francesconi (2018).

known that standard errors are underestimated in difference-in-difference frameworks because of intra-cluster correlation. Common practice for difference-in-difference methods is to use robust standard errors that are clustered at the broadest level. The resulting test-distribution is the student's-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 clusters are treated (in our study, it was only one cluster), the number of clusters is low (fourteen in our study), and the cluster sizes differ to a large extent (Mackinnon and Webb, 2018). Therefore, following Roodman et al. (2018), we calculate wild cluster bootstrap p-values, wild subcluster bootstrap p-values, 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 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 entrance into STEM occupations after graduation, driven mostly by 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 is the conclusion.

2 Related literature

In the following literature review, we focus on two types of contributions. First, we present studies that evaluated the same reform in order to compare our results to theirs. Second, we connect our analysis to the broader literature on 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, as well as in realistic and investigative vocational interests. They further considered which university field of study the individuals chose two years after completing high school. Using only data for Baden Württemberg, the state in which the reform took place, and a before-after comparison, the authors find that gender differences decreased for math achievement but increased for math self-concept and for realistic as well as investigative vocational interest. They did not find a significant effect of the reform on the choice of study subjects at university. An important difference to our study is that they only considered initial university study choices (not the successful completion of degrees) and no labor market outcomes.

Görlitz and Gravert (2016) and Görlitz 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 in high school dropouts which disappears over time among males but is persistent among females. The individuals who did not drop 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 of the reform on higher education enrollment in more detail. According to Görlitz and Gravert (2018), this higher enrolment did not go along with an increase in females entering STEM subjects. Only males exhibited a robust positive effect for STEM. Our study differs from Görlitz and Gravert (2016) and Görlitz and Gravert (2018) in that we use microdata, that we consider not only 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 analyzing the effects of differences in school curricula, especially with regard to math and the natural sciences, on 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 choosing STEM subjects in later high school years is not driven by prior differences in mathematical achievement but that female students require stronger signals of 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 in later grades. However, this effect was restricted to men, and was stronger for pupils from lower social and academic backgrounds. Domina and Saldana (2012) examined the intensification of mathematics curricula in American high schools over the period from 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 and 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 in educational outcomes. He found that math ability (but not verbal ability) is important for selecting 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) studied 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 a few studies have used quasi-experimental variation 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 typically do 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 in 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 in high school increases 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). Conditioning on already having completed a STEM degree, Sassler et al. (2017a) study the transitions of STEM graduates into STEM occupations. They find that the highest share of the variance in 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 share.

Finally, we want to point out a number of studies focusing 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 females' intentions to enroll or not 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 make clear distinctions 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 differences in math 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 of 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 onto *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; however, these are chosen only by a minority of students. The HEEQ from *Gymnasium* generally allows individuals to enroll in all fields of study at university. Specialized and vocational high schools' HEEQ usually allow students to study only a subset of degree programmes related to the secondary school's area of specialization. In Germany, two options for tertiary education 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 in all German states. In this curriculum, all students completed similar courses during the first ten or eleven school years. In 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. Out of the chosen classes, two had to be on an advanced level and several others on a basic level. In Baden-Württemberg, for example, at least one basic math and one basic German class had to be taken, in addition to at least one natural science class. If a math, German or science class at an advanced level was chosen, students could fill their basic courses with other subjects. If they chose their math, German and 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. 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 to earn an HEEQ.

In 1999, Baden-Württemberg announced a reform affecting students starting their second to last school year in a 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 onwards and specialized/vocational high school graduates from 2005 onwards were affected by the reform. The post-reform high school curriculum forced all students to attend a mandatory advanced class in mathematics, German and one foreign language. In addition, two more advanced courses in one natural science and/or another foreign language had to be taken. This means that the total number of mandatory natural science courses increased from one to two (with one of them potentially at the basic level). Because of the larger number of required 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 the reform also had some additional aspects, it is fair to say that its essential ingredient was a significant shift towards more instruction time in math and the natural sciences for a large number of students who previously would not have chosen these subjects at all or would have only chosen them at the basic level (and thus 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 compared to the situation in other federal states. The figures refer to the second qualification phase of the German Abitur at *Gymnasium*. Administrative data is available from 2002 onwards, but a number of missing values and differences in coding make some values before 2003 unusable. As figure 1 shows, advanced math participation in 2004 varied from around

.08 for Niedersachsen 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 no gender-specific administrative data is 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. The highest percentage of females taking non-mandatory advanced math classes was .4, but was typically 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 instruction time for math during the last two years of high school. For over 80 percent of students, instruction time doubled as students who would have enrolled in 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 receiving 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 the natural sciences, the reform had a similarly strong impact 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

We employ a difference-in-difference setup for our estimations with gender interactions in order to obtain gender-specific difference-in-difference effects. This setup compares the situation before and after the reform with the non-treated federal states serving as a counterfactual for the treated state. Our regression model is

$$\begin{aligned} y_{ist} = {} & \alpha + \gamma_1 \cdot \mathit{After}_t + \gamma_2 \cdot \mathit{BaWu}_s + \rho \cdot \mathit{Treatment}_{ist} \\ & + \gamma_3 \cdot \mathit{Female}_{ist} + \gamma_4 \cdot (\mathit{After}_t \cdot \mathit{Female}_{ist}) + \gamma_5 \cdot (\mathit{BaWu}_s \cdot \mathit{Female}_{ist}) \\ & + \lambda \cdot (\mathit{Treatment}_{ist} \cdot \mathit{female}_{ist}) \\ & + X'_{ist} \cdot \beta + \eta_s + \nu_t + \epsilon_{ist} \end{aligned} \tag{1}$$

where the index s indicates in which federal state individual i obtained their higher education entrance qualification (HEEQ) and t denotes the year the individual obtained their 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 of 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 $BaWu_s$ and indicates 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 η_s and ν_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, ρ 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 λ represents the gender difference of the reform effect. Overall, this setup identifies the reform effects ρ and $\rho + \lambda$ by comparing individuals before and after the reform in Baden Württemberg with the situation before and after the reform year in other federal states taken as a counterfactual scenario. There may be general time-constant differences between treatment and control states η_s as well as common time trends in STEM participation ν_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 ρ 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 the 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, there is clustering on both the state and the time dimension. Ignoring intra-cluster dependence will bias standard errors downward and lead to over-rejection rates (Bertrand et al., 2004). Even when correcting for group dependency using cluster-robust standard errors, bias may remain due to finite cluster sizes, unbalanced numbers of treated and untreated 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 approaches infinity (Cameron and Miller, 2015). This assumption is clearly violated in setups such as ours, with just a few clusters and a very small number of treated clusters (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') different cluster sizes. Mackinnon and Webb (2017) introduced the wild subcluster bootstrap and showed that it is superior to other inference procedures in the case of only a 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 comparing restricted and unrestricted standard errors (i.e., with and without imposing the null hypothesis) as a diagnostic test for the validity of standard errors. If the two do not coincide, one can be certain that they are invalid. If they are close, there is probably no problem.

Mackinnon and Webb (2018) present simulation results for a difference-in-difference setup similar to ours with just one treated cluster and varying cluster sizes. 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, the restricted and unrestricted versions of the individual wild bootstrap coincide well. In our empirical application below, we follow this procedure as well as Roodman et al. (2018) by computing conventional $t(G-1)$ clustered standard errors and 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 effects only in a number of cases, whereas the conventional clustered p-values in particular would suggest an extremely high level of statistical significance for all our results.

As explained in the previous section, we include a large set of time-varying covariates at the state level in our difference-in-difference regressions to pick up potentially differential time trends across states. One might hope that this additionally takes out intra-cluster correlations, mitigating problems of cluster inference. Therefore, it is interesting to see how the inclusion of such variables changes cluster inference. We found that including such variables generally did not change the conclusions from cluster inference very much, or if it did, then in unsystematic ways.

5 Data

The data for our study were 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 2005, 2009 and 2013 cohorts.[2] The survey includes rich information on parental background, the individual's higher education entrance qualification, choices during university study, and labor market entry. In our analysis, we exclude individuals with a HEEQ obtained before 1995 for the 2005 cohort, before 1999 for the 2009 cohort, and before 2003 for the 2013 cohort in order to drop unrepresentative long-term students. For the same reason, we exclude individuals born before 1970 in 2005 cohort, before 1974 in the 2009 cohort and before 1978 in the 2013 cohort.

— Table 1 here —

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

— Table 2 here —

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

[2] We use only the first wave of each cohort because subsequent waves suffer from considerable attrition and because only the first wave for the 2013 cohort is available at this point.

degree but count them as STEM if at least one of their degrees is in STEM. Following common practice, we include into STEM all fields in science, technology, engineering and mathematics. More precisely, our *STEM* category includes the 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 particulary pronounced mathematical or technical orientation (i.e. physics, computer science, engineering and mathematics). We also consider the smaller STEM subsets of the *life sciences* (biology and chemistry), *math and physics*, as well as *engineering and computer science*.

For 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 categorization into STEM and non-STEM occupations, but only for the KldB 2010 (Bundesagentur für Arbeit, 2016). For the KldB 1992 codes, we followed a translation from KldB 1992 to KldB 2010. This 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 these occupations were more or less specific in the KldB 1992 classification than in the KldB 2010 classification). For these cases, we used a special procedure, the details of which are available on request. As table 2 shows, around 34 percent of the members of our sample obtained a STEM degree, and for 27 percent this was in PTEM. Note that PTEM includes only physics out of the sciences fields in STEM and thus is smaller by definition. Around 5.3 percent of the observations in the sample had at least one degree in life sciences and 3.97 percent in math or physics, making the latter the smallest group we investigated. Engineering and computer science is the biggest group within STEM with a share of 22.6 percent. We obtained a qualitatively similar picture for the occupational outcomes 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 assumption and in order to minimize the remaining intra-cluster correlation, we included a set of state- and time-specific variables shown in table 3. We include variables from four different areas: economy, 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 focusses 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 conventional clustered p-values (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). Like Mackinnon and Webb (2017) and Roodman et al. (2018), we chose the state-year level as subclusters.

— Table 4 and figure 3 here —

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 had diverging results in their restricted and unrestricted versions, which, according to Mackinnon and Webb (2017) and Roodman et al. (2018), signals their invalidity. However, we also observed that the difference between the restricted and unrestricted values tended to shrink as we moved 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 in specification (3), which includes state and year effects. For specification (4), which also includes the state-level variables, the p-values jump above the 10 percent level, suggesting no statistical significance at conventional levels. Note, however, that adding this 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 of Hübner et al. (2018) who also did not find statistically significant reform effects on university 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. Therefore, we 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 all 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 the life sciences (biology and chemistry). For this category, we obtain a male treatment effect close to zero which is 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 a math or physics degree by 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 for why we obtained no significant effect for the PTEM category, which combines all of these subjects (column (2) of table 5). This shows that it is important to consider heterogeneity within STEM subjects.

The lower half of table 5 shows the gender difference in 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 degrees. 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 that the reform had a positive and statistically significant effect. Again, the negative and marginally significant gender difference in the lower half of table 6 points to a non-effect of the reform on women's occupational choices after graduation. The picture for occupations is clearer in terms of statistical significance than that for degrees, possibly because STEM occupations represent a smaller percentage of all occupations than STEM degrees represent among all degrees.

— Table 6 here —

In table 7, we differentiate between different subfields of STEM occupations. Note that some of these subfields represent very small percentages of jobs in the labor market. The reason is that occupations are harder to map to different subjects in a clear way than degrees. Interestingly, the results present a similar picture as for degrees. There is a positive (but this time slightly more significant) effect for males on PTEM occupations, and a similarly 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; however, these are very small in magnitude.

— Table 7 here —

6.4 Common trend assumption and placebo test

Graphical inspection of our raw data does not suggest obvious violations of the common trend assumption (see figures A1 to A4 in the appendix). Note that figures A1 to A4 do not reflect our multiple controls for time, state and covariate effects. In addition to this graphical check, we report the following placebo test, which defines 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 also include individuals born before 1995 in the placebo analysis (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 among men which is in the opposite direction of the actual treatment effect for this category reported in table 5. This calls into question the result for the original treatment effect, although it is comforting that the placebo effect goes in the opposite direction. A possible reason for the 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 —

The placebo results for STEM occupations and its subcategories 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 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. In addition to some other aspects, the curriculum reform doubled the instruction time and increased the level of instruction in math and the natural sciences for some 80 percent of students and an even higher proportion of the 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 significant 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 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 the natural sciences among 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 observed 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, no 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 also been deterrence effects among men, who received a clearer signal about the nature of the subject after the reform, potentially dissuading some individuals from pursuing math or physics study programs. Another possibility is that these effects are due to the fact that some individuals' math exposure was reduced by the reform (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 those by 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 the inference procedures proposed by Mackinnon and Webb (2018) and Roodman et al. (2018), which have been shown to be the most appropriate so far, drastically reduces the statistical significance of our estimates and suggests a more cautious interpretation.

References

Altonji, J., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college majors, and careers. *Annual Review of Economics 4*, 185–223.

Altonji, J. G. (1995). The Effects of High School Curriculum on Education and Labor Market Outcomes. *Journal of Human Resources 30*, 409–438.

Ardicianono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*, 343–375.

Aughinbaugh, A. (2012). The Effects of High School Math Curriculum on College Attendance: Evidence from the NLSY97. *Economics of Education Review 31*, 861–870.

Berlingieri, F. and U. Zierahn (2014). Field of Study, Qualification Mismatch, and Wages: Does Sorting Matter? *SSRN Electronic Journal*.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-In-Differences Estimates? *Quarterly Journal of Economics 119*, 249–275.

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

Outcomes. *Royal Holloway University of London Discussion Paper Series* (2010 – 1), 1–48.

Bundesagentur für Arbeit (2016). Der Arbeitsmarkt in Deutschland – MINT-Berufe. *Statistik/Arbeitsmarktberichterstattung 2016*, 1–37.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review 107*, 125–130.

Cameron, A. C. and D. L. Miller (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources 50*, 317–372.

Card, D. and A. Payne (2017). High school choices and the gender gap in stem. *NBER Working Paper No. 23769*, 1–42.

Carrell, S., M. Page, and J. West (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics 125*, 1102–1124.

Cech, E., B. Rubineau, S. Silbey, and C. Seron (2011). Professional Role Confidence and Gendered Persistence in Engineering. *American Sociological Review 76*, 641–666.

Ceci, S. and W. Williams (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science 19*, 275–279.

Cheryan, S. (2012). Understanding the Paradox in Math-Related Fields: Why Do Some Gender Gaps Remain While Others Do Not? *Sex Roles 66*, 184–190.

Corell, S. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology 106*, 1691–1730.

Cortes, K., J. Goodman, and T. Nomi (2012). Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra. *Journal of Human Resources 50*, 108–158.

Danbold, F. and Y. J. Huo (2017). Men's Defense of their Prototypicality Undermines the Success of Women in STEM Initiatives. *Journal of Experimental Social Psychology 72*, 57–66.

Daymont, T. and P. Andrisani (1984). Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources 19*, 408–428.

De Philippis, M. (2017). Stem graduates and secondary school curriculum: does early exposure to science matter? *Banca D'Italia Working Paper* (1107), 1–51.

DelCarpio, L. and M. Guadalupe (2018). More Women in Tech? Evidence from a Field Experiment Addressing Social Identity. *IZA Discussion Paper* (11876), 1 – 49.

Domina, R. and J. Saldana (2012). Does Raising the Bar Level the Playing Field? Mathematics Curricular Intensification and Inequality in American High Schools, 1982-2004. *Economics of Education Review 49*, 685–798.

Ehrenberg, R. G. (2010). Analyzing the factors that influence persistence rates in STEM field, majors: Introduction to the symposium. *Economics of Education Review 29*, 888–891.

Franceschini, G., S. Galli, F. Chiesi, and C. Primi (2014). Implicit gender–math stereotype and women's susceptibility to stereotype threat and stereotype lift. *Learning and Individual Differences 32*, 273–277.

Friedman-Sokuler, N. and M. Justman (2016). Gender Streaming and Prior Achievement in High School Science and Mathematics. *Economics of Education Review 53*, 230–253.

Friedman-Sokuler, N. and M. Justman (2018). Gender Equality and Women in STEM: the Curious Case of Arab Women in Israel. *Working Paper*, 1–46.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics 118*, 1049–1074.

Goodman, J. (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics* (forthcoming).

Görlitz, K. and C. Gravert (2016). The Effects of Increasing the Standards of the High School Curriculum on School Dropout. *Applied Economics 48*, 5314–5328.

Görlitz, K. and C. Gravert (2018). The Effects of a High School Curriculum Reform on University Enrollment and the Choice of College Major. *Education Economics 26*, 321–336.

Griffith, A. L. (2010). Persistence of Women and Minorities in STEM Field Majors: Is it the School that Matters? *Economics of Education Review 29*, 911–922.

Guiso, L., F. Monte, P. Spienza, and L. Zingales (2008). Culture, gender, and math. *Science 320*, 1164–1165.

Hübner, N., E. Wille, J. Cambria, K. Oschatz, B. Nagengast, and U. Trautwein (2018). Maximizing Gender Equality by Minimizing Course Choice Options? Effects of Obligatory Coursework in Math on Gender Differences in STEM. *Journal of Educational Psychology 109*(7), 993–1009.

Jia, N. (2016). Do Stricter High School Math Requirements Raise College STEM Attainment? *Working paper 8983, Mimeo*.

Joensen, J. and H. Nielsen (2016). Mathematics and gender: Heterogeneity in causes and consequences. *Economic Journal 126*, 1129–1163.

Joensen, J. S. and H. S. Nielsen (2009). Is there a Causal Effect of High School Math on Labor Market Outcomes? *Journal of Human Resources 44*, 171–198.

Justman, M. and S. J. Méndez (2018). Gendered choices of STEM subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of Education Review 64*, 282–297.

Kahn, S. and D. Ginther (2018). Women and Science, Technology, Engineering, and Mathematics (STEM): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family? In S. Averett, L. Argys, and L. Hoffmann (Eds.), *Oxford Handbook on the Economics of Women*, Chapter 31. Oxford: Oxford University Press.

Kokkelenberg, E. C. and E. Sinha (2010). Who succeeds in STEM studies? An analysis of Binghamton University undergraduate students. *Economics of Education Review 29*, 935–946.

Levine, P. B. and D. J. Zimmerman (1995). The Benefit of Additional High-School Math and Science Classes for Young Men and Women. *Journal of Business & Economic Statistics 13*, 137–149.

Mackinnon, J. G. and M. D. Webb (2017). Pitfalls when Estimating Treatment Effects Using Clustered Data. *Political Methodologist 24*, 20 – 31.

Mackinnon, J. G. and M. D. Webb (2018). The Wild Bootstrap for Few (Treated) Clusters. *Econometrics Journal 20*, 1–22.

Morgan, S., D. Gelbgiser, and K. Weeden (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research 42*, 989–1005.

OECD (2007). Women in Science, Engineering and Technology (SET): Strategies for a Global Workforce.

OECD (2010). OECD Information Technology Outlook.

Parey, A. and M. Francesconi (2018). Early Gender Gaps Among University Graduates. *European Economic Review 109*, 63–82.

Roodman, D., M. Ø. Nielsen, M. D. Webb, and J. G. Mackinnon (2018). Fast and Wild: Bootstrap Inference in Stata using boottest. *Working paper* (1406), 1–48.

Rose, H. and J. R. Betts (2004). The Effect of High School Courses on Earnings. *Review of Economics and Statistics 86*, 497–513.

Sassler, S., J. Glass, Y. Levitte, and K. Michelmore (2017a). The missing women in stem? assessing gender differentials in the factors associated with transition to first jobs. *Social Science Research 63*, 192–208.

Sassler, S., J. Glass, Y. Levitte, and K. M. Michelmore (2017b). The Missing Women in STEM? Assessing Gender Differentials in the Factors Associated with Transition to First Jobs. *Social Science Research 63*, 192–208.

Schavan, A. (1999). Stellungnahme des Ministeriums für Kultur, Jugend und Sport. pp. 1–7.

Shi, Y. (2018). The puzzle of missing female engineers: Academic preparation, ability beliefs, and preferences. *Economic of Education Review 64*, 129–143.

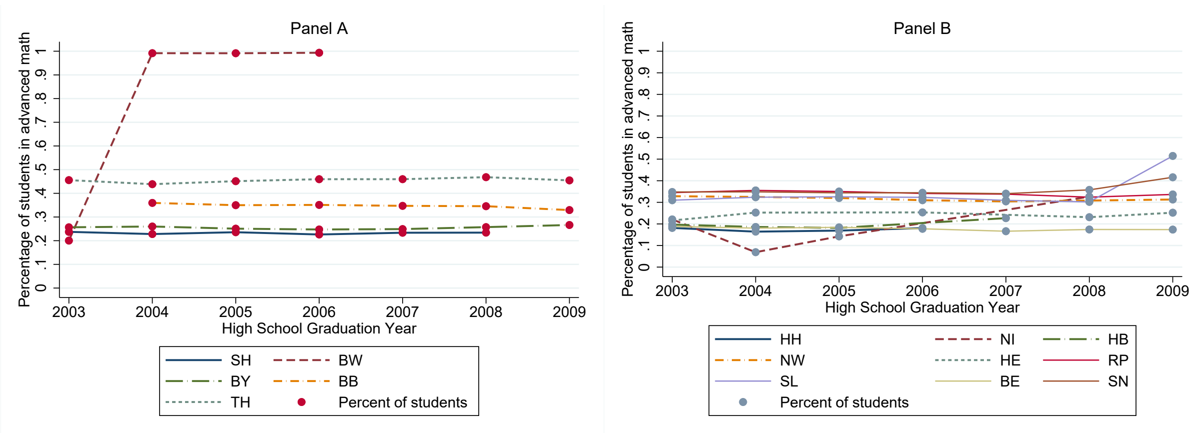
Spitz-Oener, A. and K. Priesack (2018). STEM Occupations and the Evolution of the German Wage Structure. *Mimeo*.

Wiswall, M. and B. Zafar (2015). Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies 82*, 791–824.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*, 546–593.

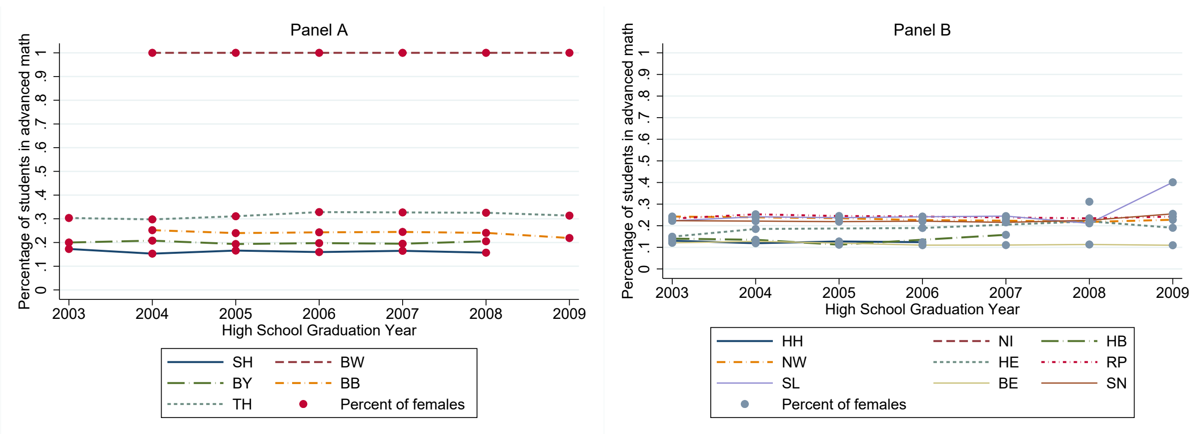
Figures

Figure 1: Students taking advanced math per state



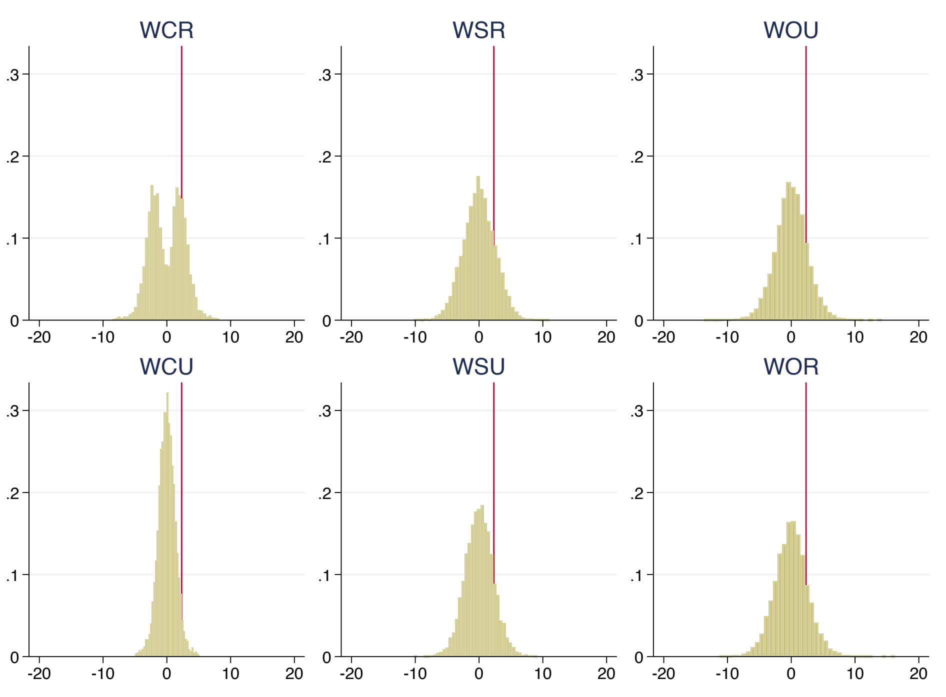
Source: Statistical Offices of the Federal States. *Notes:* Panel A includes the states Brandenburg (BB), Baden-Württemberg (BW), Bayern (BA), Mecklenburg-Vorpommern (MV), Schleswig-Holstein (SH), Sachsen-Anhalt (ST) and Thüringen (TH). Panel B shows the development for Saarland (SL) and Sachsen (SN), 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.

Figure 2: Female students taking advanced math per state



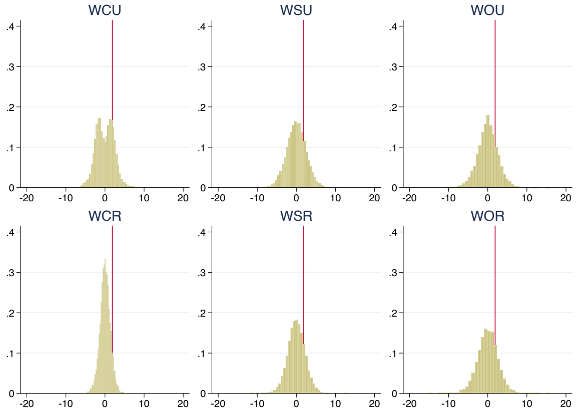
Source: Statistical Offices of the Federal States. *Notes:* Some states did not provide gender specific information for all years, which is why there are some missing years.

Figure 3: Bootstrap p-value distribution for STEM degrees

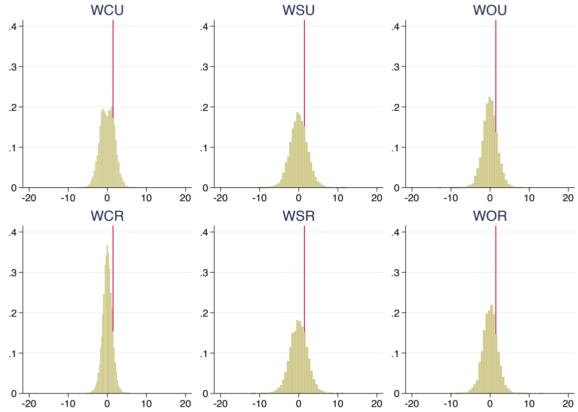


Notes: The plots show the bootstrapped t-statistics based on the particular method. The vertical line displays the t-statistic of the regression. See text for more explanations.

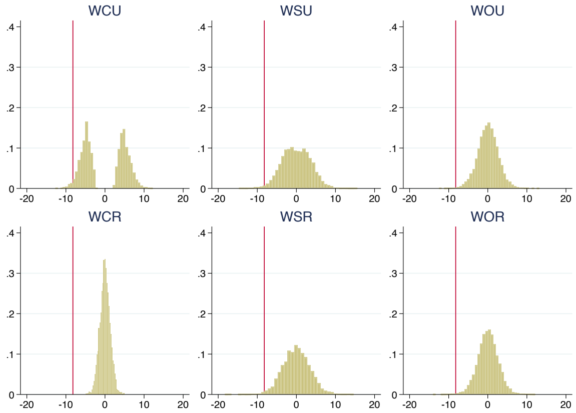
Figure 4: Bootstrap p-value distribution for subfield degrees



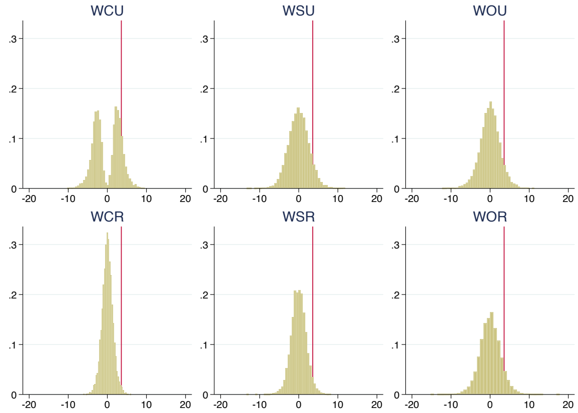
(a) PTEM



(b) Life sciences



(c) Math and physics



(d) Engineering and computer science

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.

Tables

Table 1: General sample information

	Obs	*Mean*	*SD*	*Min*	*Max*
Cohort	21,633	2008	3.1420	2005	2013
Year of HEEQ	21,633	2002	3.7460	1995	2009
Treatment period	21,633	.3475	0.4762	0	1
HEEQ from BaWu	21,633	.1664	0.3725	0	1
Treated individuals	21,633	.0587	0.2351	0	1
Female	21,633	.5366	0.4987	0	1
Treated females	21,633	.0349	0.1834	0	1
Birthyear	21,633	1982	3.9185	1970	1992
Parental education: PhD, Uni or Uni Appl. Sc.	21,633	.0152	0.1225	0	1
Parental education: HS Diploma	21,633	.3802	0.4854	0	1
Parental education: Vocational training	21,633	.0509	0.2197	0	1
Parental education: Other	21,633	.5537	0.4971	0	1
Parental occupation: White collar	21,633	.9427	0.2323	0	1
Parental occupation: Blue collar and other	21,633	.0573	0.2323	0	1

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

	Obs	Mean	SD	Min	Max
Degree outcomes					
At least one degree in STEM	21,633	.3455	0.4755	0	1
At least one degree in PTEM	21,633	.267	0.4424	0	1
At least one degree life sciences	21,633	.0557	0.2294	0	1
At least one degree math or physics	21,633	.0397	0.1953	0	1
At least one degree engineering or computer science	21,633	.2257	0.4180	0	1
Occupational outcomes					
Last or current occupation after graduation in STEM	16,009	0.2459	0.4237	0	1
Last or current occupation after graduation in PTEM	16,009	0.2275	0.4118	0	1
Last or current occupation after graduation life sciences	16,009	0.0184	0.1321	0	1
Last or current occupation after graduation math or physics	16,009	0.0063	0.0766	0	1
Last or current occupation after graduation engineering or computer science	16,009	0.2212	0.4077	0	1

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

	Obs	Mean	SD	Min	Med	Max
Economics						
Unemployment rate (OECD)	21,633	0.0461	0.0222	0.0220	0.0387	0.1129
Registered unemployment by gender	21,633	9.3585	3.8875	3.7000	8.4000	21.5000
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 FDP in percent	21,633	5.8916	3.0715	1.1000	5.9000	16.2000
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

Source: Federal Statistical Office of Germany.

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

Outcome: At least one degree in STEM	(1)		(2)		(3)		(4)	
Male treatment effect	0.0601		0.0618		0.0739		0.1069	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.11	-	0.09	-	0.01	-	0.30
Wild cluster bootstrap	25.13	0.00	21.02	0.00	16.22	0.00	32.53	2.30
Wild subcluster bootstrap	7.41	5.71	6.81	5.91	3.10	1.70	14.81	6.91
Ordinary wild bootstrap	16.02	16.42	14.71	12.51	8.01	7.71	15.12	16.22
Gender Difference	-0.0819		-0.0844		-0.0881		-0.0891	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.02	-	0.01	-	0.00	-	0.00
Wild cluster bootstrap	25.13	0.00	24.92	0.00	25.83	0.00	22.72	0.00
Wild subcluster bootstrap	4.00	1.80	1.60	1.10	1.30	0.50	1.70	0.50
Ordinary wild bootstrap	11.71	9.31	7.11	7.81	6.31	5.31	5.71	5.71
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.0776		0.0865		0.0955		0.099	

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 5: Regression results for subfield degrees and bootstraped p-values

Subfields	(2) PTEM		(3) Life Sciences		(4) Math & Physics		(5) Engineering & Computer Sciences	
Male treatment effect	0.0951		0.0054		-0.0600		0.1580	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	1.00	-	31.90	-	0.00	-	0.07
Wild cluster bootstrap	31.73	3.50	64.86	61.26	8.21	0.00	24.72	0.30
Wild subcluster bootstrap	25.93	16.32	75.58	77.68	5.41	5.71	11.41	1.50
Ordinary wild bootstrap	23.12	24.12	71.27	74.07	2.40	1.10	9.81	10.01
Gender Difference	-0.0723		-0.0073		0.0114		-0.0845	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.05	-	22.97	-	26.78	-	0.04
Wild cluster bootstrap	19.82	0.00	53.75	41.34	56.86	49.35	32.83	0.40
Wild subcluster bootstrap	4.50	3.50	66.77	68.27	83.68	84.58	7.31	4.40
Ordinary wild bootstrap	10.21	14.41	66.37	64.96	75.88	75.18	11.21	11.41
Set of covariates								
All variables included in specification (4)	Yes		Yes		Yes		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.1259		0.0237		0.014		0.1345	

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

Outcome: Occupation in STEM	(1)		(2)		(3)		(4)	
Male treatment effect	0.1022		0.1032		0.1131		0.1459	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.24	-	0.09	-	0.02	-	0.00
Wild cluster bootstrap	36.94	4.60	38.24	1.30	35.14	0.70	21.52	0.00
Wild subcluster bootstrap	33.73	26.13	26.93	23.02	29.93	22.02	17.52	8.41
Ordinary wild bootstrap	27.23	27.53	19.42	18.12	15.02	14.51	5.71	4.20
Gender Difference	-0.1131		-0.1116		-0.1181		-0.1151	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
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
Wild subcluster bootstrap	18.52	16.42	18.82	14.71	21.62	17.22	21.12	19.12
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		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		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

Subfields	(2) PTEM		(3) Life Sciences		(4) Math & Physics		(5) Engineering & Computer Sciences	
Male treatment effect	0.1088		0.0371		-0.0178		0.1266	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.05	-	0.01	-	0.02	-	0.01
Wild cluster bootstrap	24.22	0.20	23.02	0.00	26.23	0.20	18.92	0.00
Wild subcluster bootstrap	37.64	22.22	11.61	5.31	5.01	2.10	24.12	15.62
Ordinary wild bootstrap	13.31	12.61	1.40	0.90	5.31	4.50	7.91	6.31
Gender Difference	-0.1003		-0.0148		0.0091		-0.1094	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.00	-	0.95	-	0.08	-	0.00
Wild cluster bootstrap	32.63	0.00	36.04	8.61	29.73	0.00	23.52	0.00
Wild subcluster bootstrap	24.92	20.12	30.33	22.62	15.62	13.51	19.42	15.82
Ordinary wild bootstrap	9.41	8.71	24.42	23.82	14.71	14.91	6.31	6.41
Set of covariates								
All variables included in specification 4	Yes		Yes		Yes		Yes	
Additional regression information								
Number of observations	16009		16009		16009		16009	
R^2	0.1182		0.0106		0.0087		0.116	

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 8: Placebo test for STEM degrees and subfields

	(1) STEM		(2) PTEM		(3) Life Sciences		(4) Math & Physics		(5) Engineering & Computer Sciences	
Male treatment effect	-0.0367		-0.0298		0.0151		0.0725		-0.1052	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	16.32	-	18.83	-	13.41	-	0.00	-	0.37
Wild cluster bootstrap	54.85	25.93	49.75	29.03	36.94	17.42	8.31	0.00	37.44	1.60
Wild subcluster bootstrap	52.45	45.05	55.86	53.45	41.94	36.54	1.00	0.80	15.22	2.60
Ordinary wild bootstrap	57.86	61.36	64.06	62.06	46.65	47.45	2.00	1.90	13.61	15.72
Gender Difference	0.024		0.0543		-0.0208		-0.0084		0.0691	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	27.52	-	5.95	-	2.62	-	29.24	-	3.33
Wild cluster bootstrap	51.95	43.54	39.24	6.01	33.83	2.90	60.96	48.85	42.74	3.40
Wild subcluster bootstrap	74.07	74.47	42.94	42.04	23.42	17.42	87.89	88.59	24.02	14.71
Ordinary wild bootstrap	74.57	76.28	45.75	45.05	29.23	30.13	82.98	81.58	37.94	35.24
Set of covariates										
All variables included in specification 4	Yes		Yes		Yes		Yes		Yes	
Additional regression information										
Number of observations	13285		13285		13285		13285		13285	
R^2	0.1		0.1275		0.0262		0.023		0.1306	

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).

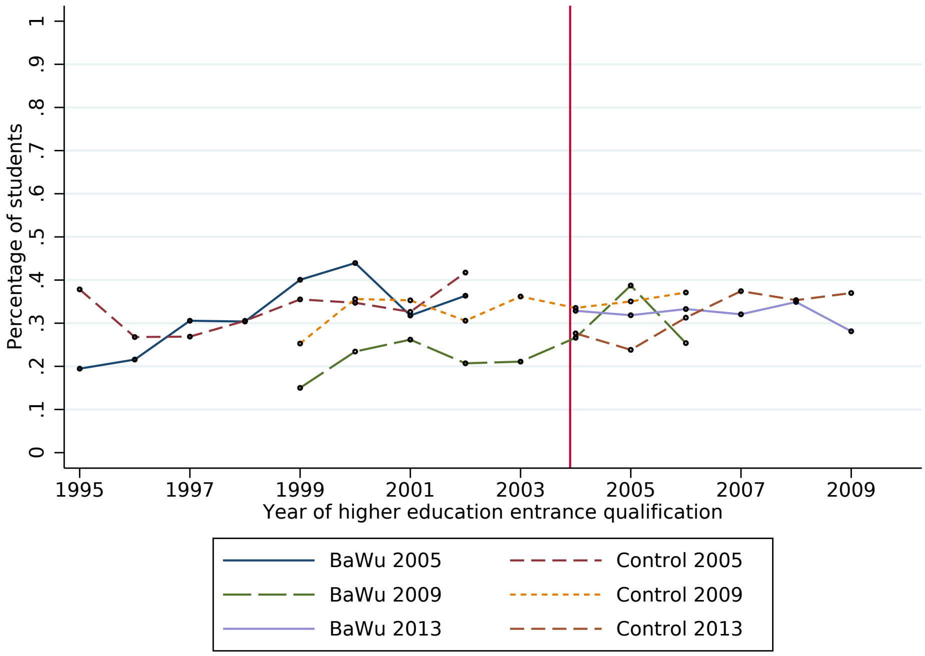
Table 9: Placebo test for STEM occupations and subfields

	(1) STEM		(2) PTEM		(3) Life Sciences		(4) Math & Physics		(5) Engineering & Computer Sciences	
Male treatment effect	-0.0654		-0.0556		-0.0098		0.0068		-0.0624	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	3.72	-	7.80	-	5.19	-	13.19	-	7.47
Wild cluster bootstrap	55.26	12.61	53.85	19.72	27.23	4.90	36.94	24.02	52.45	18.42
Wild subcluster bootstrap	35.54	25.43	41.64	33.63	20.02	15.72	48.15	50.15	36.64	29.03
Ordinary wild bootstrap	37.34	39.24	48.45	50.85	34.33	33.53	62.16	61.86	46.15	46.75
Gender Difference	0.0368		0.0166		0.0202		-0.0167		0.0334	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	12.98	-	32.72	-	1.53	-	0.00	-	20.94
Wild cluster bootstrap	44.74	20.62	73.07	71.27	32.53	4.80	24.82	0.00	48.75	38.64
Wild subcluster bootstrap	55.86	56.96	83.18	83.68	19.62	17.82	10.61	6.61	63.86	57.86
Ordinary wild bootstrap	60.86	58.96	86.29	87.39	19.72	19.62	14.51	14.41	69.77	71.87
Set of covariates										
All variables included in specification 4	Yes		Yes		Yes		Yes		Yes	
Additional regression information										
Number of observations	11089		11089		11089		11089		11089	
R^2	0.097		0.101		0.0152		0.0106		0.0983	

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).

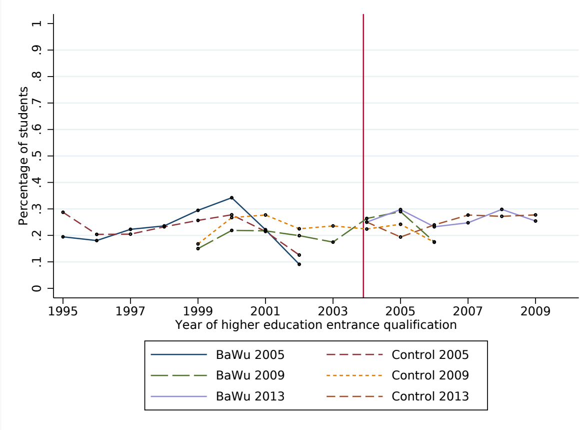
Appendix

Figure A1: Trends of STEM degrees

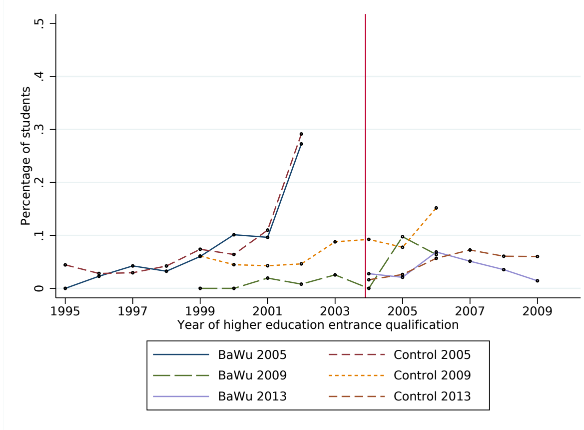


Notes: The figure shows the STEM degree percentages per HEEQ year by cohort. The solid and the long-dash lines show the means for the treatment state Baden-Württemberg (BaWu). The dash and short-dash lines show the same for the control states. The vertical red line groups the figure into before and after treatment.

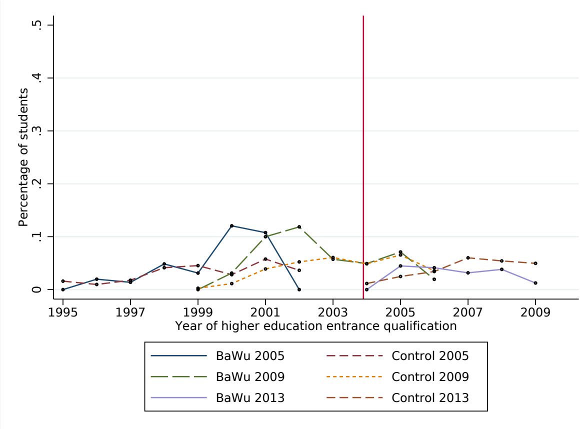
Figure A2: Trends of subfield degrees



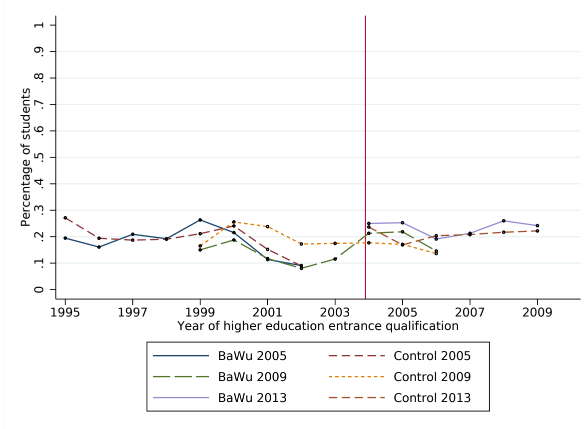
(a) PTEM



(b) Life sciences



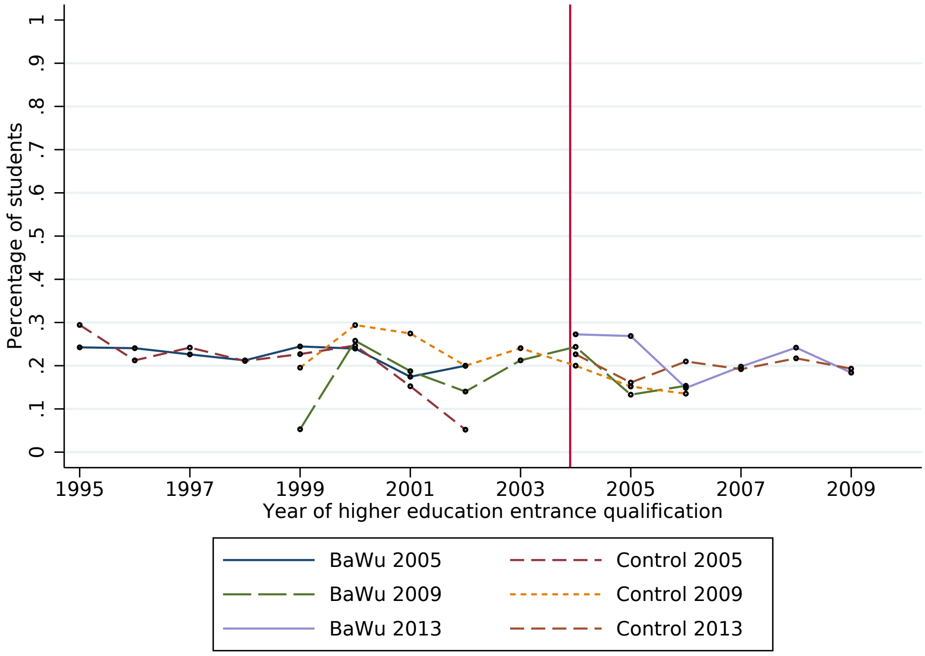
(c) Math and physics



(d) Engineering and computer science

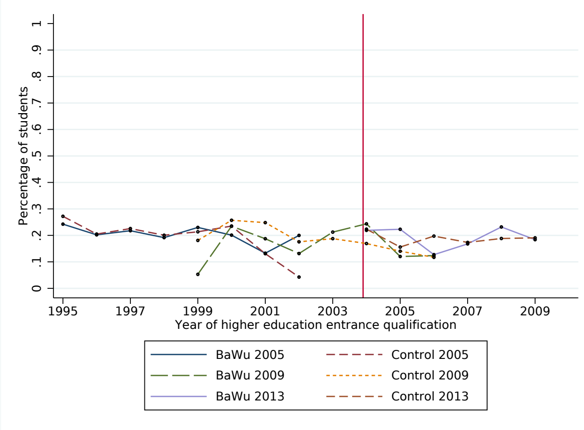
Notes: The figure shows the STEM subfield degree percentages per HEEQ year by cohort. The solid and the long-dash lines show the means for the treatment state Baden-Württemberg (BaWu). The dash and short-dash lines show the same for the control states. The vertical red line groups the figure into before and after treatment.

Figure A3: Trends of STEM occupations

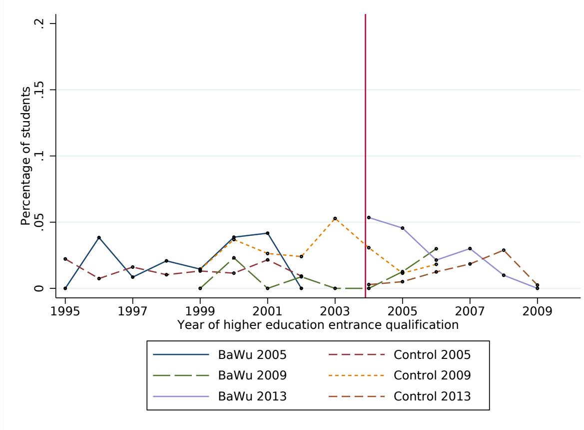


Notes: The figure shows the STEM occupation percentages per HEEQ year by cohort. The solid and the long-dash lines show the means for the treatment state Baden-Württemberg (BaWu). The dash and short-dash lines show the same for the control states. The vertical red line groups the figure into before and after treatment.

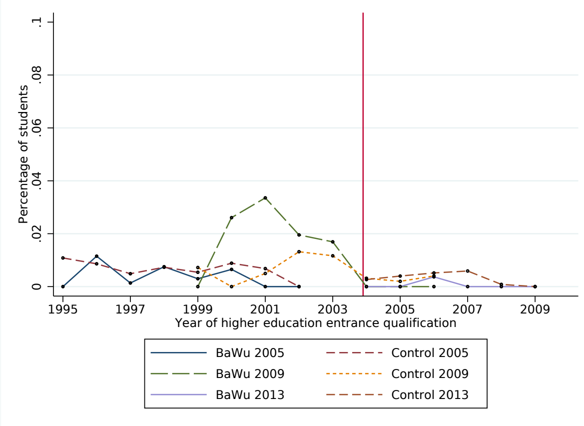
Figure A4: Trends of subfield occupations



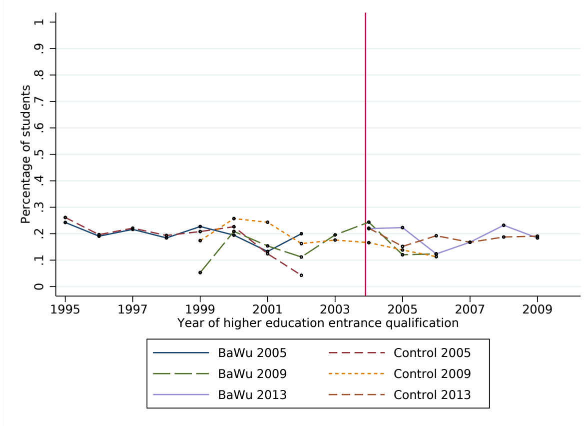
(a) PTEM



(b) Life sciences



(c) Math and physics



(d) Engineering and computer science

Notes: The figure shows the STEM subfield occupation percentages per HEEQ year by cohort. The solid and the long-dash lines show the means for the treatment state Baden-Württemberg (BaWu). The dash and short-dash lines show the same for the control states. The vertical red line groups the figure into before and after treatment.

Table A1: Regression results for PTEM degree and bootstrapped p-values

Outcome: At least one degree in PTEM	(1)		(2)		(3)		(4)	
Male treatment effect	0.0717		0.0733		0.08		0.0951	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.12	-	0.15	-	0.04	-	1.00
Wild cluster bootstrap	26.83	0.90	24.92	0.60	17.62	0.00	31.73	3.50
Wild subcluster bootstrap	10.91	6.41	11.71	7.01	8.01	4.20	25.93	16.32
Ordinary wild bootstrap	16.82	17.42	18.92	17.42	11.31	12.41	23.12	24.12
Gender Difference	-0.0685		-0.0699		-0.0731		-0.0723	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.07	-	0.05	-	0.02	-	0.05
Wild cluster bootstrap	23.02	0.00	22.22	0.00	18.72	0.00	19.82	0.00
Wild subcluster bootstrap	7.61	4.60	4.60	3.80	3.20	2.30	4.50	3.50
Ordinary wild bootstrap	15.32	13.51	12.01	15.02	9.41	12.01	10.21	14.41
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.1084		0.1126		0.123		0.1259	

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

Outcome: At least one degree in Life Sciences	(1)		(2)		(3)		(4)	
Male treatment effect	-0.0101		-0.011		-0.0076		0.0054	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	2.68	-	4.27	-	11.39	-	31.90
Wild cluster bootstrap	32.63	1.80	38.74	6.01	39.04	14.71	64.86	61.26
Wild subcluster bootstrap	29.73	30.13	35.84	41.74	53.45	55.96	75.58	77.68
Ordinary wild bootstrap	26.83	26.03	35.64	34.33	48.35	46.55	71.27	74.07
Gender Difference	-0.0042		-0.005		-0.0054		-0.0073	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	34.93	-	32.12	-	28.61	-	22.97
Wild cluster bootstrap	69.87	62.66	63.86	54.55	59.06	49.95	53.75	41.34
Wild subcluster bootstrap	83.28	84.28	77.68	78.28	73.47	75.28	66.77	68.27
Ordinary wild bootstrap	83.88	85.09	78.28	77.88	75.28	70.67	66.37	64.96
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.0038		0.013		0.0205		0.0237	

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

Outcome: At least one degree in Math or Physics	(1)		(2)		(3)		(4)	
Male treatment effect	-0.0405		-0.0397		-0.0409		-0.06	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.00	-	0.01	-	0.00	-	0.00
Wild cluster bootstrap	23.52	0.00	26.13	0.00	26.23	0.00	8.21	0.00
Wild subcluster bootstrap	16.52	12.71	17.52	14.71	16.52	14.51	5.41	5.71
Ordinary wild bootstrap	7.81	7.81	9.01	8.91	6.31	6.41	2.40	1.10
Gender Difference	0.0113		0.0107		0.0112		0.0114	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	25.02	-	26.66	-	26.06	-	26.78
Wild cluster bootstrap	52.25	48.15	57.56	49.55	55.06	47.95	56.86	49.35
Wild subcluster bootstrap	80.38	82.48	84.68	83.98	82.28	83.68	83.68	84.58
Ordinary wild bootstrap	74.27	72.87	75.18	76.98	74.27	74.47	75.88	75.18
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.0027		0.0077		0.0119		0.014	

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 bootstraped p-values

Outcome: At least one degree in Engineering or Computer Science	(1)		(2)		(3)		(4)	
Male treatment effect	0.115		0.1157		0.1237		0.158	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.01	-	0.01	-	0.00	-	0.07
Wild cluster bootstrap	24.02	0.40	21.02	0.00	18.92	0.00	24.72	0.30
Wild subcluster bootstrap	7.51	3.90	6.61	4.00	7.41	2.20	11.41	1.50
Ordinary wild bootstrap	7.41	7.41	8.81	9.11	4.90	5.11	9.81	10.01
Gender Difference	-0.0812		-0.0819		-0.0851		-0.0845	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.02	-	0.01	-	0.01	-	0.04
Wild cluster bootstrap	30.73	0.40	28.63	0.00	30.03	0.20	32.83	0.40
Wild subcluster bootstrap	5.11	4.20	5.51	3.80	5.31	3.00	7.31	4.40
Ordinary wild bootstrap	9.81	11.11	6.81	9.11	8.31	9.31	11.21	11.41
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	21633		21633		21633		21633	
R^2	0.1165		0.1183		0.1315		0.1345	

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

Outcome: Occupation in PTEM	(1)		(2)		(3)		(4)	
Male treatment effect	0.0911		0.0907		0.1004		0.1088	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.19	-	0.09	-	0.02	-	0.05
Wild cluster bootstrap	37.04	1.00	32.13	0.30	29.73	0.00	24.22	0.20
Wild subcluster bootstrap	31.53	31.23	26.43	23.62	29.63	28.23	37.64	22.22
Ordinary wild bootstrap	23.82	23.52	18.62	19.32	12.41	11.21	13.31	12.61
Gender Difference	-0.0989		-0.0977		-0.1047		-0.1003	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.01	-	0.01	-	0.00	-	0.00
Wild cluster bootstrap	25.63	0.00	23.92	0.00	24.32	0.00	32.63	0.00
Wild subcluster bootstrap	21.62	20.02	22.02	19.82	26.53	19.32	24.92	20.12
Ordinary wild bootstrap	11.21	11.91	10.91	11.81	8.61	8.81	9.41	8.71
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	16009		16009		16009		16009	
R^2	0.0959		0.0986		0.1148		0.1182	

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

Outcome: Occupation in Life Sciences	(1)		(2)		(3)		(4)	
Male treatment effect	0.0111		0.0124		0.0127		0.0371	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	7.62	-	5.05	-	4.63	-	0.01
Wild cluster bootstrap	41.24	24.22	38.84	20.62	35.64	22.32	23.02	0.00
Wild subcluster bootstrap	54.05	48.85	53.35	45.75	55.46	51.95	11.61	5.31
Ordinary wild bootstrap	43.24	44.34	36.74	38.24	37.04	34.63	1.40	0.90
Gender Difference	-0.0142		-0.0138		-0.0134		-0.0148	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.64	-	0.87	-	1.03	-	0.95
Wild cluster bootstrap	34.83	3.80	35.74	4.90	38.64	7.31	36.04	8.61
Wild subcluster bootstrap	22.72	21.22	26.93	24.22	26.93	24.12	30.33	22.62
Ordinary wild bootstrap	21.12	16.62	22.72	22.62	23.62	24.12	24.42	23.82
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	16009		16009		16009		16009	
R^2	5e-04		0.0028		0.0071		0.0106	

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

Outcome: Occupation in Math or Physics	(1)		(2)		(3)		(4)	
Male treatment effect	-0.0132		-0.0132		-0.0129		-0.0178	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.01	-	0.02	-	0.02	-	0.02
Wild cluster bootstrap	33.03	0.40	31.63	1.30	32.63	0.60	26.23	0.20
Wild subcluster bootstrap	13.91	10.31	11.71	8.41	10.11	5.71	5.01	2.10
Ordinary wild bootstrap	9.81	10.51	10.41	12.41	9.31	9.11	5.31	4.50
Gender Difference	0.0082		0.0083		0.0084		0.0091	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.12	-	0.11	-	0.14	-	0.08
Wild cluster bootstrap	28.33	0.60	27.53	0.70	27.13	0.20	29.73	0.00
Wild subcluster bootstrap	22.32	20.32	20.32	18.92	18.82	19.42	15.62	13.51
Ordinary wild bootstrap	18.52	19.62	16.02	17.52	17.72	18.32	14.71	14.91
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	16009		16009		16009		16009	
R^2	0.0036		0.0052		0.0074		0.0087	

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

Outcome: Occupation in Engineering or Computer Sciences	(1)		(2)		(3)		(4)	
Male treatment effect	0.1043		0.1039		0.1134		0.1266	
P-values for male treatment effect	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.03	-	0.01	-	0.00	-	0.01
Wild cluster bootstrap	33.73	0.30	30.53	0.00	28.83	0.00	18.92	0.00
Wild subcluster bootstrap	23.52	21.02	17.62	13.01	20.72	16.52	24.12	15.62
Ordinary wild bootstrap	14.41	14.11	10.91	9.01	5.81	6.31	7.91	6.31
Gender Difference	-0.1071		-0.106		-0.1131		-0.1094	
P-values for gender difference	Res.	Unres.	Res.	Unres.	Res.	Unres.	Res.	Unres.
t(G-1)	-	0.00	-	0.00	-	0.00	-	0.00
Wild cluster bootstrap	23.02	0.00	21.62	0.00	24.32	0.00	23.52	0.00
Wild subcluster bootstrap	16.32	14.41	14.91	12.41	19.52	12.11	19.42	15.82
Ordinary wild bootstrap	8.31	6.61	6.61	6.81	6.11	5.41	6.31	6.41
Set of covariates								
Age, Cohort, Parents	No		Yes		Yes		Yes	
State and Year Fixed Effects	No		No		Yes		Yes	
State Variables	No		No		No		Yes	
Additional regression information								
Number of observations	16009		16009		16009		16009	
R^2	0.0925		0.0953		0.1125		0.116	

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).