

Peter, Frauke H.; Zambre, Vaishali

Article — Accepted Manuscript (Postprint)

Intended college enrollment and educational inequality: Do students lack information?

Economics of Education Review

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Peter, Frauke H.; Zambre, Vaishali (2017) : Intended college enrollment and educational inequality: Do students lack information?, Economics of Education Review, ISSN 0272-7757, Elsevier, Amsterdam, Vol. 60, pp. 125-141, <https://doi.org/10.1016/j.econedurev.2017.08.002>, <http://www.sciencedirect.com/science/article/pii/S0272775716304113>

This Version is available at:

<https://hdl.handle.net/10419/204655>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Intended College Enrollment and Educational Inequality: Do Students Lack Information?[☆]

Frauke H. Peter, Vaishali Zambre*

August 4, 2017

Abstract

Despite increasing access to university education, students from disadvantaged or non-academic family backgrounds are still underrepresented in universities. In this regard, the economics literature has focused on the role of financial constraints as a cause of these observed differences in educational choices. Our knowledge of potential effects of other constraints regarding university education is more limited. We investigate the causal relationship between information and educational expectations using data from a German randomized controlled trial in which students in high schools were given information on the benefits of as well as on different funding possibilities for university education. We find that the provision of information increases intended college enrollment for students from a non-academic family background, both two to three months and one year after the intervention. In contrast, it leads students from academic backgrounds to lower their enrollment intentions in the short run. However, this effect does not persist. Our results suggest that educational inequality can be reduced by providing students from non-academic families with relevant information.

Keywords: Randomized Controlled Trial, Information Deficit, Educational Inequality, Educational Expectation, College Enrollment

JEL: I21, I24, J24

[☆]We are grateful to the editor and two anonymous referees for helpful feedback and suggestions to improve the manuscript. We especially thank our colleagues from the Best Up project team at DIW Berlin: C. Katharina Spieß, Johanna Storck, and Mathias Huebener; and at WZB: Heike Solga, Alessandra Rusconi, Claudia Finger, and Martin Ehlert. Moreover, we thank Susan Dynarski, Brian McCall and Astrid Würtz Rasmussen as well as participants of the 6th IWAEE conference, the 4th SOLE/EALE world conference, the 2015 EEA Annual Congress and the 2016 AEA Annual Meeting for valuable comments. We gratefully acknowledge funding from the Einstein Foundation Berlin (A-2010-025 (FU)). The usual disclaimer applies.

*Corresponding author

Email addresses: fpeter@diw.de (Frauke H. Peter), vzambre@diw.de (Vaishali Zambre)

1. Introduction

Around the world, post-secondary educational decisions are consistently related to individuals' socio-economic background. In Germany, the probability of starting university education is 37 percent for students with a university entrance qualification from non-academic backgrounds,¹ but the probability is 84 percent for students with a university entrance qualification from academic backgrounds (Middendorff et al., 2013). The economics literature has focused on the role of financial constraints as a cause of these observed differences in educational choices. This focus stems partly from the fact that most studies are based on English-speaking countries where tuition fees present a high financial burden. In countries like Germany, however, university education is free of charge² and the government provides means-tested financial support to finance living expenses. Thus, financial constraints are less likely to explain the observed differences in enrollment rates. The results of Steiner and Wrohlich (2012) support this argument, as they find only a small elasticity of student aid (BAföG) on participation in tertiary education in Germany.³

A relatively understudied explanation for the differing decisions to enroll in college for students from different socio-economic background is a potential lack of information. Given that educational choices are usually modeled as the result of cost-benefit considerations, it is essential that students know about costs and benefits of university education and how they compare to the alternatives. Since the odds of success and the returns to education are uncertain, students must base their decisions on the expectations they form using the information available to them at the time. These expectations are, in turn, shaped by the socio-economic environment of students (Manski, 1993a,b; Oxoby, 2008; Bifulco et al., 2014). Consequently, expectations and information sets may differ by students' educational backgrounds. Heterogeneous information sets at the time of the decision making may explain why students from different educational backgrounds arrive at different educational choices. Thus, directly providing information may help students to make a more informed and background independent decision.

This paper investigates how students' intended college enrollment changes as a result of expanding their information set. We use data from a randomized controlled trial in Germany in which high school students were provided with information about the benefits and funding possibilities of university education one year prior to their graduation exams. During this in-class information intervention, labor market benefits of university education were compared to vocational education. The presentation was given using a standardized script in order to ensure that information was consistently presented across the random sample of high schools.

A growing number of studies investigate the relationship between information and educational choices based on field experiments. Some studies provide information about costs and benefits of education (Oreopoulos and Dunn, 2013; McGuigan et al., 2016; Kerr et al., 2015),

¹Students are considered to come from a non-academic family background if none of their parents holds a university degree.

²In 2006, seven out of sixteen states in Germany introduced tuition fees (around EUR 1000 per year), which triggered a lively discussion about fairness in access to university education. However, by 2014 all states had abolished tuition fees.

³Even in the English-speaking world the effect of financial aid programs is mixed (for an overview see Dynarski (2002)).

while other studies focus on specific information, i.e. provide students solely with information on financing possibilities (Booij et al., 2012; Herber, 2015) or examine the effect of information on the application process for college and financial aid (Bettinger et al., 2012; Hoxby and Turner, 2014) or the admissions process (Castleman et al., 2014). Furthermore, there are studies exploring the influence of (general) information on educational decision making in developing countries (Nguyen, 2008; Loyalka et al., 2013; Jensen, 2010; Dinkelman and Martínez, 2014) where the lack of information may be even more severe as obtaining information is more difficult.

This existing evidence shows that providing information improves students' knowledge. As we would expect, these improvements are larger for students from low socio-economic backgrounds indicating that *ex ante* students might underestimate the returns to post-secondary education or their probabilities of succeeding in higher education. Yet, it is still unclear under which circumstances and in which contexts the provision of information impacts educational choice. The type of information, the mode of presenting information, as well as the duration and the level of interaction varies greatly across studies. Correspondingly, results are mixed, allowing neither the conclusion that information impacts educational choices nor that it does not. Most existing studies, however, find a significant effect on students' knowledge, some find an effect on their educational aspirations, but few studies find an effect on actual behavior. In addition, most evidence refers to countries with comparatively high tuition fees. In these countries the extent to which information can affect educational decisions may be restricted as financial constraints might likely outweigh the lack of information.

Hence, looking at data from a German randomized controlled trial may shed further light on the effectiveness of information provision in a tuition free context. We analyze the differential effects of providing information on intended college enrollment for students' from different educational backgrounds. We estimate the treatment effect on intended college enrollment (1) two to three months after the information provision, i.e. one year prior to high school graduation and (2) one year after the intervention, i.e. shortly after students graduated from high school.⁴

We argue that students' intended college enrollment is a valid indicator for their actual enrollment, especially the closer enrollment intentions are measured to students' actual post-graduation decision. By analyzing intended college enrollment shortly after high school graduation, i.e. closer to the actual decision making, we might get at the potential effect of providing information on actual college enrollment. In support of this argument the empirical correlation between stated enrollment intentions and actual enrollment is very strong. Based on data from a German panel of high school students, 95 percent of students who state an enrollment intention half a year prior to high school graduation do enroll within three and a half years after graduation (Heine, 2010; Spangenberg et al., 2011).⁵

Additionally, examining intended college enrollment one year prior to high school graduation, i.e. two to three months after the information intervention, can yield further insights on the

⁴Hereafter we refer to students' intended college enrollment one year prior to high school graduation as short run since these enrollment intentions are measured shortly after the information provision (two to three months later); similarly, we refer to students' intended college enrollment shortly after high school graduation as enrollment intentions one year later as these are measured one year after the information intervention.

⁵Although this correlation is not necessarily informative about trajectories for treated students in this paper, it corroborates the predictive power of intentions for actual behavior.

effectiveness of providing information as it partly abstracts from supply side restrictions. This is because these enrollment intentions are more likely to reflect students' individual preferences for university education that are less dependent on the number of places available at universities or enrollment restrictions based on grade point averages. Thus, while intended college enrollment measured a year prior to high school graduation may already give us an indication about actual choices, enrollment intentions measured shortly after high school graduation, i.e. at the time students make their post-secondary educational decisions, are likely to be linked to actual enrollment.

Our results indicate that students process the information provided and adjust their subjective beliefs on benefits of college education accordingly. The information treatment also affects students' intended college enrollment. We show that the information intervention increases intended college enrollment for students from non-academic family backgrounds by 8 percentage points in the short run, i.e. two to three months after the intervention. This effect persists when measuring intended college enrollment one year later, suggesting that the provision of information might also increase their college enrollment. For students from academic family backgrounds, we find a marginally statistically significant decrease in intended college enrollment two to three months after the intervention. However, this negative effect disappears one year later, indicating that information provision is unlikely to play a role for these students' post-secondary educational choices.

Our study relates to the information treatments assessed by Oreopoulos and Dunn (2013); McGuigan et al. (2016) and Kerr et al. (2015). Yet, to the best of our knowledge, the study by Kerr et al. (2015) is the only other study providing information on the costs and benefits of university education in a tuition free country. Kerr et al. (2015) focus on students' choice of major in Finland and, thus, provide students close to graduation with major-specific information. They find no significant effect on major-specific applications or enrollment rates. The authors conclude that a potential lack of information on labor market success may not be important for educational choices. Complementing their analysis, our study adds to the existing literature by examining the effect of providing information on the decision about the level of education that students pursue after graduating from high school with a specific focus on educational inequality. Furthermore, the way in which the information was presented to students differs between the two studies. While in the study by Kerr et al. (2015) student counselors were provided with information material, our paper looks at the effect of an information workshop that was given by a trained person with a precise script followed by a short summary video at the end. This ensures a consistent provision of information across schools without risking any potential biases that could occur from student counselors or teachers presenting the information material to students.⁶

The remainder of the paper is structured as follows: Section 2 describes the institutional context in Germany. Section 3 presents the randomized controlled trial, the intervention as well as the data, while Section 4 introduces the empirical strategy. In Section 5 we report our estimation results and briefly discuss some robustness tests. Section 6 concludes.

⁶Teachers and/or student counselors who are provided with information material, may present this material with their own interpretation and/or present only a selection of the material to students.

2. Institutional context

In Germany, education policy is the responsibility of each individual federal state (*Bundesland*). Thus, education systems differ across the sixteen states. The data used in this paper stem from a randomized controlled trial conducted in the federal state of Berlin, where students complete six years of primary school⁷ before being assigned to different tracks of secondary schooling based on their performance. Secondary school tracks can be differentiated into a vocational and a university track.⁸ Only at university track schools can students earn the general university entrance diploma, which in Germany is called *Abitur*, that allows students to immediately start university following graduation. This study uses data on students working toward the *Abitur* qualification;⁹ excluding those striving for other specialized high school diplomas. In Berlin students can earn their *Abitur* at 137 schools. These 137 schools are divided into three school types: (1) general high schools (*Gymnasium*); (2) comprehensive high schools (*integrierte Sekundarschule*); and (3) vocational oriented high schools (*berufliches Gymnasium*).

Post-secondary educational decisions in Germany differ from other countries. After earning the *Abitur* almost all students stay in post-secondary education, with a very small share deciding not to seek any further education. Given the tracking system, students studying for the *Abitur* are, in general, on track to pursue a subsequent education at university. However, approximately a quarter of students graduating with the *Abitur* choose a vocational education instead (Autorengruppe Bildungsberichterstattung (2016): 127). The German vocational education system constitutes an attractive alternative to university education, as it is a highly recognized dual system that offers good employment prospects. Although primarily designed for students with a lower or middle secondary schooling degree, a range of vocational apprenticeship programs now require the *Abitur*. In addition, the probability of admission to white collar vocational programs is very low without the *Abitur*. As the number of students pursuing an apprenticeship after obtaining the *Abitur* has increased over the years, students who would have left school with a (very) good middle secondary schooling degree might decide to pursue the university entrance qualification only to enter profitable vocational education programs. If policy makers aim to increase enrollment rates at universities, targeting this group may be most effective because these students are already equipped with the necessary academic performance.

Students from low educational backgrounds are more likely to pursue vocational education than peers from an academic family background. Conditional on earning the *Abitur*, the transition probability to university education is between 10 and 20 percentage points lower for students with lower educated parents, i.e. parents without university degree (Autorengruppe

⁷The transition to secondary schooling after six years occurs in three federal states (Berlin, Brandenburg, and Mecklenburg-West Pomerania); in all other federal states children transit to secondary school following the completion of grade four.

⁸We subsume *Hauptschule* and *Realschule* as vocational track schools and *Gymnasium* and schools with upper secondary level (*gymnasiale Oberstufe*) as university track schools.

⁹Given the early school tracking in Germany after grade four (or six), students attending university track schools represent a selected group, who may already be better informed than students attending other school tracks. Hence, focusing only on university track schools may lead to an underestimation of the potential effect of information provision. Treating students with information on the benefits of university education earlier in the school career could also affect students' high school track choice, as in most federal states teachers' track recommendations are not binding.

Bildungsberichterstattung (2016): 127). Given tracking after primary school and the associated selectivity of students who earn the *Abitur*, the observed differences in post-secondary decisions by students' educational background is an additional source of concern: If the inequalities at earlier stages are taken into account, the probability of starting university education are more than three times as high for students from academic compared to non-academic family backgrounds (77% vs. 23%) (Middendorff et al., 2013). One immediate benefit of vocational education in the dual system is its remuneration, which renders students somewhat more independent of other financial sources to cover their living expenses than students attending college. Some authors argue that having a vocational education system that offers students an attractive alternative to university education may partly explain why students from low educational backgrounds are underrepresented at German universities (Becker and Hecken, 2008).

3. Randomized controlled trial

In this Section the setup of the randomized controlled trial (RCT hereafter) and the data used are described in more detail. The information intervention was conducted as part of a larger project called *Berliner-Studienberechtigten-Panel* (Best Up).¹⁰ In this project, randomly selected high schools in Berlin were treated with an in-class presentation providing information on benefits of university education as well as on potential financing strategies.

“Best Up” project setup. The project aimed to obtain a sample of 27 schools (20% of all upper secondary schools in Berlin) that have a large share of students from non-academic family backgrounds. High schools without intakes in fifth grade¹¹ were stratified using (1) school type; (2) share of population aged 25 and older with low education (ISCED 0-2) per district; (3) cohort size one year prior the *Abitur* exams; (4) share of students with a migration background; and (5) share of female students as stratifying variables. With the exception of the share of low educated individuals within a district, all variables are measured at the school level. The Best Up project aimed at oversampling students from lower educated backgrounds. Since there is no school-level information available on students' parental educational background, we included district-level information. This allowed us to identify schools in areas with a higher share of low educated individuals and subsequently increased the likelihood of sampling students from non-academic family backgrounds. Stratification was implemented using coarsened exact matching (CEM) as proposed by Iacus et al. (2009).¹²

Based on the results of the stratification, a set of potential schools – 30 preferred schools and 20 replacement schools – was identified that was similar in terms of the stratifying variables. Schools in the preferred set were subsequently contacted and asked whether they would like to participate in a survey aiming to gain knowledge on how students can be better supported

¹⁰The project is a co-operation between the German Institute for Economic Research (DIW Berlin) and the Berlin Social Science Center (WZB). The Best Up project is funded by the Einstein Foundation Berlin. For further information on the project see: http://www.diw.de/en/diw_01.c.409542.en.

¹¹Out of the 137 Berlin high schools, 33 schools which admit high performing students in grade five are excluded from the target population, since students with a non-academic background are likely to be underrepresented in these schools.

¹²Stratification was only used to draw the school sample and played no role in randomization.

in choosing their post-secondary educational path. During the recruitment process nine out of the 30 preferred schools had to be replaced with schools from the replacement sample.¹³ Table A.1 in the Appendix presents descriptive statistics for the different sets of schools, comparing all Berlin high schools without intakes in grade five to potential schools (showing the set of preferred and replacement schools separately), contacted schools, and schools that participate in the *Best Up* study. This comparison shows that, in line with the aim to oversample students from non-academic families, the set of preferred schools are more frequently located in districts with a higher share of low educated individuals, comprise a larger share of vocational high schools and exhibit a higher share of students with a migration background than the average Berlin high school. Table A.1 further shows that the set of contacted schools and that of participating schools are similar in terms of the stratifying variables.

After schools had agreed to participate, schools within school types were randomly assigned into treatment and control groups. In the sample, nine schools out of 27 are treatment schools. After allocating schools into treatment and control groups, headmasters were contacted again to schedule a date for the survey. Treatment schools were asked for an additional class session (45 minutes), to accommodate the information workshop. A few weeks before the survey, an invitation to participate in the survey was distributed among all students who were on track to take *Abitur exams* the following year.¹⁴ Among the nine treatment schools, one school did not receive the information workshop due to a miscommunication between the headmaster and its teaching staff. Nevertheless, it was possible to survey some students in this school. We further address the non-compliance of this school in our empirical strategy in Section 4.

Information intervention. The information workshop was composed of a 20-minute in-class presentation on benefits of post-secondary education as well as on different funding possibilities of university education. The information on labor market returns comprised visualized information on earnings differences, career perspectives, unemployment risk and the gain in lifetime earnings. Students received “tailored information,” meaning information relevant for students with *Abitur*. The general numbers available on differences in earnings do not differentiate by highest achieved schooling degree. While *Abitur* is a prerequisite for university enrollment, most vocational degrees can also be obtained with a lower schooling degree. Consequently, the returns to a vocational degree largely depend on the highest achieved schooling degree.¹⁵ Thus, during the information workshop, labor market benefits of university education were compared to vocational education conditional on holding the *Abitur*. Through the comparison of labor market benefits between a university and a vocational degree, the information workshop also conveyed information on labor market benefits of vocational education. The presentation also pointed toward gender differences in earnings and differences across fields of study.

¹³Six of the nine schools that had to be replaced in the “preferred set” were general high schools and three were comprehensive high schools.

¹⁴As part of the setup of the randomized controlled trial, power analyses were conducted to judge the feasibility of the intervention. Taking the full cohorts of the 27 schools as our potential sample (2,500 students) and assuming a response rate of at least 60 percent, the minimum detectable treatment effect was equal to 6 percentage points (with α equal to 0.05 and β equal to 0.20). Additionally accounting for a panel mortality of 20%, increased the minimum detectable treatment effect size to 7 percentage points in the overall sample.

¹⁵Students holding a lower secondary schooling degree do not qualify for all vocational education programs.

With respect to the possibilities to finance university studies, the main sources of funding in Germany – BAföG (student aid), scholarships and students jobs – were introduced. The information on student aid also included basic information about repayment conditions, stressing that only half of the amount received as student aid must be repaid and repayment obligations only start once earnings exceed a certain threshold. The information on direct costs of university education emphasized that no tuition fees need to be paid (anymore) and, consequently, monthly average costs equal living expenses, which have to be financed irrespective of the educational path taken. Hence, the costs of university education boil down to the opportunity costs, which correspond to the remuneration of vocational trainees. Most of the information was visualized in order to make the information more accessible to students. Figure A.1 in the Appendix shows some example slides of the material presented in the information intervention.

The information workshop was not designed to advertise university education but rather to provide students with information relevant to making a more informed decision. In addition, the presentation was given by a trained person with a precise script from the RCT team. This type of treatment ensures a more consistent provision of information compared to other studies that give out information materials to schools or student counselors (see for example the studies by McGuigan et al., 2016; Kerr et al., 2015, for this type of treatment), who might present this material with their own interpretation and/or present only a selection of the provided information material. Another component of the information treatment was a 3-minute video at the end of the intervention summarizing the provided information and thereby further guaranteeing standardization of treatment.

Data. We use data from the *Berliner-Studienberechtigten-Panel* (Best Up) with pre- and post-treatment surveys. The pre-treatment survey was administered in schools one year prior to the *Abitur* exams using a paper-based questionnaire. It was executed in schools under exam conditions. Teachers were only present to provide their obligatory supervision. In treated schools, the survey directly preceded the information workshop. A total of 1,578 students participated in the first survey.¹⁶ Approximately two to three months and one year later follow-up online surveys were carried out. The response rates for the post-treatment surveys, each compared to the baseline number of students (1578), were 70% and 67%, respectively, which is higher than in comparable studies (see e.g. Booij et al., 2012; Oreopoulos and Dunn, 2013). More importantly, the response rate is virtually identical between treatment (69.69%) and control (70.71%) groups.¹⁷ Yet, to obtain an unbiased estimator of the treatment effect it is important that intended college enrollment and background characteristics do not influence drop out differently by treatment status. Based on a Chow-test, we do not find any evidence for differential

¹⁶Taking the full cohort of each school as a reference, this corresponds to an overall response rate of 60%.

¹⁷These numbers refer to the first post-treatment survey, i.e. students who participated in the survey two to three months after the information workshop. Due to the change in survey mode, attrition is highest between the pre-treatment and the first post-treatment survey. 96% of students, who participated in the first post-treatment survey also responded to the second post-treatment survey one year later.

attrition.¹⁸

Analyzed sample. We restrict our analysis to students participating in both pre- and post-treatment surveys. Further, we keep only students with information on pre- and post-treatment enrollment intention as well as information on parental educational background. Intended college enrollment one year prior to high school graduation is measured by asking students what education they plan to pursue after earning their *Abitur*.¹⁹ Students can choose between university education (at different types of universities²⁰), vocational education, or no education. We define intended college enrollment as a binary variable, such that it equals one if the student intends to go to college and zero otherwise. The vast majority of students who do not intend to enroll, plan to pursue a vocational education.²¹ The final sample for the analysis focusing on short-run treatment effects comprises 988 observations. Out of these students, valid information on intended college enrollment shortly after high school graduation is available for 842 students.

Given the variety of post-secondary educational paths, intended college enrollment measured shortly after high school graduation is determined as follows: For students, who already applied to study programs at the time of the survey, i.e. directly in the summer after high school graduation, this enrollment intention reflects their applications.²² For other students, it reflects either their plan to apply/enroll in the same year or their enrollment intention after taking a gap year in order to, for example, travel, do an internship or voluntary work.

Further, we define parental educational background to be either academic or non-academic. Students are from a non-academic family background if no parent (genetic or social) holds a university degree, or from an academic family background if at least one parent holds a university degree. For students who did not answer the question addressing education of both parents, we made the following assumption to determine their educational background: Students either stating that they do not know their mother or father or students with missing information on the level of education of one parent were classified according to the valid information on the one (the other) parent.²³ In specifications where we control for additional covariates, we deal with missing information by setting these variables to a constant value and including a dichotomous variable indicating missing covariates.²⁴ Missing information on the key variables does not differ significantly between treatment and control groups.

¹⁸Tested covariates comprise age, gender, migration background, non-academic family background, school types, enrollment intention, math and German grades as well as two measures of cognitive skills and again refer to participation in the first post-treatment survey; $F_{(12,1545)} = 0.68, p - value = 0.7725$.

¹⁹The translated survey question reads: Think of everything you know today: Which type of education will you most likely pursue after graduating from high school?

²⁰The institutions comprise universities, universities of applied sciences, field specific universities, and vocational oriented universities.

²¹Only around two percent of the students who participated in the pre-treatment survey plan to obtain no further education.

²²In Germany, college applications are only required for some study programs.

²³If information on parental education is completely missing, we use the education of older siblings (if available) to proxy educational background; otherwise we dropped the observation from the sample.

²⁴Estimating the treatment effect using only students with non-missing information on all covariates, does not change our conclusions.

Covariate balance. We test whether randomization was successful by comparing the balance of covariates between treatment and control groups. As is common in RCTs in the field of education, schools instead of individuals were randomized to best mimic a potential policy measure and avoid spillover effects within schools. As we are interested in the differential effect by parental educational background, we examine treatment effects at the individual level. However, the composition of students within schools is usually non-random, such that the probability of balancing covariates at the individual level is lower if entire schools instead of individuals are randomized. We assess randomization in the combined sample as well as for the subgroups by parental educational background. Table 1 displays the covariate balance by treatment status and indicates that randomization successfully balanced most of the covariates.²⁵ The only exception is detected in the subsample of students from academic backgrounds, where students in the treatment group are more likely to state that lifetime income is higher with a university degree than with a vocational degree. Conducting F-tests in a regression of individual characteristics and measures of performance and skills (as listed in Table 1) on treatment status, we cannot reject the null hypothesis that these variables are jointly equal to zero in all three samples.²⁶

[Table 1 about here]

4. Empirical framework

When analyzing data from a RCT it is generally sufficient to compare the average post-treatment outcomes by treatment status in order to identify the causal effect of the treatment. Randomization ensures that the estimates do not suffer from selection into treatment. However, based on the information of the pre-treatment survey, we see that (conditional on the sample used for the analysis) pre-treatment intentions to enroll in college are almost three percentage points lower in the treatment group than in the control group. If we look at the subsample of students with a non-academic background this difference is even larger and amounts to five percentage points (see Table 1).

Although these differences are statistically insignificant, the size of the difference cannot be ignored. If, for example, the true effect of the information intervention for students from non-academic families is less than five percentage points, by only comparing post-treatment outcomes, we would conclude that the information intervention had no effect on intended college enrollment. Further, even if the true effect is larger than five percentage points, we would still underestimate the treatment effect for students from a non-academic family background – the group of major interest in our study.

In addition to the differences in pre-treatment enrollment intentions, one school that was randomly assigned to the treatment group did not receive the information workshop (see Section 3). It was, however, possible to survey at least some of the students in this school. Thus, to obtain a causal effect of providing information, we compare post-treatment intended college

²⁵This also applies to the covariate balance in the baseline sample as well as for the sample used to analyze the treatment effects one year after the information intervention. See Table A.2 and A.3 in the Appendix.

²⁶Corresponding p-values of the F-tests are: in the combined sample 0.5229; in the sample of non-academics 0.2968; and in the sample of academics 0.7488. F-test are based on regressions with standard errors clustered at the school level.

enrollment by treatment status controlling for students' pre-treatment intention combined with a two stage least squares approach. In the first stage, we use the original classification of schools into treatment and control groups (based on randomization) as an instrument to predict actual treatment status, which is whether a school actually received the information workshop. The first stage is given by:

$$T_{is} = \mu + \eta Z_s + \gamma y_i^{(t_0)} + \delta X_i + \tau_{is} \quad (1)$$

where T_{is} indicates actual treatment status and Z_s indicates the treatment status obtained from randomization prior field start. We account for differences in students' pre-treatment enrollment intentions by including $y_i^{(t_0)}$, a binary variable indicating student i 's pre-treatment intended college enrollment. In order to account for the slight imbalance of covariates between treatment and control groups, we further include a vector of additional pre-treatment individual level controls X_i . X_i includes age, gender, migration background, school type, (standardized) pre-treatment math and German grades as well as cognitive skills measured by a verbal and a figural test.

After obtaining the predicted treatment status \hat{T}_{is} , we estimate Equation 2 for the whole sample as well as separately for students from a non-academic and academic family background:

$$y_{is}^{(t_\omega)} = \beta_0 + \beta_1 \hat{T}_{is} + \beta_2 y_i^{(t_0)} + X_i' \beta_3 + \epsilon_{is} \quad (2)$$

where $y_{is}^{(t_\omega)}$ equals 1 if student i in school s intends to enroll in college at time t_ω ($\omega = 1, 2$), and 0 otherwise. $\omega = 1$ indicates the first post-treatment survey, i.e. two to three months after the information provision, and $\omega = 2$ indicates the second post-treatment survey, i.e. one year after the treatment. \hat{T}_{is} is the predicted treatment group indicator as estimated from Equation 1. $y_i^{(t_0)}$ and X_i are defined as before in Equation 1. The error term ϵ_{is} captures the remaining variation. To account for potential dependence of observations within schools we cluster standard errors at the school level. For the mean comparison of post-treatment intentions, β_1 is the coefficient of interest and identifies the effect of the information treatment.

However, controlling for students' pre-treatment intended college enrollment (see Equation 2) cannot completely resolve the pre-treatment difference, as it only adjusts the estimates for a fraction of these differences (Allison, 1990). Therefore, in our main specification we compare the *change* in students' intended college enrollment by treatment status and examine the difference between pre- and post-treatment enrollment intentions. Our outcome variable is given by $\Delta y_{is}^{(t_\omega)} = y_{is}^{(t_\omega)} - y_{is}^{(t_0)}$, where again $\omega = 1$ represents post-treatment intended college enrollment two to three months after the intervention and $\omega = 2$ indicates enrollment intention one year later. We estimate the following Equation and use the predicted treatment status \hat{T}_{is} from Equation 1 as our treatment indicator. Our preferred specification is given by:²⁷

$$\Delta y_{is}^{(t_\omega)} = y_{is}^{(t_\omega)} - y_{is}^{(t_0)} = \gamma_0 + \gamma_1 \hat{T}_{is} + X_i' \gamma_2 + v_{is} \quad (3)$$

²⁷Results based on estimating Equation 2 are reported in Table A.5 in the Appendix. Note that Equation 3 is a version of Equation 2 where we restrict β_2 to be equal to one.

where $\Delta y_{is}^{(t_\omega)}$ depicts the change in intended college enrollment of student i in school s between time t_ω and t_0 . We also add a vector of additional covariates, X_i (defined as before), to this specification.²⁸ The error term v_{is} is clustered at the school level.²⁹ The effect of the information treatment is given by γ_1 .

By using the change between pre- and post-treatment intended college enrollment in our main specification, we not only fully account for the pre-treatment imbalance in enrollment intentions but also for any time invariant observables and unobservables that might influence intended college enrollment and differ by treatment status.

5. Results

Before we present the effect of the information workshop on students' intended college enrollment, we first provide some descriptive evidence on the lack and relevance of information using pre-treatment data. We then show that the information workshop successfully conveyed information to students and subsequently turn to our main results.

5.1. Pre-treatment survey evidence

Intended college enrollment in our sample is (pre-treatment) around 13 percentage points lower for students from non-academic compared to students from academic backgrounds. In Table 2 we differentiate between students from different educational backgrounds with and without intentions to enroll and investigate whether information sets are related to their intended college enrollment.³⁰ Thus, we present mean and mean differences based on regressing each variable on an indicator variable for intended college enrollment, i.e. $X_i = \alpha + \beta y_i^{(t_0)} + \varepsilon_i$, where X_i represents the variable in the left most column of Table 2 and $y_i^{(t_0)}$ is an indicator variable for pre-treatment intended college enrollment.³¹

[Table 2 about here]

Table 2 shows that students from an academic background who intend to enroll in college are five percentage points more likely to rely on their parents and perceive this information source as more helpful than students having no enrollment intentions. In contrast, this does not apply to students from non-academic backgrounds.

Comparing the information set by intended college enrollment for students from a non-academic family background (columns 1 and 2 of Table 2) reveals that students with an intention to enroll feel significantly better informed about university education than their peers

²⁸ Male/female students, students attending different school types, or students located in different parts of the skill distribution might differ regarding a change in their intended college enrollment depending on what other information they acquire or experiences they gain in the meantime. For example, traditional academic track high school (Gymnasium) may be more likely to provide information about university education, while vocational oriented university track high schools may be more likely to inform students about traineeships in companies.

²⁹ Accounting for the small number of clusters does not change our conclusions (see Table B.1 in Section 6).

³⁰ Table A.4 in Appendix A further shows that students' information sets differ by parental educational background.

³¹ The columns in Table 2 present estimates of α and β .

without an enrollment intention.³² These students are also more likely to have investigated the possibilities of financing university attendance and perceive the cost burden of university education as lower.³³ Note that for students from a non-academic family background the subjective income premium associated with a higher degree is not correlated with educational aspirations. However, perceiving the unemployment risk to be lower, the prospects of finding a well-paid job and lifetime earnings to be higher with a university degree compared to a vocational degree is highly correlated with students' intended college enrollment. Thus, a lack of information on returns to tertiary education could potentially affect college enrollment.

Students from academic backgrounds, who intend to enroll in college, likewise perceive the returns to university education as higher. However, only the prospect of finding a well-paid job is (marginally) statistically different between students with and without an enrollment intention in this subgroup. Looking at the “costs” of university education shows that feeling well informed about university education and having dealt with financing possibilities is also positively correlated with enrollment intentions for students from academic family backgrounds.

Overall, Table 2 suggests that information is relevant in forming an enrollment intention for students from both, non-academic and academic, family backgrounds. However, it seems even more important for students from a non-academic family background.

5.2. *The effect of information provision on intended college enrollment*

Before we turn to the treatment effects on intended college enrollment, we briefly discuss whether the information workshop successfully conveyed information that was adequately processed by students. We compare students' perceived labor market benefits of university education pre- and post-treatment. We consider the subjective unemployment risk, the subjective prospects of finding a well-paid job, and the subjective income premium of university education. We are only able to assess a small subset of subjective beliefs of labor market returns. The information treatment, however, consisted of a bundle of information on post-secondary education among which labor market returns were just one aspect. Unfortunately, the post-treatment surveys do not contain questions about funding possibilities, making it difficult to disentangle the effects of the information regarding returns from that regarding financing.

The estimates in Table 3 are based on Equation 3 and suggest that students absorbed the provided information. Treated students updated their subjective beliefs in the expected way and all estimates have the expected sign and are, with one exception, statistically significant. Students in the treatment group are significantly more likely to expect their unemployment risk to be smaller and their prospects of finding a well-paid job to be higher with a university degree than with a vocational degree. As such, the information workshop seems to have provided students with relevant information that may influence their educational decision making.

[Table 3 about here]

³²This is based on the question of whether students feel well-informed about the general rules and possibilities of university.

³³The perception of “how difficult financing university education” is also varies significantly by parental educational background. Almost half of the students from a non-academic family background state that bearing the costs during university education is very difficult or mostly difficult (see Table A.4 in the Appendix).

After providing evidence that students process the information from the in-class presentation, we now turn to the main results. Table 4 presents the treatment effects of the information intervention on intended college enrollment (1) one year prior to high school graduation, i.e. two to three months after the intervention and (2) shortly after high school graduation, i.e. one year later. We argue that this second outcome is likely to be linked to students’ actual enrollment behavior. Table 4 further shows the treatment effect for the whole sample as well as separately by parental educational background. We report estimates of the treatment effect based on Equation 3, i.e. our preferred specification, in which we analyze the change in students’ intended college enrollment.³⁴

[Table 4 about here]

Looking at the effects of the information workshop on intentions one year prior to graduation (see upper panel of Table 4), we find a positive, but statistically insignificant effect of the information intervention on intended college enrollment in the whole sample. However, this result masks considerable effect heterogeneities by parental educational background. Considering students from non-academic and academic backgrounds separately shows that the information intervention increases intended college enrollment for students from non-academic backgrounds by around 8 percentage points in the short run (p-value < 0.05).

In contrast, students from an academic background decrease their enrollment intentions 2-3 months after the information workshop by 5.6 percentage points (p-value < 0.10). Although the negative effect for students from academic families is only marginally statistically significant, the sign of the effect might at first be a rather surprising finding. However, the fact that students from different educational backgrounds respond in opposite direction to the information treatment in the short term, suggests that information sets of students may indeed be biased towards the educational level that prevails in their environment. Where students from non-academic family backgrounds may lack information about university education, students from academic backgrounds may have an information deficit about options other than university education.³⁵

Focusing on the change in intended college enrollment one year after the intervention, i.e. shortly after high school graduation (see bottom panel of Table 4), reveals that the marginally statistically significant negative treatment effect on students from academic backgrounds does not persist. The information intervention has no statistically significant impact on students’ enrollment intentions one year later. Among treated students from academic family backgrounds, more than two thirds revert to their pre-treatment intention to enroll one year later. For this

³⁴Estimates based on Equation 2 are reported in Table A.5 in the Appendix. In this mean comparison, the estimates of the short-run effects decrease in size but are contained in the 95% confidence interval of the effects presented in Table 4. The estimates of the effect on intended college enrollment directly after high school graduation (one year after the treatment) are similar to Table 4, but the effect for students from non-academic backgrounds is not statistically significant.

³⁵After the information workshop, some students from academic family backgrounds might have regarded vocational education to be more attractive than they originally assumed and, for the first time, considered vocational degree as a valid “outside” option. It might be that for these students raising the awareness for alternatives to university education and providing further information on vocational education may indeed induce them to choose this path.

group of students family expectations are likely to outweigh the information treatment, since path dependency might be even stronger for this subgroup as downward mobility in educational attainment rarely occurs in Germany (see for example Heineck and Riphahn, 2009; Mueller and Pollak, 2015; Schnitzlein, 2016).³⁶

For students from non-academic backgrounds, Table 4 shows that the information intervention still affects students' intended college enrollment shortly after graduation, i.e. one year after the workshop. The estimates of the effect on these intentions that are likely linked to actual enrollment remain similar in size and statistical significance level compared to the findings on enrollment intentions in the short run. Adding control variables only marginally changes our estimates. The information workshop increases students' intended college enrollment measured shortly after high school graduation, i.e. one year after the information workshop, by 8 percentage points. Given students' baseline enrollment intention, this effect corresponds to an overall boost in the share of students with a non-academic family background intending to go to university of about 11 percent.³⁷ Within the control group, the share of students from non-academic families intending to enroll in college decreases by 2.3 percentage points.

Our results further imply that the information treatment successfully decreases the gap in students' intended college enrollment by parental educational background. Prior to the information treatment this "education gap" in enrollment intentions was 15 percentage points in the treatment group and 12 percentage points in the control group. By increasing intended college enrollment for students from non-academic family backgrounds, the information workshop reduces the gap measured shortly after high school graduation in the treatment group by 4 percentage points (by 11 percentage points); while the gap in the control group only decreases by 6 percentage points.³⁸

In sum, while our findings on intended college enrollment one year prior to students' high school graduation yield valuable insights on the effectiveness of providing information in the absence of supply side restrictions, we argue that by analyzing enrollment intentions shortly after high school graduation, i.e. closer to the actual decision making, we might get at the potential effect of providing information on college enrollment. Our results show that the information workshop increased intended college enrollment for students from non-academic backgrounds. Thus, we similarly expect the information provision to increase college enrollment rates for these students. In contrast, it seems unlikely that enrollment rates for students from academic family backgrounds will be affected.

5.3. *Adjustments to pre-treatment educational plans*

In order to better understand the effect of the information workshop, we disaggregate the change in intended college enrollment into three further outcomes. Between periods,³⁹ students

³⁶Note that enrolling in the German vocational education system, especially in the dual system, might be more difficult as it requires more timely effort and initiative from students than enrolling in college. This may further explain why some of these students revert to their enrollment intentions shortly after graduating from high school.

³⁷Pre-treatment intended college enrollment for students from non-academic family backgrounds in the treatment group is equal to 69.9 (see Table 1).

³⁸These numbers are calculated without the one non-compliant school.

³⁹This either compares the period between the pre-treatment survey and the survey two to three months after the treatment or between the pre-treatment survey and the survey one year after treatment.

can either adjust their educational expectations upward, downward or remain within their educational plan. We define *upward adjustment* as a binary variable equal to one if a student has no intention to enroll in college pre-treatment and changes her intention towards pursuing a college degree post-treatment, and zero otherwise. Similarly, *downward adjustment* indicates students who change from having an enrollment intention (pre-treatment) to having no intention anymore (post-treatment). Finally, if students maintain their educational intentions, either to enroll in college or to obtain a vocational degree, we refer to this as *stable intentions*. This disaggregation is particularly interesting for students from non-academic backgrounds as it is shown in the literature that these students have more difficulties in forming and maintaining high educational expectations (see e.g. the literature reviewed in Engle, 2008). Based on Equation 3, we estimate the effect of the information intervention on these three outcomes separately and present the results in Table 5.

[Table 5 about here]

Focusing on adjustments one year after the information intervention, Table 5 shows that for students from non-academic families the information treatment significantly decreases the probability to adjust enrollment intentions downward. Treated students from non-academic family backgrounds are 6.3 percentage points less likely to change from intended college enrollment to no intention (see column 3 of Table 5). The corresponding mean in the control group is equal to 13.6 percent, which implies that the information intervention cuts the share of students who adjust their enrollment intentions downward almost in half. Moreover, Table 5 shows that the short-run effect of the information intervention on enrollment intentions of students from non-academic family backgrounds can be attributed to a statistically significant upward adjustment in intended college enrollment. The information treatment almost doubles the share of students who adjust their intentions upwards in the short run. In addition, among those students who moved from no intention to intended college enrollment in the short run, approximately 71 percent maintain their changed enrollment intentions one year later.

In contrast, the marginally statistically significant negative treatment effect for students from academic backgrounds is caused by averting an upward adjustment rather than by provoking a downward adjustment in intended college enrollment (see column 2 of Table 5). For students from academic families the information treatment decreases students' likelihood of an upward adjustment by 4.1 percentage points in the short run.

The results in Table 5 suggest that overall the information workshop mainly worked through fostering enrollment intentions for students from a non-academic family background.⁴⁰ These students are more likely to maintain their intended college enrollment due to the information provision.

⁴⁰Further analyses on students *with* enrollment intentions support this finding (Ehlert et al., 2017).

5.4. Robustness tests

We also investigate how sensitive our estimates are to different specifications. We briefly summarize them here and provide a more detailed discussion of each sensitivity test in Appendix B. None of the sensitivity tests change our conclusion.⁴¹

First, selective attrition or other differences between treated and control group students may bias our results. Therefore, we tested whether our results were similar when we use either inverse probability weights or entropy balancing weights. Second, the non-compliance of one treatment school could affect our estimates of the treatment effect; thus we rerun our estimations excluding this non-compliant school from the sample and reassigning it to the control group. Third, we account for the small number of clusters in our analyses and apply the wild cluster bootstrap-procedure as suggested by Cameron and Miller (2015). Fourth, we estimate a difference-in-difference type of regression, which allows us to include school fixed effects to account for time invariant school level omitted variables that might affect students' enrollment intentions. Fifth, we want to eliminate the possibility that our results are driven by selection into (student-level) survey participation, as we observe school-level differences in response rates to the pre-treatment survey. We therefore drop the schools with the five lowest school-level response rates from our estimation sample. Sixth, to refute concerns about potential spill-over effects of the information workshop across schools, we exclude all control schools within a two kilometer radius of a treatment school. Finally, we account for the fact that we made several assumption to proxy students' educational background and restrict our sample to students with unambiguous information on their parents' education.

6. Conclusion

This paper contributes to the growing economic literature on the effect of information provision on educational decisions. We present results using data from a randomized controlled trial in Germany. Students in randomly selected schools were treated with information about labor market benefits of university education as well as about different funding possibilities. Students seem to comprehend the information they were given. Our results show that students in the treatment group are significantly more likely to expect their unemployment risk to be smaller and their prospects of finding a well-paid job to be higher with a university degree than with a vocational degree. We find that the provision of information increases intended college enrollment for students from non-academic backgrounds, both two to three months and one year after the information treatment. For these students, the information treatment prevents a downward adjustment of their enrollment intentions, i.e. it avoids that these students might be discouraged, if peers and parents based on their own preferences support a differing educational trajectory.

In contrast, the information treatment leads students from academic family backgrounds to lower their enrollment intentions in the short term (albeit this effect is only marginally statistically significant). The treatment may have led these students to re-consider their options after

⁴¹ However, the short-run point estimates for students from academic backgrounds vary slightly more given the smaller sample size.

graduation instead of routinely following the expectations of their surroundings. However, our results show that for students from academic families the change in intended college enrollment is only temporary, since we do not find a treatment effect on their enrollment intentions one year later. Thus, we argue that the information provision is likely to increase college enrollment rates for students from non-academic family backgrounds, while it seems unlikely that the information treatment will affect college enrollment of students from academic family backgrounds.

Given the evidence from the U.S. on the so called “summer melt” (e.g. Castleman et al., 2014; Castleman and Page, 2015) it may, however, not suffice to foster higher educational expectations of students from disadvantaged or non-academic backgrounds to increase their enrollment. Castleman et al. (2014) show that, given the complex admission process in the U.S., students from disadvantaged backgrounds need further assistance to follow through on their educational plans. However, in Germany the matriculation process is comparatively less complicated. In Germany, students, who intend to enroll in college, face fewer challenges in the summer following high school graduation than in the U.S., i.e. less forms to fill out, hardly any placement test amongst other things. Thus, we argue that they are more likely to translate their enrollment intentions into actual matriculation.

The fact that we find a statistically significant effect on intended college enrollment for students from non-academic family backgrounds shows that pre-treatment plans do not reflect optimal choices and that those students indeed lack information. If students’ intentions were already optimal prior to treatment, receiving information should have no effect. However, although we find a causal effect of information provision, the question of which specific information triggered this result, is less clear. Further research is needed to obtain a better understanding of what particular type of information helps students from non-academic family backgrounds to make an informed decision and encourages them to pursue university education.

In contrast to the study by Kerr et al. (2015), our results indicate that providing (general) information has the potential to impact educational choices, especially for students from non-academic families. One explanation for the differing results, despite the similar context, may be that the general educational decision, i.e. students’ choice between university education and an alternative, may be more responsive to information than students’ choice of college major. Another possibility may be that teachers/counselors in the RCT by Kerr et al. (2015) differ in their presentation of the information materials and thus no significant treatment effect can be identified. In addition, the mere fact that information is provided in school by an external person, i.e. a person outside the school context, may further contribute to the effectiveness of the information workshop analyzed in this paper.

The gap in educational attainment by family background is mostly discussed from the angle of inequality of opportunities, whereas the loss of efficiency through an underutilization of human capital is often neglected. However, the efficient use of these resources is important, especially in countries with a shrinking labor force. The findings of this paper show that educational inequality – measured by the differences in students’ intended college enrollment by parental educational – can be reduced by providing students with relevant information. A tailored information workshop may indeed be an appropriate and inexpensive policy tool to narrow the gap in take up of university education.

References

- Allison, P. D., 1990. Change scores as dependent variables in regression analysis. *Sociological Methodology* 20, 93–114.
- Amt f?r Statistik Berlin-Brandenburg, 2011. Regionaler Sozialbericht Berlin und Brandenburg 2011. Tech. rep., Federal statistical office Berlin-Brandenburg (Amt f?r Statistik Berlin-Brandenburg).
- Autorengruppe Bildungsberichterstattung, 2016. Bildung in deutschland 2016. Ein indikatorengest?tzter Bericht mit einer Analyse zur Bildung und Migration.
- Becker, R., Hecken, A., 2008. Warum werden arbeiterkinder vom studium an universit?ten abgelenkt? eine empirische ?berpr?fung der “ablenkungsthese” von m?ller und pollak (2007) und ihrer erweiterung durch hillmert und jacob (2003). *K?lner Zeitschrift f?r Soziologie und Sozialpsychologie* 60 (1), 3–29.
- Bettinger, E. P., Long, B. T., Oreopoulos, P., Sanbonmatsu, L., Aug 2012. The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *The Quarterly Journal of Economics* 127 (3), 1205–1242.
- Bifulco, R., Fletcher, J. M., Oh, S. J., Ross, S. L., 2014. Do high school peers have persistent effects on college attainment and other life outcomes? *Labour Economics* 29, 83 – 90.
- Booij, A. S., Leuven, E., Oosterbeek, H., 2012. The role of information in the take-up of student loans. *Economics of Education Review* 31 (1), 33–44.
- Cameron, A. C., Miller, D. L., 2015. A practitioner’s guide to cluster-robust inference. *Journal of Human Resources* 50 (2), 317–372.
- Castleman, B. L., Page, L. C., 2015. Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization* 115, 144–160.
- Castleman, B. L., Page, L. C., Schooley, K., 2014. The forgotten summer: Does the offer of college counseling after high school mitigate summer melt among college-intending, low-income high school graduates? *Journal of Policy Analysis and Management* 33 (2), 320–344.
- Dinkelman, T., Mart?nez, C., May 2014. Investing in schooling in chile: The role of information about financial aid for higher education. *Review of Economics and Statistics* 96 (2), 244–257.
- Dynarski, S., 2002. The behavioral and distributional implications of aid for college. *The American Economic Review* 92 (2), 279–285.
- Ehlert, M., Finger, C., Rusconi, A., Solga, H., 2017. Applying to college: Do information deficits lower the likelihood of college-eligible students from less-privileged families to pursue their college intentions?: Evidence from a field experiment. *Social Science Research* Available online 29 April 2017.

- Engle, J., 2008. Postsecondary access and success for first-generation college students. In: American Academic Volume Three. Pell Institute for the Study of Opportunity in Higher Education, pp. 25–48, boulder: Westview Press.
- Hainmueller, J., 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20 (1), 25–46.
- Hainmueller, J., Xu, Y., 2013. Ebalance: A stata package for entropy balancing. *Journal of Statistical Software* 54 (7), 1–18.
- Heine, C., 2010. Soziale ungleichheiten im zugang zu hochschule und studium. Arbeitspapier 213, Hans B?ckler Stiftung.
- Heineck, G., Riphahn, R. T., 2009. Intergenerational transmission of educational attainment in germany - the last five decades. *Jahrb?cher f?r National?konomie und Statistik* 229 (1), 36–60.
- Herber, S. P., 2015. The role of information in the application for merit-based scholarships: Evidence from a randomized field experiment. BERG Working Paper 95, University of Bamberg.
- Hoxby, C., Turner, S., 2014. Expanding college opportunities for high-achieving, low income students. Stanford Institute for Economic Policy Research (SIEPR) Discussion Paper No. 12-014.
- Iacus, S., King, G., Porro, G., 2009. Cem: Software for coarsened exact matching. *Journal of Statistical Software* 30 (9), 1–27.
- Jensen, R., May 2010. The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics* 125 (2), 515–548.
- Kerr, S. P., Pekkarinen, T., Sarvim?ki, M., Uusitalo, R., September 2015. Post-secondary education and information on labor market prospects: A randomized field experiment. IZA Discussion Paper Series No. 9372.
- Loyalka, P., Song, Y., Wei, J., Zhong, W., Rozelle, S., 2013. Information, college decisions and financial aid: Evidence from a cluster-randomized controlled trial in china. *Economics of Education Review* 36, 26–40.
- Manski, C. F., 1993a. Adolescent econometricians: How do youth infer the returns to schooling? In: Clotfelter, C. T., Rothschild, M. (Eds.), *Studies of Supply and Demand in Higher Education*. University of Chicago Press, Ch. 2, pp. 43–60.
- Manski, C. F., 1993b. Dynamic choice in social settings. *Journal of Econometrics* 58 (1-2), 121–136.
- McGuigan, M., McNally, S., Wyness, G., 2016. Student awareness of costs and benefits of educational decisions: Effects of an information campaign. *Journal of Human Capital* 10 (4), 482–519.

- Middendorff, E., Apolinarski, B., Poskowsky, J., Kandulla, M., Netz, N., 2013. Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2012. 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das HIS-Institut für Hochschulforschung.
- Mueller, W., Pollak, R., 2015. Bildung und soziale Mobilität in Deutschland. AStA Wirtschafts- und Sozialstatistisches Archiv 9 (1), 5–26.
- Nguyen, T., 2008. Information, role models and perceived returns to education: Experimental evidence from Madagascar, mimeo.
- Oreopoulos, P., Dunn, R., 2013. Information and college access: Evidence from a randomized field experiment. *The Scandinavian Journal of Economics* 115 (1), 3–26.
- Oxoby, R. J., 2008. Skill uncertainty and social inference. *Labour Economics* 15 (3), 400 – 405.
- Schnitzlein, D. D., 2016. A new look at intergenerational mobility in Germany compared to the US. *Review of Income and Wealth* 62 (4), 650–667.
- Spangenberg, H., Beuße, M., Heine, C., 2011. Nachschulische Werdegänge des Studienberechtigtenjahrgangs 2006. Dritte Befragung der studienberechtigten Schulabgänger/innen 2006 - 3 1/2 Jahr nach Schulabschluss im Zeitvergleich. HIS: Forum Hochschule 18|2011.
- Steiner, V., Wrohlich, K., 2012. Financial student aid and enrollment in higher education: New evidence from Germany. *The Scandinavian Journal of Economics* 114 (1), 124–147.

Tables

Table 1: Covariate balance by treatment status

	All		Non-academic background		Academic background	
	Control Group Mean	Treatment Group Difference	Control Group Mean	Treatment Group Difference	Control Group Mean	Treatment Group Difference
Intended college enrollment	0.792 (0.028)	-0.028 (0.047)	0.749 (0.032)	-0.050 (0.065)	0.865 (0.027)	-0.002 (0.040)
Individual characteristics						
Age	18.591 (0.155)	-0.128 (0.231)	18.739 (0.138)	-0.237 (0.252)	18.340 (0.189)	0.064 (0.242)
Female	0.588 (0.034)	0.012 (0.060)	0.599 (0.037)	0.054 (0.048)	0.570 (0.040)	-0.051 (0.091)
Migration background	0.465 (0.058)	0.055 (0.121)	0.501 (0.071)	0.065 (0.138)	0.402 (0.043)	0.047 (0.115)
Non-academic background	0.629 (0.030)	-0.026 (0.051)				
Performance and skills						
German Grade	8.775 (0.211)	-0.154 (0.348)	8.467 (0.217)	0.000 (0.379)	9.296 (0.251)	-0.436 (0.376)
Math Grade	8.034 (0.190)	0.344 (0.324)	7.845 (0.178)	0.368 (0.283)	8.353 (0.301)	0.280 (0.486)
Cognition test (verbal)	9.796 (0.251)	0.295 (0.495)	9.413 (0.276)	0.145 (0.459)	10.447 (0.241)	0.454 (0.533)
Cognition test (figural)	11.014 (0.186)	0.159 (0.301)	10.749 (0.213)	0.518 (0.407)	11.463 (0.172)	-0.433 (0.314)
School type						
School type I (<i>Gymnasium</i>)	0.307 (0.126)	-0.001 (0.204)	0.278 (0.116)	0.054 (0.211)	0.357 (0.148)	-0.089 (0.214)
School type II (<i>Integrierte Sekundarschule</i>)	0.368 (0.133)	0.008 (0.220)	0.377 (0.131)	-0.010 (0.218)	0.352 (0.143)	0.037 (0.233)
School type III (<i>berufliches Gymnasium</i>)	0.325 (0.127)	-0.007 (0.214)	0.345 (0.127)	-0.044 (0.209)	0.291 (0.134)	0.053 (0.231)
Perceived returns						
Unemployment risk smaller	0.402 (0.026)	-0.004 (0.042)	0.382 (0.033)	0.010 (0.055)	0.438 (0.038)	-0.028 (0.049)
Prospects for well paid job higher	0.712 (0.020)	0.004 (0.026)	0.712 (0.026)	0.016 (0.035)	0.711 (0.034)	-0.014 (0.049)
Relative income premium (bachelor's/vocational degree)	1.525 (0.035)	-0.021 (0.051)	1.546 (0.039)	-0.043 (0.053)	1.487 (0.042)	0.018 (0.078)
Life time income higher	0.644 (0.017)	0.008 (0.035)	0.66 (0.021)	-0.048 (0.044)	0.607 (0.025)	0.096* (0.052)
N	658	330	414	199	244	131
N (total)	988		613		375	

Notes: This table presents control group means and treatment-control differences for the analyzed samples used to investigate short-run treatment effects measured 2-3 months after the information treatment. Means and mean differences are derived by separately regressing each variable on the treatment group indicator, i.e. $X_i = \alpha + \beta Z_s + \varepsilon_i$, where X_i represents the variable in the left most column and Z_s is an indicator variable for treatment status as obtained from randomization. Standard errors are clustered at the school level and shown in parentheses. Source: Best Up, wave 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Relevance of information by educational background

	Non-academic background		Academic background	
	No Intention	Intention to enroll (Difference)	No Intention	Intention to enroll (Difference)
Information source				
Information source: Parents/Family	0.870 (0.024)	0.003 (0.027)	0.902 (0.032)	0.054* (0.031)
Parents/Family helpful as information source (1-5)	3.674 (0.124)	-0.175 (0.128)	3.609 (0.182)	0.375* (0.201)
Costs				
Feeling well informed about university education	0.236 (0.040)	0.129*** (0.043)	0.200 (0.082)	0.204** (0.075)
Problem: obtaining info	0.264 (0.045)	0.025 (0.052)	0.294 (0.054)	-0.053 (0.052)
Not/hardly dealt with financing possibilities	0.608 (0.051)	-0.165*** (0.049)	0.636 (0.065)	-0.203** (0.078)
Perceived cost burden high	0.593 (0.041)	-0.148*** (0.052)	0.314 (0.054)	-0.040 (0.055)
Perceived returns				
Unemployment risk smaller	0.274 (0.047)	0.152*** (0.045)	0.353 (0.050)	0.087 (0.051)
Prospects for well paid job higher	0.571 (0.028)	0.201*** (0.034)	0.580 (0.075)	0.147* (0.072)
Relative income premium (bachelor's/vocational degree)	1.531 (0.036)	0.002 (0.044)	1.392 (0.067)	0.117 (0.092)
Life time income higher	0.580 (0.033)	0.095** (0.038)	0.588 (0.084)	0.061 (0.085)
N	613		375	

Notes: This table depicts the relevance of information separately for students from different educational backgrounds. It presents mean and mean differences based on regressing each variable on an indicator variable for intended college enrollment, i.e. $X_i = \alpha + \beta y_i^{(t0)} + \varepsilon_i$, where X_i represents the variable in the left most column and $y_i^{(t0)}$ is an indicator variable for pre-treatment intended college enrollment. Standard errors are clustered at the school level and shown in parentheses. The numbers reflect the share of students whose answers are in accordance with the statements listed in the left column. Source: Best Up, wave 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Treatment effect on perceived labor market returns

	Treatment effect	Control Group Mean	N
Unemployment risk is smaller	0.096** (0.048)	-0.087	[966]
Prospects for finding a well paid job are higher	0.079*** (0.028)	0	[952]
Relative income premium (bachelor's/vocational degree)	0.050 (0.052)	-0.027	[752]

Notes: This tables presents the effect of information provision on students' perceived labor market returns to university education. Each row represents a separate regression with the outcome specified in the most left column. Estimates are based on Equation 3, i.e. using changes in subjective labor market benefits as dependent variables. In all estimations we control for age, gender, migration background, school types, standardized math and German grades as well as two measures for cognitive skills measured by a verbal and a matrix test. The number of observations is shown in square brackets in the utmost right column and varies across estimations due to item non-response. Standard errors in parentheses are clustered at the school level. Source: Best Up, wave 1 and 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Treatment effect on the change in students' intended college enrollment

	All		Non-academic background		Academic background	
<u>2/3 months after the intervention</u>						
Change in intended college enrollment	0.031 (0.024)	0.030 (0.025)	0.080** (0.033)	0.082** (0.035)	-0.047 (0.029)	-0.056* (0.029)
<i>Control group mean</i>	<i>-0.026</i>		<i>-0.034</i>		<i>-0.012</i>	
N	988		613		375	
<u>1 year after the intervention</u>						
Change in intended college enrollment	0.059** (0.029)	0.058* (0.033)	0.080** (0.032)	0.078** (0.035)	0.041 (0.051)	0.033 (0.047)
<i>Control group mean</i>	<i>-0.045</i>		<i>-0.023</i>		<i>-0.082</i>	
N	827		510		317	
Controls	No	Yes	No	Yes	No	Yes

Notes: This table presents the effect of information provision on students' change in intended college enrollment 2/3 months after the intervention as well as one year after the intervention based on Equation 3. In all estimations school types are included as control variables. In columns 2, 4, and 6 additional controls include age, gender, migration background, standardized math and German grades as well as two measures for cognitive skills measured by a verbal and a matrix test. Standard errors in parentheses are clustered at the school level. Source: Best Up, wave 1-3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Adjustments to pre-treatment intended college enrollment

	2-3 months after the intervention		1 year after the intervention	
	Non-academic background	Academic background	Non-academic background	Academic background
(1) Upward adjustment	0.053** (0.022)	-0.041** (0.020)	0.015 (0.026)	0.019 (0.032)
<i>Control group mean</i>	0.063	0.049	0.113	0.048
(2) Downward adjustment	-0.029 (0.026)	0.015 (0.024)	-0.063** (0.025)	-0.015 (0.030)
<i>Control group mean</i>	0.097	0.061	0.136	0.130
(3) Stable intention	-0.024 (0.033)	0.026 (0.034)	0.048 (0.037)	-0.004 (0.040)
<i>Control group mean</i>	0.841	0.889	0.751	0.822
N	613	375	510	317

Notes: This table shows how students adjusted their pre-treatment intended college enrollment in response to the information provision. All estimations are based on Equation 3 and include the following control variables: school type, age, gender, migration background, standardized math and German grades as well as two measures for cognitive skills measured by a verbal and a matrix test. Standard errors in parentheses are clustered at the school level. Source: Best Up, wave 1-3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A: Additional tables and figures

Table A.1: Comparison of schools in recruitment sample and *Best Up* sample (in %)

School and district characteristics	All schools	Preferred schools	Replacement schools	Contacted schools	<i>Best Up</i> schools
School type:					
General high schools (<i>Gymnasium</i>)	53.8	33.3	55.0	41.0	33.3
Comprehensive high schools (<i>Integrierte Sekundarschule</i>)	31.7	36.7	30.0	35.9	33.3
Vocational high schools (<i>berufliches Gymnasium</i>)	14.4	30.0	15.0	23.1	33.3
District information:					
Share of low educated aged 25 and older	17.1	23.0	20.3	22.3	21.2
School information:					
Cohort size (number of students)	104	109	94	108	102
Share of students with migration background	13.9	18.4	15.8	18.4	17.6
Share of female students	52.4	53.4	49.9	53.9	54.2
Number of schools	104	30	20	39	27

Notes: This table presents descriptive characteristics of university track high schools in Berlin from which the final *Best Up* sample of schools was drawn. The share of low educated individuals aged 25 and older ranges from 7.1% to 30.3% across Berlin and all 104 schools. For the sample of contacted schools this range goes from 9.1% to 30.3% and from 12.2% to 30.3% in the *Best Up* sample. Source: Federal statistical office of Berlin-Brandenburg (Amt für Statistik Berlin-Brandenburg 2011/12); and regional data from Amt für Statistik Berlin-Brandenburg (2011).

Table A.2: Covariate balance by treatment status based on the baseline sample

	Baseline sample			
	Control Group Mean		Treatment Group Difference	
Intended college enrollment	0.77	(0.021)	-0.016	(0.042)
Individual characteristics				
Age	18.704	(0.152)	-0.131	(0.248)
Female	0.570	(0.031)	-0.007	(0.054)
Migration background	0.518	(0.063)	0.035	(0.120)
Non-academic background	0.623	(0.028)	-0.022	(0.052)
Performance and skills				
German Grade	8.558	(0.192)	-0.061	(0.329)
Math Grade	7.676	(0.158)	0.253	(0.275)
Cognition test (verbal)	9.464	(0.274)	0.350	(0.500)
Cognition test (figural)	10.681	(0.205)	0.212	(0.284)
School type				
School type I (<i>Gymnasium</i>)	0.280	(0.116)	0.021	(0.200)
School type II (<i>Integrierte Sekundarschule</i>)	0.376	(0.129)	0.013	(0.219)
School type III (<i>berufliches Gymnasium</i>)	0.345	(0.127)	-0.034	(0.207)
Perceived returns				
Unemployment risk smaller	0.390	(0.017)	-0.000	(0.030)
Prospects for well paid job higher	0.700	(0.017)	0.008	(0.020)
Relative income premium (bachelor's/vocational degree)	1.542	(0.032)	0.009	(0.040)
Life time income higher	0.622	(0.017)	0.002	(0.028)
N	1059		519	
N (total)	1578			

Notes: This table presents control group means and treatment-control differences for the baseline sample. Means and mean differences are derived by separately regressing each variable on the treatment group indicator, i.e. $X_i = \alpha + \beta Z_s + \varepsilon_i$, where X_i represents the variable in the left most column and Z_s is an indicator variable for treatment status as obtained from randomization. Standard errors are clustered at the school level and shown in parentheses. Source: Best Up, wave 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Covariate balance by treatment status based on sample one year after the intervention

	All		Non-academic background		Academic background	
	Control Group Mean	Treatment Group Difference	Control Group Mean	Treatment Group Difference	Control Group Mean	Treatment Group Difference
Intended college enrollment	0.796 (0.024)	-0.037 (0.047)	0.751 (0.029)	-0.054 (0.066)	0.870 (0.025)	-0.017 (0.049)
Individual characteristics						
Age	18.553 (0.160)	-0.086 (0.244)	18.704 (0.138)	-0.195 (0.262)	18.303 (0.206)	0.101 (0.267)
Female	0.588 (0.029)	0.029 (0.057)	0.588 (0.032)	0.072 (0.048)	0.587 (0.039)	-0.036 (0.097)
Migration background	0.459 (0.058)	0.075 (0.119)	0.494 (0.071)	0.058 (0.143)	0.400 (0.043)	0.105 (0.103)
Non-academic background	0.624 (0.034)	-0.022 (0.051)				
Performance and skills						
German Grade	8.831 (0.226)	-0.198 (0.380)	8.569 (0.244)	-0.115 (0.414)	9.263 (0.244)	-0.357 (0.402)
Math Grade	8.057 (0.198)	0.410 (0.339)	7.959 (0.191)	0.317 (0.308)	8.218 (0.324)	0.539 (0.533)
Cognition test (verbal)	9.937 (0.224)	0.165 (0.415)	9.554 (0.263)	0.113 (0.404)	10.572 (0.228)	0.189 (0.486)
Cognition test (figural)	10.971 (0.197)	0.233 (0.256)	10.664 (0.229)	0.615* (0.339)	11.481 (0.190)	-0.389 (0.258)
School type						
School type I (<i>Gymnasium</i>)	0.327 (0.131)	-0.013 (0.208)	0.296 (0.121)	0.038 (0.212)	0.380 (0.155)	-0.095 (0.220)
School type II (<i>Integrierte Sekundarschule</i>)	0.358 (0.134)	0.054 (0.223)	0.380 (0.134)	0.026 (0.225)	0.322 (0.140)	0.100 (0.231)
School type III (<i>berufliches Gymnasium</i>)	0.315 (0.125)	-0.041 (0.196)	0.325 (0.123)	-0.064 (0.193)	0.298 (0.138)	-0.004 (0.213)
Perceived returns						
Unemployment risk smaller	0.409 (0.027)	-0.036 (0.044)	0.38 (0.034)	-0.021 (0.056)	0.449 (0.043)	-0.062 (0.054)
Prospects for well paid job higher	0.704 (0.022)	-0.000 (0.029)	0.704 (0.029)	-0.005 (0.041)	0.703 (0.035)	0.007 (0.053)
Relative income premium (bachelor's/vocational degree)	1.519 (0.034)	-0.025 (0.054)	1.531 (0.037)	-0.043 (0.058)	1.498 (0.050)	0.006 (0.088)
Life time income higher	0.634 (0.020)	0.014 (0.046)	0.653 (0.023)	-0.040 (0.061)	0.602 (0.032)	0.099* (0.050)
N	553	274	345	165	208	109
N (total)	827		510		317	

Notes: This table presents control group means and treatment-control differences for the analyzed samples used to investigate treatment effects on intended college enrollment one year after the information treatment. Means and mean differences are derived by separately regressing each variable on the treatment group indicator, i.e. $X_i = \alpha + \beta Z_s + \varepsilon_i$, where X_i represents the variable in the left most column and Z_s is an indicator variable for treatment status as obtained from randomization. In addition to the marginally statistically significant difference regarding the perception on lifetime income (see also Table 1), in this sample treated students from non-academic backgrounds score slightly higher on the figural cognition test. However, the absolute size of the difference corresponds to around a fifth of a standard deviation, which we consider negligible. Standard errors are clustered at the school level and shown in parentheses. Source: Best Up, wave 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Descriptive statistics by students' parental educational background

	Non-academic background	Academic background	Difference
Intended college enrollment	0.732	0.864	-0.132***
Individual characteristics			
Age	18.662	18.363	0.300***
Female	0.617	0.552	0.065*
Migration background	0.522	0.419	0.103**
Performance and Skills			
German Grade	8.467	9.144	-0.677***
Math Grade	7.965	8.450	-0.485*
Cognition test (verbal)	9.460	10.605	-1.145***
Cognition test (figural)	10.917	11.312	-0.395*
School types			
School type I (<i>Gymnasium</i>)	0.295	0.325	-0.030
School type II (<i>Integrierte Sekundarschule</i>)	0.374	0.365	0.008
School type III (<i>berufliches Gymnasium</i>)	0.331	0.309	0.022
Information sources			
Internet	95.402	94.879	0.524
Friends	89.256	88.679	0.577
Central study counseling	36.913	38.859	-1.946
Job information center/Employment agency	60.738	52.162	8.576**
Parents/Family	87.273	94.879	-7.606***
Parents/Family helpful as information source (1-5)	3.545	3.935	-0.389***
Costs			
Feeling well informed about university education	0.331	0.377	-0.046
Problem: obtaining information	0.282	0.248	0.034
Not/hardly dealt with financing possibilities	0.485	0.458	0.027
No scholarships known	0.367	0.281	0.085*
Perceived cost burden high	0.484	0.280	0.205***
Perceived returns			
Unemployment risk smaller	0.385	0.428	-0.043
Prospects for well paid job higher	0.717	0.706	0.011
Relative income premium (bachelor's/vocational degree)	1.532	1.493	0.040
Life time income higher	0.650	0.641	0.009
N	613	375	

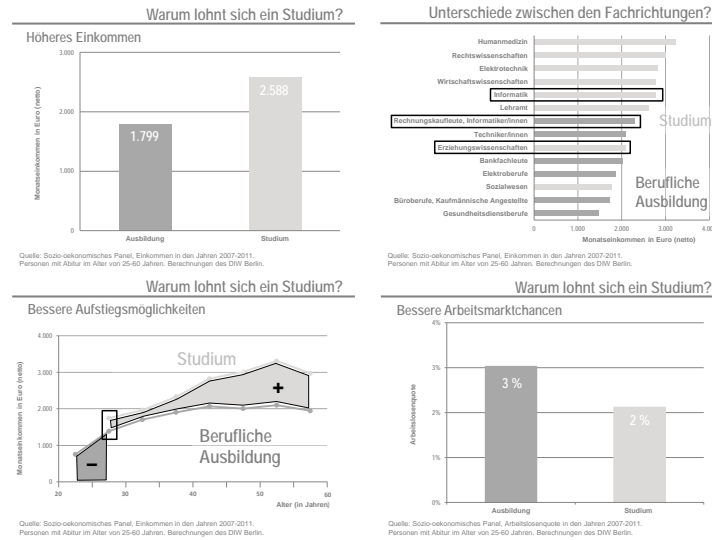
Notes: This tables documents differences of students by educational background with regard to various characteristics. Differences are based on a two-sided t-test. Source: Best Up, wave 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Treatment effect on intended college enrollment: Mean comparison

	All		Non-academic background		Academic background	
<u>2/3 months after the intervention</u>						
Intended college enrollment	0.019 (0.026)	0.015 (0.025)	0.055* (0.031)	0.055* (0.030)	-0.046 (0.031)	-0.052* (0.031)
<i>Control group mean</i>	<i>0.792</i>		<i>0.749</i>		<i>0.865</i>	
N	988		613		375	
<u>1 year after the intervention</u>						
Intended college enrollment	0.035 (0.032)	0.030 (0.034)	0.041 (0.036)	0.041 (0.037)	0.029 (0.045)	0.022 (0.043)
<i>Control group mean</i>	<i>0.796</i>		<i>0.751</i>		<i>0.870</i>	
N	827		510		317	
Controls	No	Yes	No	Yes	No	Yes

Notes: This table presents the effect of information provision on students' intended college enrollment 2/3 months after the intervention as well as one year after the intervention based on Equation 2. In all estimations school types and pre-treatment intended college enrollment ($y_i^{(t_0)}$) are included as control variables. In columns 2, 4, and 6 additional controls include age, gender, migration background, standardized math and German grades as well as two measures for cognitive skills measured by a verbal and a matrix test. Standard errors in parentheses are clustered at the school level. Note that the treatment effect for students from non-academic backgrounds 2-3 months after the treatment is very similar in terms of magnitude, as the differences in pre-treatment intended college enrollment shown in Table 1. The difference in statistical significance stems from the decrease in residual variance in the treatment models, where we control for school types and students' pre-treatment intended college enrollment. Source: Best Up, wave 1-3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: Presentation slides used in the information treatment: Examples



Note: This figure provides examples of the presentation slides used in the information treatment, in which college and vocational education were compared conditional on having earned the *Abitur*. The slide in the upper left panel shows the difference in average earnings between individuals with a university degree (*Studium*) and a vocational degree (*Ausbildung*). The upper right panel shows earnings differences across different university majors and occupations in vocational education. The slide in the lower left panel shows a comparison of lifetime earnings with a university degree and a vocational degree, while in the lower right panel the unemployment rate for individuals with a university degree and a vocational degree are depicted.

Appendix B: Detailed discussion of the sensitivity analyses

In this Section we discuss the various robustness tests summarized in Section 5.4 and shown in Table B.1 in more detail. Columns 1 and 2 of Table B.1 report the robustness of the estimates with respect to short-run intended college enrollment (2-3 months after the intervention) and columns 3 and 4 for enrollment intentions one year after the information workshop. The first row of Table B.1 shows the main estimates as a reference point.

Accounting for attrition. Attrition is a common problem in RCTs that rely on survey data to measure the outcome of interest. Generally, attrition poses a threat to the estimation of the treatment effect only if there are non-random differences between treatment and control groups. This may result in biased estimates of the treatment effect. As outlined in Section 3 differential attrition is of no concern for our estimations. However, even if there is no differential attrition between treatment and control groups, we might still be worried if certain types of students are over- or underrepresented in the analyzed sample and treatment effects vary for these groups. For example, if the information intervention is more (less) effective for underrepresented groups, our estimates will be biased downward (upward). It is well known that individuals with certain characteristics are more likely to respond to surveys than others. Comparing attritors and non-attritors in our sample shows that students who are younger, female, have no migration background, and have higher math grades, German grades, or have higher scores on cognitive measures are more likely to participate in the post-treatment surveys.

In order to investigate how this affects our estimates, we predict the subgroup-specific probability to participate in each of the post-treatment surveys and rerun our estimation using the inverse of these probabilities as sampling weights. To predict post-treatment participation we use the same set of covariates as in our main specifications as well as pre-treatment intentions to enroll in college. Additionally, we include a binary variable indicating whether we have valid data on the contact information (email, address, phone) that was used to contact students for the post-treatment survey and collected in the pre-treatment survey.⁴² Using inverse probability weights slightly decreases our point estimate for students from non-academic backgrounds, whereas it increases (in absolute values) for students from academic backgrounds in the short run. Nevertheless, effect size and statistical significance remain mostly comparable.

Accounting for covariate imbalance. In Table 1 we show the covariate balance for the sample that we use for our analysis as well as for the subgroups by parental educational background. Most of the covariate differences are statistically insignificant. However, irrespective of the statistical significance of these differences the size of some of the differences may trigger concerns about the comparability of treatment and control group students. In order to increase the similarity of treated and control group students we rerun our estimation using entropy balancing weights (Hainmueller, 2012; Hainmueller and Xu, 2013). Entropy balancing reweights control group students such that a set of pre-specified moment conditions are equal across treatment status. This procedure selects the set of weights that satisfies the pre-specified moment conditions but remains as close as possible to uniform weights (Hainmueller, 2012). In our estimation we require the first moment of all variables included as controls to be the same

⁴²This information was updated in the first post-treatment survey.

in the control group as in the treatment group. As shown in row 3 of Table B.1 the short-run result for students from academic backgrounds is unaffected. For students from non-academic backgrounds we find a similar treatment effect one year after the information intervention and a slightly smaller effect with similar significance level in the short run.

Dealing with non-compliance. As pointed out in Section 4 one school in the treatment group did not receive the information workshop. For our estimations, as presented in Section 5, we therefore follow a two stage least squares approach. Yet, in order to assess the sensitivity of the results to the non-compliant school, we run two further analyses. We first examine the treatment effect without the non-compliant school in the sample and then estimate a specification in which we assign the non-compliant school to the control group. Compared to our main specification, the changes in short-run point estimates are only marginal (see row 4 and 5 of Table B.1), while the effect size and the significance level decrease slightly for students from non-academic backgrounds one year after the information intervention.

Wild cluster bootstrap t-procedure. In our main specification we cluster standard errors at the school level. To account for the small number of clusters (27 schools), we also apply the wild cluster bootstrap-t procedure to calculate alternative p-values as suggested by Cameron and Miller (2015).⁴³ The corrected p-values are depicted in row 6 of Table B.1 and do not change our conclusions.⁴⁴

Including school fixed effects. Given the design of the RCT in which entire schools were randomized, it is advisable to include school fixed effects to account for any time invariant school level omitted variables that might affect students' enrollment intentions. In order to strengthen our results, we estimate a difference-in-difference type of regression, which allows us to additionally include school fixed effects.⁴⁵ Other than in our main specification, we do not use predicted treatment status (as obtained from Equation 1) but use T_s as the treatment group indicator instead, where the non-compliant school is assigned to the control group. Table B.1 shows that with school fixed effects and further control variables the short-run effects remain very similar. However, the effect for students from non-academic backgrounds one year after the intervention decreases in size and is no longer statistically significant (p-value: 0.145).

Discarding selected schools. In Table B.1 we further investigate how sensitive our results are to considering specifics of the project setup, i.e. student level participation and geographical proximity of schools.

⁴³We use the Stata command *clustse* (provided by Andrew Menger) and specify the wild option (1000 replications), which implements the program *cgmwildboot* created by Judson Caskey (available from his website at: <https://sites.google.com/site/judsoncaskey/data>).

⁴⁴Although we have not found any other studies implementing the wild cluster bootstrap t-procedure in a two stage least squares (2SLS) setting, we calculate the corrected p-values in the second stage of the 2SLS approach. Nonetheless, we are confident to report these values, since we also calculated the corrected p-values in the sample without the non-compliant school as well as in the sample with the reassigned non-compliant school; in all cases the statistical significance level in the sample of students from non-academic families decreases to 10%; for students from academic families the statistical significance level of 10% only holds in the case of reassigning or excluding the non-compliant school but not for the specification shown here.

⁴⁵We estimate the following equation: $y_{ist} = \alpha + \beta_1(T_s * post_{it}) + \beta_2 post_{it} + X'_i \beta_3 + \kappa_s + \varepsilon_{ist}$, where y_{ist} is the intended college enrollment of student i in school s at time t ($t=0,1,2$, i.e. before, 2-3 months or one year after the treatment). T_s is the treatment indicator and $post_{it}$ indicates whether it is the post-treatment period. X_i is a vector of additional (pre-treatment) individual level controls (as defined before) and κ_s represents school fixed effects. As before, we cluster standard errors at the school level.

First, although the information workshop as well as the pre-treatment survey were conducted during school hours, participation for students was still on a voluntary basis due to strict data protection regulations in Germany. As a result, we observe school-level differences in response rates to the pre-treatment survey. If students' decision to participate is correlated with intended college enrollment, our results will be biased. Thus, to limit the possibility that our results are driven by selection into (student-level) survey participation, we drop those schools with the five lowest school-level response rates from our sample. As shown in row 8, this yields almost no changes regarding short-run estimates. The effect for students from non-academic backgrounds one year after the intervention, however, decreases in size but remains statistically significant at the 10% significance level.

And second, given that the project's focus was to conduct its RCT in districts with a high share of low educated individuals in Berlin, one concern may be that students of treatment schools potentially inform control school students of the information workshop leading to spillover effects. We rerun our estimations excluding all students from control schools that are close, i.e. within a two kilometer radius of a treatment school (see row 9 of Table B.1). For students from non-academic backgrounds the change in the short-run point estimate is minimal and the effect one year after the intervention even increases; for students from academic family backgrounds the short-run estimate slightly increases in absolute values implying a downward bias (in absolute values) of our main estimate of 1.2 percentage points.

Defining educational background. To cope with missing information on students' educational background we made several assumptions in order to approximate students' background (described in Section 3), thereby minimizing the loss of observations. Therefore, as a last robustness check, we investigate whether a potential misclassification of students affects our estimates. We restrict our sample to students for whom we have *complete* information on parental education only. This approach slightly changes the estimated effect sizes, but increases the statistical significance level of our estimates for students from non-academic backgrounds (p-value < 0.01). For students from a non-academic background the effect increases by around one percentage point in the short run and by around two percentage points one year later; whereas the short-run estimate for students of parents with a college degree remains nearly identical to our main specification (see row 10 of table B.1).

Overall, the sensitivity analyses confirm our findings as presented in Section 5.2. While for students from academic family backgrounds the information intervention decreases intended college enrollment in the short run, this effect is no longer detectable based on enrollment intentions one year after the intervention. For students from non-academic families, however, the information treatment increases intended college enrollment, both in the short run and one year later, i.e. shortly after graduating from high school.

Table B.1: Sensitivity Analyses

	Change in students' intended college enrollment			
	2-3 months after the intervention		1 year after the intervention	
	Non-academic background	Academic background	Non-academic background	Academic background
(1) Main	0.082** (0.035) [613]	-0.056* (0.029) [375]	0.078** (0.035) [510]	0.033 (0.047) [317]
(2) Inverse probability weighting	0.076** (0.038) [613]	-0.085** (0.039) [373]	0.063* (0.036) [510]	0.014 (0.047) [315]
(3) Entropy balancing	0.076** (0.033) [613]	-0.059* (0.031) [375]	0.080** (0.035) [510]	0.024 (0.053) [317]
(4) Without non-compliant school	0.084** (0.036) [606]	-0.057** (0.027) [358]	0.068* (0.033) [503]	0.023 (0.046) [303]
(5) Reassigning non-compliant school to control group	0.085** (0.036) [613]	-0.057** (0.027) [375]	0.062* (0.034) [510]	0.021 (0.046) [317]
(6) Wild cluster t-procedure <i>corrected p-value</i>	0.082* <i>0.068</i> [613]	-0.056 <i>0.104</i> [375]	0.078** <i>0.046</i> [510]	0.033 <i>0.508</i> [317]
(7) Including school fixed effects	0.083** (0.036) [613]	-0.050* (0.028) [375]	0.064 (0.043) [510]	0.013 (0.056) [317]
(8) Without low response schools	0.083** (0.036) [559]	-0.056* (0.029) [324]	0.065* (0.034) [461]	0.022 (0.050) [271]
(9) Without potential 'spill-over schools'	0.077** (0.030) [532]	-0.068*** (0.026) [352]	0.085** (0.037) [446]	0.009 (0.045) [299]
(10) Strict definition on educational background	0.095*** (0.034) [567]	-0.058** (0.027) [367]	0.106*** (0.034) [474]	0.033 (0.048) [311]

Notes: This table shows how sensitive our estimates are to different specifications. All estimates are based on Equation 3 and include the following control variables: school type, age, gender, migration background, standardized math and German grades as well as two measures for cognitive skills measured by a verbal and a matrix test. The number of observations is shown in square brackets. Standard errors in parentheses are clustered at the school level. To address the issue of school level randomization we include school fixed effects in row 7 of this table; this specification estimates the following equation: $y_{ist} = \alpha + \beta_1(T_s * post_{it}) + \beta_2 post_{it} + X'_i \beta_3 + \kappa_s + \varepsilon_{ist}$, where y_{ist} is the intended college enrollment of student i in school s at time t ($t=0,1,2$, i.e. before, 2-3 months or one year after the treatment). T_s is the treatment indicator and $post_{it}$ indicates whether it is the post-treatment period. X_i is a vector of additional (pre-treatment) individual level controls (as defined before) and κ_{ps} represents school fixed effects. For this specification we do not use the predicted treatment status but use the treatment indicator where the non-compliant school is assigned to the control group. Source: Best Up, wave 1-3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.