Economics of Education Review 95 (2023) 102409

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Preferred field of study and academic performance ☆

Francesco Berlingieri [a,*], André Diegmann [b], Maresa Sprietsma [c]

[a] European Commission, Joint Research Centre (JRC), Italy
[b] IWH Halle, ZEW Mannheim, Germany
[c] Hochschule der Bundesagentur fuer Arbeit (HdBA), Germany

ARTICLE INFO

JEL classification:
I21
I23
J24

Keywords:
Field of study
Preferences
Academic performance
University dropout

ABSTRACT

This paper investigates the impact of studying the first-choice university subject on dropout and switching field of study for a cohort of students in Germany. Using detailed survey data, and employing an instrumental variable strategy based on variation in the local field of study availability, we provide evidence that students who are not enrolled in their preferred field of study are more likely to change their field, delay graduation and drop out of university. The estimated impact on dropout is particularly strong among students of low socio-economic status and is likely to be driven by lower effort and motivation.

1. Introduction

Enrollment in tertiary education has increased significantly over the past few decades in OECD countries (OECD, 2020). In Germany, the share of an age cohort starting tertiary education increased from 37% in 2005 to 57% in 2020 (Destatis, 2021). A considerable number of students, however, drop out of higher education before completing a degree (see Vossensteyn et al., 2015). Reducing dropout rates in higher education, therefore, remains an important policy goal in many European countries. Descriptive evidence shows that besides high school grades, enrollment in the preferred field of study is an important determinant of successful degree completion (Heublein et al., 2017; Larsen et al., 2013; and Lassibille & Gomez, 2008).

In European countries, students apply for a specific field of study before entering higher education. Since different fields of study have substantially different labor market payoffs (Kirkeboen et al., 2016), and students have limited information on their ability to perform well in different subjects when entering higher education, the choice of the field is an important and complex decision to make for young adults. In addition, not all individuals are able to enroll in their preferred field of study. In our sample, 30% of students are not enrolled in their first choice field of study.

One reason for not enrolling in the preferred field is that students typically attend university in a location near to their secondary school and that most universities offer only a subset of fields. In Germany, only 27% of first year students attend a university that is in another federal state than the state in which they completed secondary education. As a result of low geographical mobility, the regional supply of study places is likely to restrict educational decisions. Distance to the nearest university, for instance, was shown to significantly affect the decision to enroll at university rather than to pursue vocational training (Spiess & Wrohlich, 2010). A second reason for not being able to enroll in the preferred field of study is that some fields have strict admission rules and admit only a limited number of students. As a result, a descriptive analysis of the effect of enrolling in the preferred field of study on

☆ We thank the editor, Daniel Jones, an anonymous referee, Max Bach, Lorenzo Cappellari, Bernd Fitzenberger, Marta Golin, Guido Heineck, Sarah McNamara, Jan Marcus, Friedhelm Pfeiffer, Stephan Thomsen, as well as participants at the LERN 2019, IWAEE 2019, NEPS 2019, the ZEW workshop on the Economics of Higher Education, and seminars at ZEW for valuable comments. Christina Leroch, Anne Kuhn and Patrick Schulze provided outstanding research assistance. We gratefully acknowledge financial support from the German Federal Ministry of Education and Research's framework program "Academic success and dropout phenomena" (BMBF, research grant 01PX16018A). This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 5 - Students, doi:10.5157/NEPS:SC5:10.0.0. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. Opinions expressed herein are those of the authors only and do not reflect the views of, or involve any responsibility for, the institutions to which they are affiliated. We are responsible for all remaining errors.

* Corresponding author.
E-mail addresses: francesco.berlingieri1@ec.europa.eu (F. Berlingieri), Andre.Diegmann@iwh-halle.de (A. Diegmann), Maresa.Sprietsma@arbeitsagentur.de (M. Sprietsma).

https://doi.org/10.1016/j.econedurev.2023.102409
Received 11 July 2022; Received in revised form 18 March 2023; Accepted 22 April 2023
Available online 18 May 2023

dropout rates can be misleading. Students who enroll in their preferred field are likely to be of higher ability, more geographically mobile and more motivated than those that do not enroll in their preferred field of study.

In this paper, we investigate how a mismatch between the preferred field of study and the actual field of enrollment affects academic performance. We measure academic performance in several ways, such as by the probability of dropping out of the program, the probability of graduating on time, and the probability of exiting higher education entirely. Our identification strategy exploits regional variation in the availability of fields of study. We use an instrumental variable approach, in which the instrument consists of the supply of study places in the preferred field of study in German universities, weighted by the distance to the place of high school graduation. By accounting for detailed field preferences at the individual level, we compare students with the same preferences who are exposed to a different availability of study places in the area surrounding the location where they graduated from secondary education.

Our analysis is based on survey data from the starting cohort 5 of the National Educational Panel (Blossfeld et al., 2011). This survey follows the cohort of students who entered German universities in 2010 over time and provides detailed information on study paths, academic outcomes, as well as on preferences and aspirations. Compared to the related studies based on administrative sources (see, e.g., Heinesen, 2018), we can thus provide more detailed information about how studying the preferred field impacts the academic performance and on the channels beyond the observed patterns.

Our IV estimates show that students able to enroll in their preferred field of study because of its regional availability perform significantly better compared to students who cannot study their preferred field. OLS results first show that studying the preferred field reduces the probability of program-specific dropout by 8% points. The causal effect on program dropout implied by our instrumental variable strategy turns out to be even larger: studying the preferred field reduces program dropout by 20% points. Similarly, we document significant and sizeable effects on the probability to change the field of study (−20% points), timely graduation (25% points) and university dropout (−11% points). We provide an economic interpretation by calculating counterfactual rates for our sample of students. If all students were able to enroll in their preferred field of study, program dropout would decrease by 8%, timely graduation would increase by 9% and dropout rates would decrease by 23%. These results are robust to several robustness checks. Most notably, we show robustness related to concerns about sample attrition, parents' judgement about the subject choice, students' personality traits, and different definitions of the instrument.

We further show that males and students from a low socio-economic background are over-represented in our complier group. Consistent with these results, the estimated impact of attending the preferred field of study on academic performance is particularly strong among students with lower socio-economic status. Interestingly, males are driving the results on program dropout while there are no gender differences for the results on leaving university without a degree. Finally, we present evidence suggesting that passion and motivation are important drivers behind our results. Students who enroll in their preferred field show a higher identification with their field of study, report a higher satisfaction with their studies and spend four hours more per week studying two to three years after university entry.

Causal evidence on the returns to admission to the preferred field of study is scarce. Using administrative data on applications to different fields of study in Denmark, Heinesen (2018) finds that being admitted to the first-choice field of study increases the probability of completing a master's degree in this field, but has no significant impact on completing a master's degree overall. Daly et al. (2022) show that studying the preferred field of study yields higher returns on the labor market than studying the second choice field of study. We find that, in Germany, studying the preferred field leads to lower university dropout. This could be an important channel behind the positive wage returns found in the literature.

Our study relates to several other threads of the empirical literature on higher education. Firstly, it relates to the literature on the choice of fields of study or the choice of college majors. Expected earnings (Arcidiacono et al., 2012; Kirkeboen, 2012), parents educational level (Boudarbat & Montmarquette, 2009), risk aversion (De Paola & Gioia, 2012), and individual taste (Hilmer & Hilmer, 2012; Wiswall & Zafar, 2015) are shown to play a significant role in the choice of major or field of study. Evidence on the role of the local supply of study fields is scarce. One exception is the study by Denzler and Wolter (2010), who show that, in Switzerland, the local supply of fields of study has an impact on the chosen field of study.

Secondly, we contribute to the evidence on the determinants of university dropout and delayed graduation. High school grades and perceived ability are the most important determinants of successful tertiary education completion (Danilowicz-Oesele et al., 2017; Lassibille & Gomez, 2008; Stinebrickner & Stinebrickner, 2014). Students face uncertainty as to their skills and abilities when entering university. Credit constraints (Lassibille & Gomez, 2008; Modena et al., 2020; Stinebrickner & Stinebrickner, 2008) and non-cognitive skills (De Paola & Scoppa, 2015) are also related to the probability of graduation. We show causally that enrolling in the preferred field of study is a further important driver of academic performance.

The rest of the article is organized as follows. In Section 2 we present the data and descriptive statistics. In Section 3 we introduce our instrumental variable approach, while in Section 4 we assess the quality of the instrument. We then present estimation results in Section 5, and we investigate potential mechanisms in Section 6.

2. Data and descriptive statistics

2.1. Data sources and sample selection

The Starting Cohort First-Year Students of the National Educational Panel Study (NEPS) is a panel survey of university entrants for the winter semester 2010/2011 covering about 18,000 individuals (see Blossfeld et al., 2011, for more details). Individuals are first surveyed shortly after entering university and generally on a half-year basis afterwards. However, many students participate only in a subset of all surveys. The data include information about study paths and academic outcomes, education and family background, personality traits, preferences, their school history as well as the district (*Kreis*) of secondary school graduation and the district of their university.

In the first wave, students are asked to report their preferred field of study. Students are asked to list their first and second preferred field by answering the question: "Which two fields of study best meet your preferences?". This information is then coded to the two-digit field of study classification from the Federal Statistical Office (Destatis), which consists of 59 fields. Given that some of these fields are very small or are relatively broad groupings of other fields in the list, we rearrange these 59 fields to build a consistent classification of 35 narrow fields (see appendix Table A.1 for the list of narrow fields and the details about the field reclassification).[1]

We then use this classification to construct a binary variable indicating whether students are enrolled in their first preferred field when entering university. Note that students are surveyed shortly after university entry and their stated preferences might be influenced by

[1] We regroup 8 small fields and drop 1 field with no relative preference information (Industrial Engineering with Economics Focus), 5 interdisciplinary fields (broad groupings of several other fields in the list), as well as further 9 small fields with less than 20 observations in the sample. This turns our sample in regrouping fields for about 2.5% of the students and dropping 6.4% of the observations.

ex-post rationalizations or available information at the time of the interview. Ideally one would observe preferences before university entry. To test for the consistency of our data with existing data, we compare the share of students studying their preferred field in our sample to that resulting from a different survey, the DZHW 2008 school graduate survey, where field preferences are surveyed in the final year of high school. Using sample weights, we find that 71% of individuals study their preferred field when using the DZHW data compared to 70% in our data. The fact that these shares are very similar between the two datasets suggests that preference adjustments shortly after the start of university enrollment are not likely to play an important role in the present analysis.

The information about the preferred field of study is only available for about 45% of students in the survey. Among the remaining 55% of students, 13% are not asked about their preferred field of study at all, 7% are asked but do not answer any question about field of study preferences and 35% provide an answer to another question asking whether they are studying their preferred field but do not provide information on which field that is. In our baseline sample, we only include students who explicitly specify their first preferred field. However, for students who state to be studying the preferred field and do not provide the exact field, we can impute the actual field of study as the first preference. Appendix Table A.11 shows that the main results are not substantially different when including students with imputed preferred field in the estimation.

Our sample is based on all students who are interviewed during their first semester of university. We then exclude all individuals with missing information on the main control variables (educational and parental background information) and construct our outcome variables for students who are interviewed again four years after university entry.[2] Section 5.4 provides evidence that the preferred field of study is uncorrelated with the probability of staying in the sample. Our results are also robust when constructing the outcome variables of program dropout and field of study change 1.5 years (survey wave 3) or 2.5 years (survey wave 5) after university entry. The final sample covers 3916 students.

We consider several outcome variables. Program dropout is a binary variable that is equal to one for individuals who leave the study program they first enrolled in without completing it. Field change before completion is a binary variable that is equal to one for individuals who change their field of study prior to completing this main program of first enrollment. It consists, therefore, of students who dropout from their first program but subsequently enroll in a different narrow field of study. Timely graduation is a binary variable equal to one for individuals who successfully graduate within the standard duration of university programs. Finally, university dropout is a binary variable that is equal to one for students who leave university without earning a degree and are not studying at the time of the interview.

The NEPS sample covers university entrants of different study programs. We distinguish between three main program types: universities of applied sciences (*Fachhochschulen*), teaching programs (*Lehramt*) and non-teaching programs at traditional universities (*Universitäten* and *Technische Universitäten*). Universities of applied sciences concentrate on more applied programs in a subset of fields. Moreover, some school-leaving exams grant access only to those universities and not to traditional ones. It is also important to separately consider teaching and non-teaching programs since our subject classification does not enable such a distinction. Students in programs leading to a teaching accreditation often study two or more fields at the same time.[3] Most university entrants in the sample are enrolled in bachelor programs typically lasting 3 years. Almost all students in universities of applied sciences are enrolled in bachelor programs. As regards the students in the regular track, 82% are enrolled in bachelor programs and 18% in programs leading to a state examination (e.g. law, medicine).[4] Among students in the teaching track, about one-third are enrolled in a bachelor degree and two-thirds are enrolled in a teacher-specific degree (*Staatsexamen*).

In order to account for regional-specific conditions, we exploit data on the labor market region of high school graduation for 2010 which we gather from official statistics of the Federal Statistical Office and the Federal Employment Agency.[5] We include population size and density, the unemployment rate, per capita GDP, the share of regional migrants and employment shares in seven broad industries. In order to account for the overall demand and supply of higher education, we also include the number and the share of high school graduates with a university entrance examination, the number of first-year university students, as well as the number of narrow fields of study offered in the region.

Table 1
Number of fields of study by district and by labor market region.

	Mean	Std. Dev	Min	Max	25%	50%	75%
Districts	12.0	10.4	1	35	4	7	21
Labor market regions	17.3	10.9	1	35	6	18	28

Note: Summary statistics for the 190 districts (out of 402) and 112 labor market regions (out of 141) with a university. Data on first-year students in 2010 by university and field of study from the Federal Statistical Office.

2.2. Descriptive statistics: Regional supply of field of studies and probability to attend the preferred field

As a result of the large expansion in higher education in the past 50 years in Germany, almost half of regional districts (190 out of 402) now have a university or applied university. However, there was still large variation in the number of field of studies offered at the regional level in 2010. Table 1 shows descriptive statistics of the number of fields of study available at the district and at the labor market region level. We observe that, on average, only 12 narrow fields out of 35 were offered in districts with a university. At the level of labor market regions, the average number of fields offered in regions with a university increases to 17. However, it is noticeable that large spatial variation exists also at this regional level of aggregation. While 28 fields are available in labor market regions in the 75th percentile, only 6 fields of study are available in labor market regions of the 25th percentile. While bigger cities offer a large number of fields, only Berlin and Munich offer programs in all 35 narrow field of studies (see appendix Fig. A.1 for a detailed map on the number of field offered by German districts).

The availability of certain fields in the region of school graduation is thus likely to influence whether students enroll in their preferred field of study. Our empirical strategy exploits these regional differences that are arguably exogenous conditional on preferences and relevant regional characteristics (see Section 3).

In the following section, we look at the overall availability of study places by field of subject and the share of students attending their preferred field. Aggregating field of study availability at the broad field

[2] About 32% of students are not interviewed in wave 9 which corresponds to four years after university entry in the NEPS data.

[3] Note that the data oversamples students of teaching programs. These amount to circa 31% of the students in the sample. According to official statistics, the share of university entrants in teaching programs was approximately 8% in 2010.

[4] Few students in the sample (less than 0.2%) are enrolled in traditional *Diplom* and *Magister* programs, which we consider equivalent to state examination programs because of their similar length.

[5] We follow the classification of Kosfeld and Werner (2012) specifying 141 labor market regions in Germany, which are aggregations of the 402 districts based on commuting patterns.

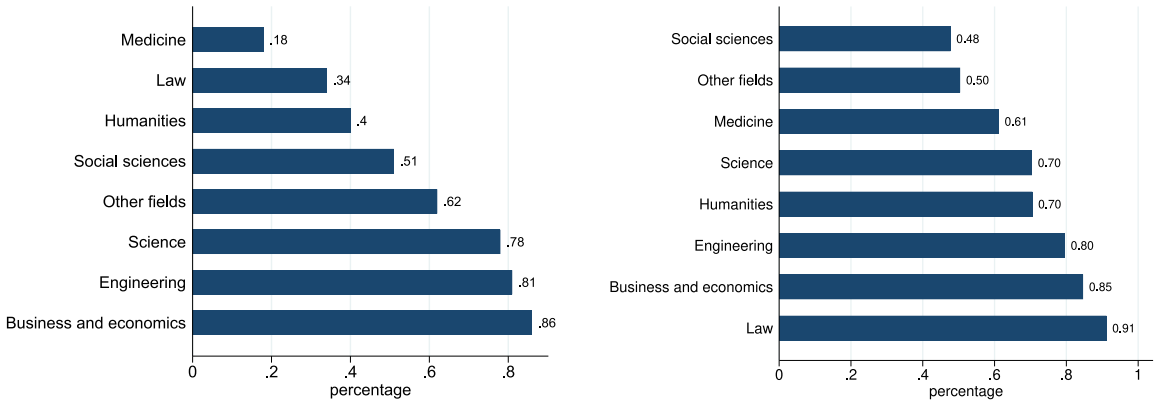

Fig. 1. Field supply and share of students in preferred field by broad field classification.
Notes: The figure shows the share of districts where a university offer the respective field in Panel A conditional on having a university in the district. Out of 401 districts, 190 districts have at least one university or applied university. Panel B shows the share of students in the preferred field.

classification, Panel A of Fig. 1 shows the share of districts with at least one university offering a specific field of study. For example, out of the 190 districts with a (applied) university, 86% of districts offer study places in business and economics, whereas only 18% of the districts offer study places in medicine.

In our sample, on average 67% of students (unweighted) are able to enroll in their first preferred subject.[6] Panel B of Fig. 1 shows the share of students in their preferred field by broad preferred field classification. As can be observed in Fig. 1, there seems to exist a correlation between the overall availability of the field across districts and the share of students that study this field as their preferred field. Students willing to study business and economics, science, engineering and humanities groups are very likely to enroll in these programs as their preferred field.

On the contrary, students who want to study medicine, social sciences and other fields more often end up studying a different field. For medicine, the lower than average share is also driven by the strict admission rules. For fields within social sciences and other fields, the lack of supply at universities nearby is likely to be a more important driver of these figures. Law represents an interesting exception as a there is a rather low share of districts that offer study places, but about 91% of the students study law as their preferred field. This might be related to a relatively lower demand. Among all students, about 4% study law compared to 22% in engineering or 18% in business and economics. Table A.2 provides an overview on the number of districts that offer the respective field and the number of students who study the field by broad field of study classification.

Table 2 presents summary statistics for students that enroll in their preferred field and those who do not. It also shows a *t*-test of difference in means between the two groups.

The table shows that the two groups are different with respect to several variables other than the broad field preference. In particular, students not studying the preferred field are more likely to be enrolled in teaching programs. The reasons for this are not stricter admission

Table 2
Summary statistics by studying preferred field.

	Studying preferred field		*t*-test	
	No	Yes	diff.	(t-stat.)
Panel A: University type				
University of applied science	0.16	0.16	−0.00	(−0.01)
Teaching program	0.49	0.33	0.16***	(9.65)
Bachelor degree	0.62	0.65	−0.03	(−1.74)
Panel B: Demographics and school background				
Age (university entrance)	20.64	20.73	−0.10	(−1.03)
Female	0.63	0.59	0.05**	(2.77)
Born in Germany	0.96	0.96	−0.00	(−0.27)
School final grade (z-score)	−0.16	0.21	−0.36***	(−10.79)
Repeated grade	0.15	0.11	0.04**	(3.17)
Middle secondary track	0.29	0.29	0.01	(0.41)
Applied university track	0.07	0.08	−0.01	(−1.10)
High school: abroad	0.03	0.04	−0.01	(−1.01)
Apprenticeship before 2011	0.14	0.16	−0.02	(−1.38)
Panel C: Parental background				
Mother: foreign born	0.11	0.10	0.00	(0.28)
Mother: employed at age 15	0.73	0.72	0.00	(0.19)
Mother: tertiary degree	0.25	0.29	−0.04**	(−2.58)
Mother: no upper secondary qual.	0.07	0.05	0.02*	(2.18)
Father: foreign born	0.10	0.11	−0.01	(−0.56)
Father: employed at age 15	0.94	0.95	−0.00	(−0.34)
Father: tertiary degree	0.36	0.41	−0.05**	(−3.05)
Father: no upper secondary qual.	0.04	0.03	0.01	(1.85)
Panel D: Regional characteristics				
A-level school graduates: number	4143.72	4289.97	−146.25	(−1.03)
A-level school graduates: share	30.10	30.67	−0.56*	(−2.51)
First-year students	6153.80	6502.49	−348.69	(−1.48)
Number of narrow fields	27.79	28.52	−0.74	(−1.24)
Population density	861.82	876.95	−15.13	(−0.53)
Unemployment rate (in %)	7.17	7.37	−0.21*	(−1.99)
GDP per capita	32.22	31.80	0.42	(1.63)
Regional migrants (in %)	39.80	39.32	0.49	(1.89)
Observations	1300	2616		

Note: The table refers to the sample of 3916 students with non-missing information on the preferred field of study and the main variables in the analysis. Regional characteristics refer to the labor market region of high school graduation.

[6] When using sampling weights, the share of students enrolled in the first preferred subject is 70%, as reported in Section 2.1.

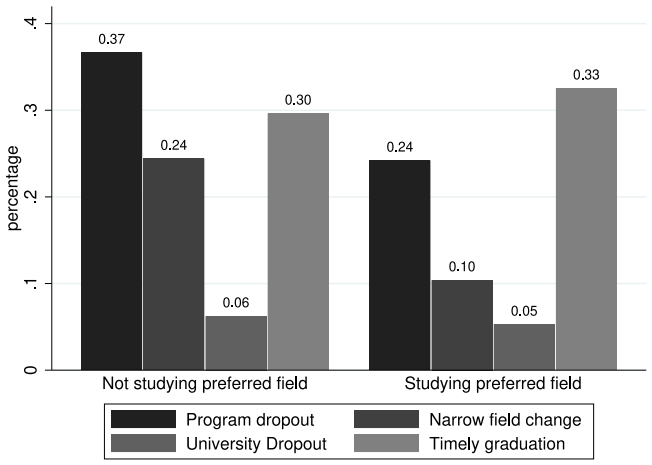

Fig. 2. Preferred field and study success.

rules in the latter programs, given that non-teaching programs in traditional universities do restrict admission more often and have stricter requirements. One potential explanation is that teaching programs do not cover all potential subjects. Moreover, students in these programs are less likely to move to another region or federal state for university. The share of individuals enrolling to university in the same federal state of high school is 80% for teaching programs and slightly less than 70% for non-teaching programs at traditional universities. A further explanation is that the specific subject studied is less important for students in teaching programs. To account for these differences we control for the type of program in the empirical analysis. Moreover, we show that the results are robust to excluding students from teaching programs and universities of applied science.

Table 2 further shows that students not studying their preferred field are more likely to be women, have on average worse grades in the high school examination and are more likely to have repeated a grade during school. Moreover, their parents are less likely to have a university degree and are more likely to have no upper secondary qualification. Regional level characteristics such as GDP per-capita and the unemployment rate do not differ significantly between field of study fulfillment. Importantly, regions are not significantly different in terms of the number of university students and the number of narrow field of study offered. However, students studying their preferred field went to schools in regions with a slightly higher share of school graduates with a traditional university entrance examination (A-level or *Abitur*).

Fig. 2 shows the unconditional means of the four key outcome variables for students studying the preferred field and those that do not. Among individuals studying their preferred field, 24% leave the first university program without completion, 10% change the narrow field of study and 6% switch to a field outside the broad subject group. These outcomes are on average at least 10% points larger for those not studying the preferred field. The difference between the two groups in terms of university dropout is smaller on average. While 5% of individuals studying the preferred field leave the university without earning a degree, this share amounts to 6% for those with unfulfilled study preferences.

3. Empirical strategy

3.1. Instrumental variable: regional offer of preferred field

Identifying the causal effect of attending the preferred field of study on academic performance implies solving several selection and endogeneity issues. More able students (e.g. students with better school grades or higher cognitive or non-cognitive skills) or students with a higher intrinsic motivation are less likely to dropout of university, but are also more likely to study their preferred field of study. For some subjects with excess demand, grades are used as entry requirement. Students with better grades thus have a higher probability of attending their preferred field of study.

The failure to control for such characteristics will lead to an omitted variable bias, which is likely to bias upward the estimates. Moreover, our measure for studying the preferred field of study is arguably subject to measurement error due, for instance, to mistakes in the reported preferred or actual fields or in the classification of specific fields into broader groups. This would lead to a possible underestimation of the true effect.

To deal with these issues, we employ an instrumental variable strategy, where we use the regional variation in the subject-specific availability of university spots. Most secondary school graduates in Germany have a strong preference for attending a university near to where their family and friends live. Spiess and Wrohlich (2010), for instance, show that the distance to the closest university affects the likelihood of enrolling in tertiary education. In our data, 42% of first year students enroll in a university that is located in the district of secondary school graduation, 52% stay in the same labor market region, while 73% stay in the same federal state. Similarly to Kamhöfer et al. (2019), we expect that the probability to study a given university subject will depend both on whether the subject is offered in universities nearby and on how many university places are offered by local universities in the given subject.

In Germany, admission to higher education is mostly decentralized and administered by universities themselves.[7] In practice, for many subjects at most universities every applicant with a university entrance qualification (*Abitur*) is granted access. Whenever demand exceeds supply of college places, universities select applicants based on university-specific criteria. The final grade of the university entrance qualification is by far the most common and important admission criterion.

We construct our instrumental variable at the level of regional districts (*Kreise*) and preferred field of study. We thereby assign the instrument for each individual based on the first preferred field using the consistent 35 categories, and the region of secondary school graduation. Our instrument combines information on the distance to university offering the preferred field with the number of places in the preferred field (measured by the number of first-year students in the field) as:

$$Z_{ijl} = \sum_{k=1}^{402} \left[K(dist_{ik}) \times students_{jkl} \right] \quad (1)$$

where $students_{jkl}$ represent the number of university entrants of field j in program type l and district k, weighted by the Gaussian kernel distance $K(dist_{ik})$ between the centroid of the district of school graduation of individual i and the centroid of district of the university k. The kernel gives a high weight to close universities and a rather small weight to distant ones. The instrument Z_{ijl} sums up the weighted number of available college spots in the preferred field over all districts with a university offering that field. We use a bandwidth of 150 km, meaning that $K(dist_k) = exp[-(dist_k^2)/(150^2)]$. To provide an example, universities in the same district of the high school receive a weight of 1, whereas those at 100 km distance receive a weight of 0.64 and those at 150 km distance a weight of 0.37. The weight of universities further away is very small. For instance, for universities at 300 km distance the weight is smaller than 0.02. The number of students is reported in thousands to avoid reporting very small coefficients in the first stage

[7] Only for medicine, pharmacy, and veterinary medicine there are nationwide admission restrictions.

Table 3
Regional availability of preferred field by field fulfillment.

	Mean	Std. Dev	Min	Max
Panel A: Studying preferred field				
IV: preferred field availability	1.399	1.523	0.003	8.954
Preferred field in district	0.37	0.48	0	1
Distance to closest spot	23.0	24.6	0	183.1
Panel B: Not studying preferred field				
IV: preferred field availability	0.651	0.948	0	8.068
Preferred field in district	0.31	0.46	0	1
Distance to closest spot	26.5	25.5	0	181.9

Note: NEPS SC5, final sample of 3916 students.

equation. Our results are robust to the choice of the bandwidth (see Table A.11, columns (2)–(3)).

We compute the instrument within a given type of higher education program l (i.e. university of applied sciences, teaching programs or non-teaching programs at traditional universities), in which individual i is enrolled. This is because admission requirements and occupational goals differ considerably by program type. We thus assume that students have a preference for different study fields within a specific type of higher education. We indeed find that the strength of the instrument is significantly enhanced when computed by type of higher education program in comparison to an instrument based on the number of students in all higher education types pooled together. In Section 5.4, we show that the results are robust to excluding the students who are studying a program not offering their preferred field (see Table A.11, column(6)).

Table 3 presents the summary statistics of the instrumental variable depending on whether students are enrolled in the preferred field in their first semester. It provides insights on the unconditional correlation between the instrument and the endogenous variable. While the mean of the instrumental variable is 1.4 for those studying the preferred field, it is about half as large for those who are not able to fulfill their preferences in terns of field of study. The table further presents descriptive results for some of the key ingredients of the instrument. Those studying their preferred field have a higher likelihood that this field was available in the region of their school graduation. Moreover, the distance to the closest university offering the field was also shorter on average for these individuals.

3.2. Empirical model

In our main specification we estimate the following system of equations by two-stage least squares, with the second stage as:

$$y_{ijl} = \beta preferred_{ijl} + x_i'\gamma + u_l + p_j + \epsilon_i \quad (2)$$

where y_{ijl} is a binary indicator denoting the outcome variable, i.e. program dropout, field change, timely graduation, university dropout, etc. The variable $preferred_{ijl}$ is a binary variable indicating whether individual i studies his/her preferred field j in the first semester of university and x_i' is a vector of individual characteristics including demographics (age, gender and country of birth), educational background variables (secondary school grade, as well as its interaction with the federal state and the school type, grade repetition, the attended school track, apprenticeship qualification), parental background (highest qualification and employment status), federal state of school graduation fixed effects, and regional characteristics at the level of the labor market region of school graduation (number of university entrants, number and share of high school graduates with a university entrance examination, population size and density, the unemployment rate, per capita GDP and employment shares by industry). u_l is an indicator of the university program type (standard program at a regular university, teaching program at a regular university or university of applied sciences). Finally, p_j is an indicator of the preferred field of study.

Due to the discussed endogeneity concerns, the first stage can be written as:

$$preferred_{ijl} = z_{ijl}\gamma + x_i'\pi + u_l + p_j + \mu_i \quad (3)$$

where z_{ijl} is the instrumental variable that varies at the level of the preferred field, the university program type and the district of school graduation. Therefore, standard errors are clustered at the district level through the empirical analysis.

Our empirical strategy exploits variation in field of study availability at the regional level. We include a large vector of control variables and we include fixed effects for the preferred field of study, p_j. Specifically, we measure field preferences following the 35 narrowly defined fields. With this empirical strategy, we thus compare high school graduates with the same preferences for a given field of study in regions that are structurally similar (e.g. number of school graduates, the number of university entrants, the population density and the industry structure), but that differ in the distance to study places in the preferred field of study. For instance, we compare high school graduates with a preference for political science in regions with many nearby possibilities for studying political science to those in regions that are located relatively far away from universities offering political science.

It is important to underline that the estimates based on the instrumental variable approach are local average treatment effects (LATE) and only hold for those individuals who study a given subject because of the regional availability of the subject. Individuals who have a very strong interest in a given subject and no mobility or financial constraints will probably attempt to move to a region where this subject is offered. They could also decide to postpone their study to a later period or not study at all if the subject is not offered in their region. Results from IV regressions are not informative for these individuals. However, our estimates are interesting for those students that do face mobility constraints.

4. Assessment of the instrument

4.1. Instrument relevance

Fig. 3 provides a graphical representation of our first stage. In the background, we show the histogram of the distribution of our instrument controlling for detailed field of study preferences. The histogram reveals a wide spread in field availability conditional on detailed study preferences. The mean of our instrument is 1.15 with a standard deviation of 1.40. The figure further plots the probability of studying the preferred field as a function of our instrument. We plot estimates from a local linear regression which represents a flexible form of Eq. (3). The likelihood of studying the preferred field is monotonically increasing with field availability. At the upper end of the instrument distribution, the likelihood of studying the preferred field levels off. To provide an example, a student situated in the lower decile of the instrument distribution has a 50% probability of studying their preferred field compared to about 75% probability for a student in the upper decile of the distribution.

Table 4 shows the first stage estimation results. We provide point estimates of preferred field availability (measured at the region of school graduation) on studying the preferred field for different empirical specifications. Columns (1) to (6) present the results from a linear probability model regression, while in column (7) the average marginal effects from a probit regression is shown. Column (1) shows the effect of the instrumental variable, i.e. the regional availability in the preferred field, conditional on basic demographic characteristics (female indicator, age, born in Germany). Consistent with Fig. 3, the instrumental variable has a significant and positive effect on the probability to studying the preferred field.[8] The coefficient of 0.087 implies

[8] As shown in the regression tables of the results, the F-statistics on the excluded instruments is 111 in our preferred specification.

Table 4
First-Stage estimates of studying preferred field on field availability.

	LPM (1)	LPM (2)	LPM (3)	LPM (4)	LPM (5)	LPM (6)	Probit (7)
Field availability	0.087***	0.088***	0.087***	0.095***	0.095***	0.099***	0.112***
	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)
Program type	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Narrow field preference	No	Yes	Yes	Yes	Yes	Yes	Yes
Demographic characteristics	No	No	Yes	Yes	Yes	Yes	Yes
School background	No	No	No	Yes	Yes	Yes	Yes
Parental background	No	No	No	No	Yes	Yes	Yes
Regional characteristics	No	No	No	No	No	Yes	Yes
Observations	3916	3916	3916	3916	3916	3916	3916
R^2	0.081	0.176	0.179	0.216	0.218	0.225	

Note: Control variables included are demographics (age, gender and country of birth), educational background variables (final school grade and its interaction with the federal state and the school type, grade repetition, school track background, apprenticeship qualification), parental background (highest qualification and employment), university type, preferred field fixed effects, federal state of school graduation fixed effects, as well as characteristics of the labor market region of school graduation (number of university entrants, number and share of high school graduates with a university entrance examination, population size and density, the unemployment rate, per capita GDP and employment shares by industry). Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

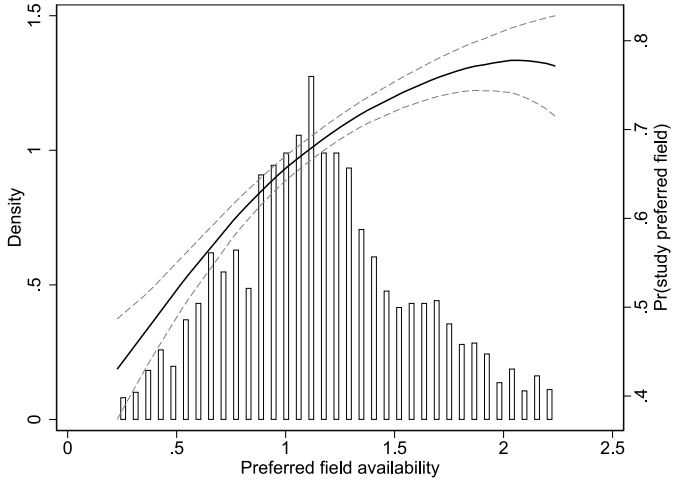

Fig. 3. First stage.
Notes: First-stage graph of studying preferred field on preferred field availability. The probability of studying the preferred field is plotted on the right Y-axis against constructed instrument. The plotted values are mean-standardized residuals from regressions on field of study preferences fixed effects. The solid line shows a local linear regression of studying preferred field on preferred field availability. Dashed lines show 90% confidence intervals. The histogram shows the density of the instrument along the left Y-axis (top and bottom 5% excluded).

that an increase of 1000 slots in the preferred field in the region of school graduation (i.e. an increase of 85% on average) would increases the probability of studying the preferred field by 8.7% points.

Important to our identification strategy, column (2) adds field preference-fixed effects. Compared to column (1), the coefficient of the instrument remains unaffected. The first stage coefficient is also unaffected when including basic demographics (column (3)) and is very similar when school background characteristics are included (column (4)). For the instrument to be valid, it must be uncorrelated to predetermined variables such as parental background information (e.g. parental qualification) and regional characteristics (e.g. local unemployment rate). If this is the case, the first stage coefficient should not change after adding these variables as further controls. Columns (5) and (6) provide the results after including additional information on school and parental background, as well as regional characteristics. Adding these controls hardly increases the first stage coefficient indicating that, conditional on preferences, the preferred field availability is rather unrelated to parental and local conditions. Finally, column (7) reports average marginal effects from a probit model. The coefficients are very similar to those of the linear probability model.

Appendix Table A.3 provides detailed estimation results of the first stage (Table 4) on the control variables. As previously discussed, one of the most important determinants of studying the preferred field is the final grade of the school-leaving qualification. Its coefficient is large and significant in all specifications indicating that better school grades increase the probability of studying the preferred field. Females are more likely to study their preferred field, when conditioning on the program type and preferred field of study. Moreover, the coefficient of the variables on the study program and preferred field confirm what is shown in the descriptive results. Students in teaching programs or who are studying at universities of applied sciences are less likely to enroll in their preferred field. Relative to students willing to study a field within the humanities group, individuals aiming to study medicine, social sciences or other, less common, fields are less likely to have their preferences fulfilled.

4.2. Exclusion restriction

Interpreting the IV coefficient as a causal effect requires an exclusion restriction. In our setting, local field availability should affect students outcomes and performance measures only through the preferred field of study take-up and not directly. A major concern with our empirical approach might be related to local labor market conditions or shocks at the local level. If, for example, local companies work closely together with local universities, this could influence the availability of certain fields and simultaneously affect student performance because these students might have better opportunities to gather practical experiences and apply theoretical knowledge. We condition in our main specification on important local labor market characteristics, such as the employment share in 1-digit industries. Appendix Table A.6 shows that the main estimated coefficient on program dropout is not affected after conditioning on regional characteristics, while the precision of the estimate slightly improves.

We further test for randomness of our instrument with respect to pre-determined variables by providing balancing results in Table 5. A correlation between our instruments and pre-determined individual and regional characteristics would indicate a potential violation of the exclusion restriction. Such correlations with observables might hint to

Table 5
Balancing Test.

	Dep. variable: Preferred field		Dep. variable: Instrument	
	Coefficient (1)	st. error (2)	Coefficient (3)	st. error (4)
Age (university entrance)	0.009**	0.004	−0.001	0.007
Female	0.039**	0.015	0.041	0.030
Born in Germany	−0.010	0.044	−0.128	0.084
School final grade (z-score)	0.107**	0.046	0.029	0.104
Repeated grade	−0.014	0.022	0.025	0.041
Middle secondary track	−0.013	0.017	0.009	0.028
Applied university track	−0.180	0.115	0.284	0.284
High school: abroad	−0.049	0.036	0.034	0.074
Apprenticeship before 2011	0.016	0.028	−0.037	0.061
Parental background				
Mother: foreign born	−0.020	0.031	−0.056	0.054
Mother: employed at age 15	−0.013	0.017	0.039	0.024
Mother: tertiary degree	0.014	0.017	−0.022	0.028
Mother: no upper secondary qual.	−0.056	0.035	0.007	0.050
Father: foreign born	0.056*	0.032	0.069	0.047
Father: employed at age 15	0.011	0.033	−0.056	0.059
Father: tertiary degree	0.018	0.015	0.025	0.025
Father: no upper secondary qual.	−0.014	0.040	−0.089	0.077
Regional characteristics				
A-level school graduates: number	−0.000	0.000	0.000	0.000
A-level school graduates: share	−0.001	0.003	0.008	0.006
First-year students (log)	0.002	0.011	0.038	0.027
Number of narrow fields (log)	−0.008	0.026	−0.007	0.062
Population density	0.000	0.000	0.000	0.000
Unemployment rate (in %)	0.003	0.008	0.008	0.016
GDP per capita	0.006*	0.003	−0.008	0.008
Regional migrants (in %)	−0.001	0.002	−0.001	0.004
Regional employment by sector				
Manufacturing	−0.012***	0.005	0.014	0.011
Low-skilled services	−0.013***	0.004	0.002	0.009
Professional services	−0.013	0.008	−0.027	0.020
Business services	−1.210	1.083	3.290	2.093
Finance/insurance	−0.012	0.010	0.032	0.022
Research intensive services	0.002	0.003	0.017**	0.007
Creative sector	0.012	0.013	0.024	0.023
		p-value		*p*-value
Joint *F*-test	2.40	0.002	1.01	0.442
Observations	3916		3916	

Note: The table shows results of OLS regressions with as dependent variable a binary variable denoting whether students are enrolled in their preferred field (column 1) and the instrument (column 3). All regressions include narrow field preferences, program type indicators and federal state of school graduation fixed effects. Regional characteristics and regional employment refer to the labor market region of high school graduation. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the fact that our instruments might be correlated with unobservable variables as well.

First of all, column (1) shows the results of a regression of the determinants of studying the preferred field of study. The probability of studying the preferred field is particularly correlated with age, gender, and high school performance which indicates that the preferred field dummy is potentially endogenous even after conditioning on observables. Column (3) provides the results by regressing the instrument on the observed characteristics. As borne out by the table, coefficient estimates provide strong evidence that our instrument is uncorrelated with a large set of observable variables. In particular coefficient estimates of age, gender and final school grades do not significantly correlate with the constructed instrument. The overall correlation with the instrumental variable is low, which is supported by the joint *F*-test at the bottom of the table. Column (3), however, shows that one out of 32 coefficients is significant. That is, the employment share in the research intensive service sector in the region of high school graduation positively correlates with the instrument. The probability, however, that one or more out of 32 coefficients will be significant at the 5% level by chance is 80.6%. We therefore consider estimation results in Table 5 along with results in Table A.6 as strong evidence for random assignment.

4.3. Monotonicity

In our setting, monotonicity requires that a student who enrolls in the preferred field – although he or she lives in an area where there are only few possibilities to study his or her preferred field – would also study the same field if he/she were located in an area where there are many possibilities to study the preferred field. This holds vice versa for students who do not study the preferred field. This assumption must be made to ensure the LATE interpretation if treatment effects are heterogeneous. That is, the 2SLS estimand represents the average causal effect among students who would have enrolled in their preferred field if they had been located in a region where the preferred field was offered closer by (Bhuller et al., 2020).

This assumption has implications that can be tested. In particular, first stage estimates should be non-negative for any subgroup. We follow Bhuller et al. (2020) and construct our instrument based on the full sample and provide estimates of the first stage on specific sub-samples. The empirical specification corresponds to column (4) of Table 4. Appendix Table A.4 provides the first stage results separated by gender, socio-economic background and broad field of study groups. Consistent with the monotonicity assumption, each reported coefficient is large, positive, and statistically different from zero.

Table 6
Compliance.

	$P[X = x]$ (1)	First stage (2)	$P[X = x \mid I_{1i} > I_{0i}]$ (3)	$\frac{P[X=x \mid I_{1i} > I_{0i}]}{P[X=x]}$ (4)
Male	0.397	0.108	0.433	1.091
Female	0.603	0.090	0.548	0.909
Low SES	0.531	0.106	0.568	1.070
High SES	0.469	0.092	0.436	0.930
Humanities & Social Sciences	0.346	0.283	0.989	2.858
Business & Law	0.116	0.021	0.024	0.206
STEM	0.362	0.205	0.749	2.069
Medicine (incl. other fields)	0.177	0.095	0.170	0.960
Overall	1.000	0.099	1.000	1.000

Note: The table reports characteristics of compliers. This includes the distribution of the population $P[X = x]$ and the distribution of compliers $P[X = x|I_{1i} > I_{0i}]$ calculated as first-stage coefficient times population share divided by the overall first-stage coefficient. The last statistic shows the relative likelihood of an individual belonging to a subgroup, in the complier group compared to the overall subgroup population.

4.4. Compliance

In order to learn more about our sample of students, we calculate the share of always takers, never takers and compliers by following Dahl et al. (2014). Similar to the definition of Imbens and Rubin (1997) and Abadie (2003), a complier is defined as a student who could not enroll in his or her preferred field because study places for this field are too far away or that could enroll in the preferred field because study places were available nearby. In order to calculate the share of compliers with a continuous instrument, we define different values of our instrumental variable. Specifically, we define $\bar{z}$ as the maximum of our instrument (high field of study availability), whereas $\underline{z}$ is the minimum (low field of study availability). Based on a linear specification of the first stage, we can calculate the share of compliers as $\pi_c = \hat{\gamma}(\bar{z} - \underline{z})$, where $\hat{\gamma}$ comes from the first stage equation. Always takers are students who study the preferred field regardless of the distance between the place of residence and the nearest study place in the preferred field. Due to monotonicity, the share of always takers can be calculated from $\pi_a = \hat{\alpha} + \hat{\gamma}\underline{z}$ ($\hat{\alpha}$ being the predicted constant). The share of never takers is calculated as the remaining group of students.

Table A.5 provides the results for different definitions of the maximum and the minimum of z. In column (1) we use as the high field availability the top 1 percentile and for the low field availability the bottom 1 percentile. Using these measure, we find that the complier share is 45.4%, whereas always takers and never takers represent 54.1% and 9.5%, respectively. Changing the definition of $\bar{z}$ and $\underline{z}$ to the top/bottom 1.5 percentile (column 2) or top/bottom 2 percentile shows that the size of the complier group decreases as the difference between $\bar{z}$ and $\underline{z}$ become smaller.

We further characterize compliers by observable characteristics. Specifically, we first separate the sample by gender and socio-economic status and recover the fraction of compliers for the different subsamples. Table 6 provides information about the characteristics of the complier group. Column (1) shows the share of the respective subgroups. Our sample consist of 60% females and 53% of our graduates are of low socio-economic status (SES), i.e. with parents who do not have an academic tertiary degree. The first stage estimate for the sub-groups are shown in column (2). We see that our instrument is stronger for males and individuals of low SES type. The proportion of the compliers ($P[X = x|I_{1i} > I_{0i}]$) is then calculated as the ratio of the first-stage coefficient for the respective subgroup relative to the overall first-stage coefficient times the population share (Angrist & Pischke, 2008; Bhuller et al., 2013). The table shows that the share of females in the population is 60%, whereas 55% of the compliers are females. This result indicates that females are underrepresented among the compliers. Similarly, we find that graduates of low SES type are over-represented among the complier group.

The lower part of the table provides the results by the broad field of study classification. Although our empirical specification always includes narrow field fixed effects that allows us to compare students with the same field preferences, first stage results by broad field classification provides valuable insights into the complier group. For sample size reasons, we group humanities and social sciences, business/economics and law, all STEM fields (science and engineering) and medicine with other fields. Again, all first stage coefficients are non-negative supporting the monotonicity assumptions. The table provides evidence that in particular students within humanities and social sciences as well as STEM fields are over-represented in the complier group, whereas business and law students are under-represented. The last column displays the relative likelihood of an individuals belonging to a particular group. Overall, our results imply that our complier group beholds a disproportionate share of males with a low socio-economic status in the broad fields of humanities, social sciences as well as STEM.

5. Results

5.1. Preferred field and study outcomes

In what follows, we turn to the analysis of the effect of studying the preferred subject on several measures of academic performance. Table 7 presents the effect on program dropout, i.e. non-completion of the first university program attended, on university dropout, changing field of study and timely graduation. Columns (1) and (3) show that, in a linear probability model (LPM), studying the preferred subject is negatively correlated with program dropout but not with university dropout. Columns (2) and (4) present the baseline IV results where we instrument whether individuals study the preferred field with the regional availability of the field. The F-statistic is relatively large, and is greater than 100 in both estimations. Therefore, weak identification issues do not apply here. The local average treatment effect (LATE) shows that attending the preferred field of study reduces the program dropout probability by about 20% points. University dropout is reduced by 11% points. These results suggest that not being able to enroll in the preferred study program implies significant costs for the students involved and the society as a whole, given the high degree of public investment in higher education in Germany. Section 5.3 puts the magnitude of the 2SLS coefficients into context, by providing estimates of the impacts for the overall sample.

After dropping out of their first university program, students may re-enroll in a different program by changing field of study. We test for the importance of this adjustment mechanism. Table 7 shows that enrolling in the preferred field of study reduces the likelihood of switching fields later on. The coefficient from the 2SLS specification is negative, larger in magnitude compared to the LPM estimate, and significantly different from zero. According to the LATE, studying the preferred field reduces the probability of changing the narrow field of study by about 20%

Table 7
Effects of studying the preferred field on dropout, field change and timely graduation.

	Program dropout		University dropout		Field change		Timely graduation	
	LPM (1)	2SLS (2)	LPM (3)	2SLS (4)	LPM (5)	2SLS (6)	LPM (7)	2SLS (8)
Preferred field	−0.081***	−0.202**	0.004	−0.113**	−0.126***	−0.204***	0.037**	0.252**
	(0.017)	(0.086)	(0.009)	(0.048)	(0.014)	(0.069)	(0.016)	(0.108)
N	3916	3916	3916	3916	3916	3916	3916	3916
R^2	0.110	0.097	0.090	0.043	0.088	0.079	0.102	0.064
F-stat. (Instrument)		113.2		113.2		113.2		113.2
Dependent v. mean	0.283	0.283	0.056	0.056	0.150	0.150	0.316	0.316

Note: Control variables included are demographics (age, gender and country of birth), educational background variables (final school grade and its interaction with the federal state and the school type, grade repetition, school track background, apprenticeship qualification), parental background (highest qualification and employment), university type, preferred field fixed effects, federal state of school graduation fixed effects, as well as characteristics of the labor market region of school graduation (number of university entrants, number and share of high school graduates with a university entrance examination, population size and density, the unemployment rate, per capita GDP and employment shares by industry). Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

points (see column 6). Changing the field of study does not need to be a negative outcome per se, if students are then able to enroll in their preferred field. However, only 6% of students not enrolled in their preferred field of study upon starting higher education switch to this field later on in their studies.

In Germany, as in most OECD countries, changing field of study entails starting another study program. Study credits earned by students are accepted for the new program only under limited circumstances, even for switches within the same university. Thus, dropping out of a study program is likely to lead to longer duration of studies and to higher direct costs and indirect costs in terms of forgone earnings. Table 7 presents evidence that studying the preferred field leads to graduating earlier. Columns (7) and (8) report results for graduating within the nominal duration of university programs (i.e. completing university within three years for bachelor programs and within 5 years for programs leading to a state examination). The point estimates are positive and statistically significant in both the LPM and 2SLS models. Studying the preferred field increases the likelihood of timely graduation by 25% points.

5.2. Heterogeneous effects

The results so far point to economically sizeable effect of studying the preferred field on academic performance measured by dropout rates and changes in field of studies. But is the impact similar for all types of students? The analysis in Section 4.4 points to the fact that in particular low SES students, i.e. students with parents without an academic degree, are over-represented among the compliers. Low SES students are more likely to be constrained by mobility costs. There is evidence that students from lower-income families are particularly disadvantaged by distance in the US (Frenette, 2006). However, the paper by Spiess and Wrohlich (2010) concludes that in Germany, the distance to university affects enrollment irrespective of income and education of parents. In this section, we therefore provide evidence on the effects of studying the preferred field by SES status. Gender may also play a role in college decisions. Astorne-Figari and Speer (2018) find that men are more likely than women to drop out of college, while women are more likely to switch majors. Program dropout also differs by field of study and gender. In a second step, we therefore also present results by gender.

Table 8 shows the 2SLS results for the sub-groups on program and university dropout. The table documents sizable differences across gender and SES status. As can be seen in columns 1–4, lower SES students have higher program and university dropout rates when they are unable to enroll in their preferred field of study than higher SES students. This is an interesting result, given that low SES students overall face a higher risk of university dropout than high SES students. Columns (5) to (8) report the results by gender. Men that could not enroll in their preferred field of study drop out of their current program with a higher probability than women, whereas the effect of studying the preferred field on university dropout does not depend on gender. Male students have a higher overall risk of university dropout in the sample, while female students are more likely to switch the study program, in line with previous results in the literature (Astorne-Figari & Speer, 2018).

5.3. Economic significance

To put the relatively large magnitudes of the 2SLS estimates into perspective, we calculate counterfactual program dropout rates, changes in field of study, timely graduation, and university drop rates that would have occurred had all graduates studied according to their preferences. To do so, we follow Angrist and Pischke (2008) and Bhuller et al. (2013) and calculate the counterfactual rates by the actual share minus the predicted effect among compliers. The predicted effect of preferred field of study on our outcome variables is calculated as the average value of our instrument times the first stage times the LATE.

Table 9 provides the results. Our estimates suggest that 25.9% of students would not have completed their first university program if all individuals would have studied their preferred field. Likewise, we estimate the counterfactual field change to be 12.7% and timely graduation to be 34.5%. These results imply that, if all individuals were able to study their preferred field, program dropout would decrease by 8%, field change by 15% and timely graduation would increase by about 9%. The counterfactual for university dropout are of particular relevance because dropping out of university likely generates high costs for the individual (Ost et al., 2018) as well as for the state in terms of budget expenses (Pfeiffer & Stichnoth, 2020). The counterfactual calculation suggests that only 4.3% would have dropped out if all individuals had studied their preferred field. This represents a reduction of 23% relative to the actual university dropout rate.

5.4. Robustness checks

In this section, we show that the main results are robust to several consistency checks. We thereby focus on the impact of studying the preferred field on program dropout.

Table A.7 firstly reports the effects of attending the preferred field of study using a probit and a bivariate probit regression including the instrumental variable in the preferred field equation. The estimates are very similar to the baseline model. This provides confidence that

Table 8
Results on program and university dropout by socio-economic status and gender.

	Program dropout		University dropout		Program dropout		University dropout	
	High SES (1)	Low SES (2)	High SES (3)	Low SES (4)	Female (5)	Male (6)	Female (7)	Male (8)
Preferred field	−0.032	−0.282***	−0.102	−0.128*	−0.092	−0.327***	−0.126*	−0.122*
	(0.161)	(0.105)	(0.064)	(0.066)	(0.133)	(0.112)	(0.074)	(0.071)
N	1837	2079	1837	2079	2362	1554	2362	1554
R^2	0.142	0.097	0.052	0.072	0.102	0.132	0.029	0.114
F-stat. (Instrument)	41.8	63.2	41.8	63.2	55.3	50.6	55.3	50.6
Dependent v. mean	0.280	0.286	0.036	0.074	0.290	0.273	0.050	0.066

Note: Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9
Economic significance.

	Program dropout (1)	Field change (2)	Timely graduation (3)	University dropout (4)
Actual rate	0.283	0.150	0.316	0.056
Counterfactual rate	0.259	0.127	0.345	0.043

Note: The table presents actual and counterfactual rates with respect to program dropout, field changes, timely graduation, and university dropout. The counterfactual rates are calculated as the actual rate minus the predicted effect. The predicted effect of preferred field of study is calculated as the average of the instrument times the first stage times the LATE.

Table 10
The potential role of motivation for the effect of studying the preferred field.

	Study identification (1)	Study satisfaction (2)	Study hours (3)
Preferred field	0.486**	1.024**	4.004*
	(0.205)	(0.444)	(2.183)
N	4181	4181	4181
F-stat (Instrument)	148.7	148.7	148.7
Dependent v. mean	3.63	6.71	13.88

Note: The dependent variables are variables measured 2–3 years after university entry. Study identification is a categorical variable (1–5 scale) denoting the extent to which students identify themselves with their degree program. Study satisfaction is measured with a 0–10 Likert scale. Study hours refer to the number of hours per week spent studying outside of classes. Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

functional form assumptions do not affect our estimates. Columns (3) and (4) show the results for a sample including only students studying their first or second preferred field. Students enrolled in their first preferred field are thus not compared now to all other students, but only to those who study their second-best alternative. If anything, the point estimates of the 2SLS in this case are larger than those of the main sample in absolute values. However, there is a loss of precision due to the smaller sample size.

Secondly, we focus on the issue of panel attrition. In our baseline sample, we select individuals observed in wave 9 or 10 of the NEPS data, i.e. 5–6 years after the first semester of university. However, there is substantial panel attrition in the data: only 56% of students are surveyed in wave 9 and 51% in wave 10. The results of the analysis may be biased if the probability of participating in the survey is related to studying the preferred field and to our outcome variables. Columns (1) and (2) of Table A.8 show that studying the preferred field is not related to the participation in the survey in the waves considered. The LPM and 2SLS estimates for studying the preferred field are small and not significantly different from zero. Moreover, columns (3)–(6) report the results on program dropout and field change for all individuals observed in either wave 3 or 5 (i.e. circa 1.5 or 2.5 years after university entry). The coefficients are statistically significant for both variables and only slightly smaller than the main estimates (in absolute value).

Third, Table A.9 shows that the results are robust to the addition of further control variables. In the baseline estimates, we condition on an extensive set of individual and regional characteristics. Nevertheless, since the preferences relative to the field of study may, to a certain extent, reflect parents' opinion, we now add information on the parents' judgement about the subject choice in the first university semester. Second, we add mathematics and German final school grades separately to capture potential differences in abilities and skills by school subject. Third, we include information on the big five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism), which are only available for a subset of students.[9] The coefficient of interest changes only little when including these variables.

Fourth, Table A.10 shows that the main results are not driven by students in teaching programs or in universities of applied sciences (*FH*), nor by other specific sub-groups. First, the point estimates are very similar when excluding students in teaching programs and the 2SLS coefficient becomes slightly larger in size when students from universities of applied sciences are excluded. Second, we exclude individuals working more than 20 h per week, because they may not have a serious intent to pursue their study program.[10] Third, we exclude students that graduated from school in East Germany to make sure that the results are not driven by this smaller group of students. Again, results are highly robust to these adjustments.

Fifth, Table A.11 shows that the results of our estimations are fairly similar when we employ different specifications of our instrumental variable. Our instrumental variable uses a bandwidth of 150 km, meaning that universities further away than 150 km receive very small weights. The results are very similar when a smaller bandwidth of 100 km or a bigger one of 200 km is employed. The results are also similar when we use a similar kernel function as Kamhöfer et al. (2019), namely $K(dist_k) = \phi(dist_k/150)$ where ϕ is the standard normal pdf.

[9] Given that personality traits are measured during the second year of university, these may be influenced both by the likelihood to study the preferred field and by academic performance during the first year of studies.

[10] In Germany, where tuition fees are very low or not existent, some individuals may have an incentive to enroll to university just to receive some benefits that students have (e.g. lower health insurance costs).

Our IV uses variation in the local number of college spots available. Exploiting the size of the local offer increases the relevance of the instrument. However, the results are qualitatively similar when we use instead a binary instrumental variable denoting whether the preferred field of study is offered at any university in the labor market region of school graduation. Finally, our IV exploits only the local field availability in the chosen program type (i.e. university of applied sciences, teaching programs and non-teaching programs at traditional universities), but not all fields are available in universities of applied sciences or teaching programs. When excluding the students who are studying a program not offering their preferred field, the size of the coefficient is almost unchanged.

Finally, Table A.11 shows that the results are not due to specific choices related to the sample considered. First, the main estimate is robust to estimating a 2SLS with sampling weights, even if the estimate is slightly less precise. Second, the results are comparable when imputing the preferred field based on a direct question on whether students are enrolled in the preferred field. While the sample is larger when doing this imputation (more than 7000 students), this does not improve the precision of our estimates given that we can impute only the preferred field for students stating that they are enrolled in that field.[11] All in all, these estimates prove that the results are robust to different choices about the instrumental variable strategy or main explanatory variable.

6. Potential mechanisms

The psychological literature has stressed that passion and motivation for studying play an important role for academic achievement (Stoeber et al., 2011). Students may have a higher intrinsic motivation when pursuing their studies in their preferred field and may find it harder to motivate themselves for a different field. This could be one mechanism through which studying the preferred field affects the probability of dropout. Descriptive evidence from our data is in line with this hypothesis.

First, we look at whether individuals report that they can identify themselves with their degree program and at their satisfaction with their studies two or three years after university entry. Columns (1) and (2) of Table 10 show that students that are able to study their preferred field are more likely to report to identify themselves with the program and are more satisfied with their studies. Second, we analyze differences in terms of the number of hours spent studying in a regular week two or three years after university entry. Students who enroll in their preferred field spend on average four hours more per week studying (column 3). This higher effort may be related to a stronger motivation, or to a higher interest and passion in the subject.

Previous studies have suggested that studying the preferred program leads to better education and labor market outcomes due to a comparative advantage in the subject studied rather than due to a higher motivation (Daly et al., 2022; Kirkeboen et al., 2016). While the results on identification and satisfaction with the study program may be consistent with students having a comparative advantage in the preferred field, this is not the case for the results on hours spent studying. If students are relatively better in the preferred field, they would need to invest less time studying in the preferred program compared to that needed for succeeding in their second-best alternative.

[11] Our main "preferred field" variable is based on a comparison on the current field to the first preferred field. Many students do not report this field ans just answer a question asking whether they are enrolled in the preferred field. This allows to impute the preferred field, but only for the students enrolled in that field. This leads to a selective increase in the share of students enrolled in the preferred field.

7. Conclusion

In this paper we investigate the impact of studying the preferred field of study on academic performance. Using detailed survey data from a cohort of students in Germany, we document that students who are enrolled in their preferred field of study in the first semester are less likely to leave the first university program before completion, and less likely to delay graduation or dropout from university. To identify the causal effect of studying the preferred field, we use an instrumental variables approach in which the probability to enroll in the preferred field of study depends on the regional supply of study spots in the preferred field, conditional on field preference and an extensive set of individual and regional characteristics. The IV results imply that, if all individuals would be able to study their preferred field, the probability that students change field of study would decrease by 15%, the probability to graduate on time would increase by 9%, and university dropout would decrease by 23%.

The impact of the regional supply of study places on the probability to enroll in the preferred field of study appears to be larger for male students and students from a lower socio-economic background. Consistent with these results, the estimated impact of attending the preferred field of study on academic performance is particularly strong among these groups of students. Our results are especially meaningful for the outcomes of students from a low socio-economic background given that they face a higher risk of dropout than students from a high socio-economic background. Finally, we present evidence suggesting that interest in the subject and motivation may be important drivers behind our results.

According to our results, ensuring a minimum number of study fields in each region as well supporting student mobility may contribute to reducing dropout rates in tertiary education. This would ensure that more students are able to study their preferred field. However, other arguments may speak in favor of regional clustering of study fields. For instance, clustering of researchers and research institutes is favorable to excellence in research and could attract international projects and researchers (see, e.g., Moretti, 2021). This again may be beneficial for students through the quality of teaching and networks for future employment. It is therefore not clear from a policy perspective what level of regional spread would be the best.

CRediT authorship contribution statement

Francesco Berlingieri: Conceptualization, Methodology, Software, Writing. **André Diegmann:** Conceptualization, Methodology, Software, Writing. **Maresa Sprietsma:** Conceptualization, Methodology, Writing.

Declaration of competing interest

Declarations of interest: none.

Data availability

The authors do not have permission to share data.

Appendix A. Additional tables and figures

See Fig. A.1. See Tables A.1–A.11.

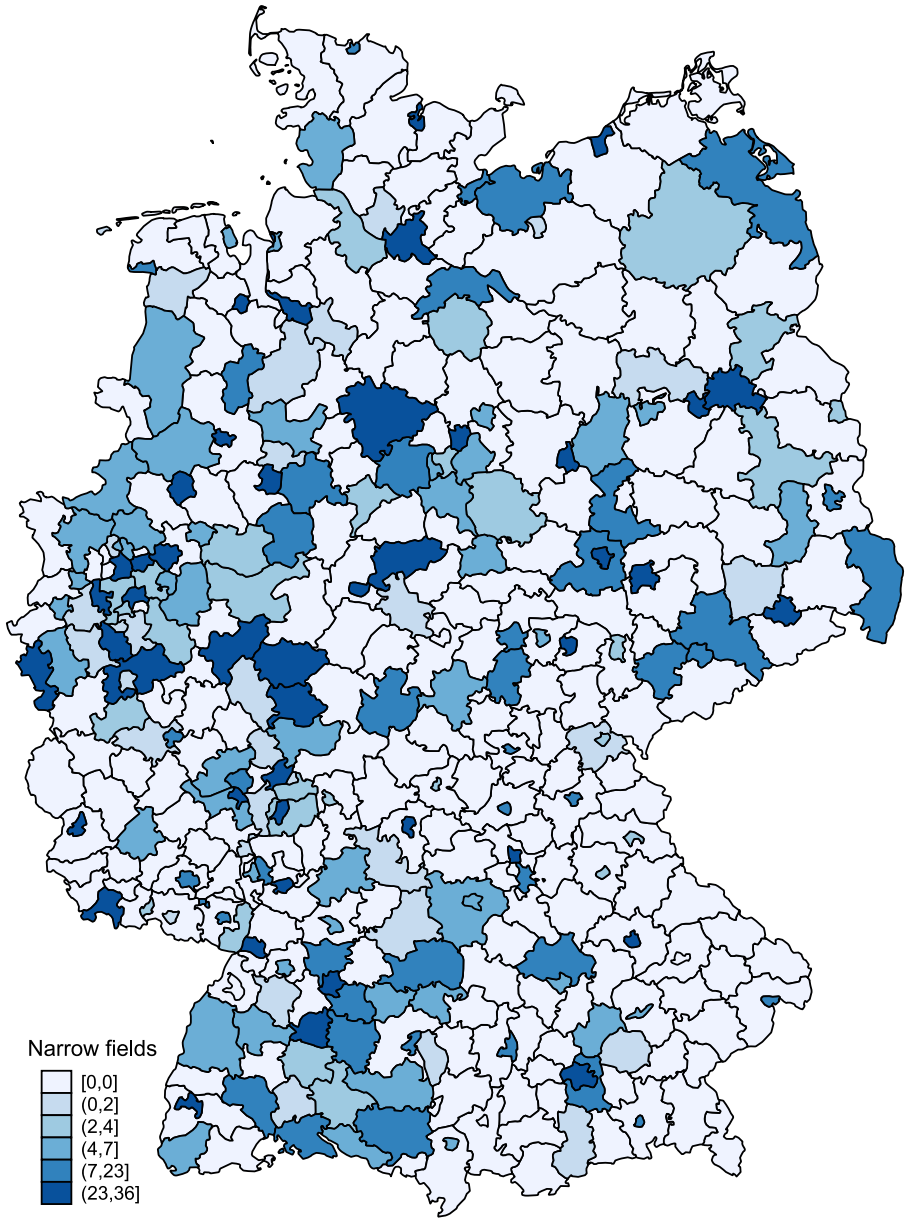

Fig. A.1. Number of field of studies offered in German districts.

Table A.1
Classification of 2-digit Destatis fields into 35 narrow fields.

2-digit Destatis fields	35 narrow fields	Dropping reason	Merging reason	N	Sample share
Linguistics and cultural studies (General)	.	<20 obs.		14	0.33
Protestant Theology	Theology		<20 obs.	14	0.33
Catholic Theology	Theology		<20 obs.	19	0.45
Philosophy	Philosophy			43	1.02
History	History			118	2.79
Information and Library Science	.	<20 obs.		1	0.02
Literature and Languages (General)	.	interdisciplinary		41	0.97
Classical Philology	Classical Philology			32	0.76
German Studies	German Studies			343	8.11
English Studies, American Studies	English Studies, American Studies			240	5.68
Romance Studies	Romance Studies			90	2.13
Slavonic, Baltic, Finno-Ugrian Studies	.	<20 obs.		4	0.09
Other Linguistics and Cultural Studies	.	<20 obs.		4	0.09
Cultural Studies	.	<20 obs.		10	0.24
Psychology	Psychology			125	2.96
Education	Education			58	1.37
Special Education	Special Education			46	1.09
Sport	Sport			71	1.68
Law, Economics and Social Sciences (General)	.	interdisciplinary		23	0.54
Regional Studies	.	no obs.		0	0
Political Science	Political Science			35	0.83
Social Sciences/Sociology	Social Sciences/Sociology			159	3.76
Social Services	Social Services			92	2.18
Law	Law			135	3.19
Public Administration	.	<20 obs.		19	0.45
Economics and Business	Economics and Business			368	8.7
Industrial Engineering (Economics Focus)	.	no preference		105	2.48
Mathematics, Natural Sciences (General)	.	interdisciplinary		7	0.17
Mathematics	Mathematics			342	8.09
Computer Science	Computer Science			162	3.83
Physics, Astronomy	Physics, Astronomy			113	2.67
Chemistry	Chemistry			112	2.65
Pharmacy	Pharmacy			28	0.66
Biology	Biology			208	4.92
Earth Sciences (excl. Geography)	Earth Sciences/Geography		<20 obs.	17	0.4
Geography	Earth Sciences/Geography			81	1.92
Health Sciences	Health Sciences			27	0.64
Human Medicine (excl. Dentistry)	Human Medicine			192	4.54
Dentistry (Clinical-Practical)	Human Medicine		<20 obs.	7	0.17
Veterinary Medicine	.	<20 obs.		14	0.33
Landscape Management/Architecture	.	<20 obs.		6	0.14
Agriculture, Food and Beverage Technology	Agriculture/Forestry Management			23	0.54
Forestry, Wood Management	Agriculture/Forestry Management		<20 obs.	15	0.35
Nutritional and Domestic Science	.	<20 obs.		18	0.43
Engineering (General)	Engineering (General)			75	1.77
Mining, Metallurgy	.	no obs.		0	0
Mechanical/Process Engineering	Mechanical/Process Engineering			235	5.56
Electrical Engineering	Electrical Engineering			89	2.1
Transport Engineering, Nautical Science	Transport Engineering			20	0.47
Architecture	Architecture/ Spatial Planning			27	0.64
Spatial Planning	Architecture/ Spatial Planning		<20 obs.	10	0.24
Civil Engineering	Civil Engineering			58	1.37
Surveying	.	<20 obs.		5	0.12
Industrial Engineering (Engineering Focus)	Industrial Engineering			26	0.61
Art, Art Theory (General)	Arts		<20 obs.	19	0.45
Fine Arts	.	no obs.		0	0
Design	Arts			20	0.47
Performing Arts	Music/Performing Arts		<20 obs.	4	0.09
Music, Musicology	Music/Performing Arts			60	1.42

Note: The table shows the regrouping of the 2-digit Destatis field classification into the classification of 35 narrow fields used in the paper. It indicates whether a field is dropped or merged to a larger field, as well as the number of observations in the final sample before the regrouping. Interdisciplinary fields are fields that are broad groupings of several other fields of the Destatis classification. ''No preference'' indicates that the given field is not part of the classification used for field preferences.

Table A.2
Field offerings and number of students by broad field classification.

	Districts with field		Students in field	
	Number (1)	Share (2)	Number (3)	Share (4)
Humanities	76	0.40	36,999	0.10
Social sciences	97	0.51	36,872	0.10
Law	65	0.16	16,701	0.04
Business and economics	163	0.86	68,179	0.18
Science	149	0.78	64,975	0.17
Engineering	154	0.81	82,112	0.22
Medicine	34	0.18	10,624	0.03
Other fields	117	0.62	27,612	0.07

Note: Column (1) of the table presents the number of districts with university or applied universities that offer programs in the respective broad field classification. Out of 402 districts, 190 districts have at least one university or applied university. Column (2) calculates the share conditional on have at least one (applied) university. Column (3) provides the total number of students in the respective broad field classification. Column (4) presents the share of students within the broad field classifications.

Table A.3
Determinants of studying the preferred field (first stage)

	LPM (1)	LPM (2)	LPM (3)	LPM (4)	LPM (5)	LPM (6)	Probit (7)
Field availability	0.087***	0.088***	0.087***	0.095***	0.095***	0.099***	0.112***
	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)
University of applied science	−0.154***	−0.226***	−0.227***	−0.205***	−0.200***	−0.203***	−0.174***
	(0.022)	(0.025)	(0.024)	(0.027)	(0.027)	(0.027)	(0.027)
Teaching program	−0.103***	−0.164***	−0.168***	−0.137***	−0.135***	−0.135***	−0.109***
	(0.019)	(0.029)	(0.029)	(0.030)	(0.030)	(0.030)	(0.027)
Bachelor degree	−0.044**	−0.084***	−0.083***	−0.092***	−0.091***	−0.096***	−0.081***
	(0.018)	(0.028)	(0.028)	(0.027)	(0.027)	(0.027)	(0.024)
Age (university entrance)			0.004	0.009**	0.009***	0.009***	0.009**
			(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Female			0.046***	0.036**	0.038**	0.036**	0.034**
			(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Born in Germany			0.031	−0.015	−0.001	0.003	0.004
			(0.033)	(0.034)	(0.042)	(0.042)	(0.041)
School final grade				0.103**	0.100**	0.107**	0.085*
				(0.045)	(0.046)	(0.046)	(0.047)
Narrow field preference	No	Yes	Yes	Yes	Yes	Yes	Yes
School background	No	No	No	Yes	Yes	Yes	Yes
Parental background	No	No	No	No	Yes	Yes	Yes
Regional characteristics	No	No	No	No	No	Yes	Yes
Observations	3916	3916	3916	3916	3916	3916	3916
R^2	0.081	0.176	0.179	0.216	0.218	0.225	

Note: Control variables included are demographics (age, gender and country of birth), educational background variables (final school grade and its interaction with the federal state and the school type, grade repetition, school track background, apprenticeship qualification), parental background (highest qualification and employment), university type, preferred field fixed effects, federal state of school graduation fixed effects, as well as characteristics of the labor market region of school graduation (number of university entrants, number and share of high school graduates with a university entrance examination, population size and density, the unemployment rate, per capita GDP and employment shares by industry). Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4
First-Stage estimates by different groups.

	SES		Gender		Preferred field			
	High (1)	Low (2)	Female (3)	Male (4)	Humanities & Social S. (5)	Business & Law (6)	STEM (7)	Medicine & Other (8)
Field availability	0.092***	0.106***	0.090***	0.108***	0.283***	0.021	0.205***	0.095**
	(0.014)	(0.013)	(0.012)	(0.015)	(0.038)	(0.014)	(0.027)	(0.041)
Observations	1837	2079	2362	1554	1353	456	1416	691
R^2	0.255	0.236	0.254	0.254	0.270	0.281	0.195	0.439

Note: Control variables included are as in Table 4. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5
Compliance types in the sample.

	1%	1.5%	2%
	(1)	(2)	(3)
Students of compliers	0.454	0.342	0.290
Students of never takers	0.095	0.173	0.205
Students of always takers	0.451	0.485	0.504

Notes: Calculation based on Dahl et al. (2014). Different columns indicate different percentiles for z.

Table A.6
Program dropout estimates controlling for different sets of variables.

	(1)	(2)	(3)	(4)
Preferred field	−0.183***	−0.164*	−0.168*	−0.202**
	(0.058)	(0.091)	(0.091)	(0.086)
Program type	Yes	Yes	Yes	Yes
Narrow field preference	No	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes
School background	No	Yes	Yes	Yes
Parental background	No	No	Yes	Yes
Regional characteristics	No	No	No	Yes
N	3916	3916	3916	3916
R^2	0.036	0.096	0.098	0.097
F-stat. (Instrument)	279.5	106.2	107.9	113.2

Note: Control variables included are as in Table 4. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7
Results on program dropout: alternative specifications.

	Program Dropout		First or second preference	
	Probit	Biv.Probit	LPM	2SLS
	(1)	(2)	(3)	(4)
Preferred field	−0.080***	−0.177***	0.057***	−0.323*
	(0.015)	(0.064)	(0.021)	(0.177)
N	3916	3916	3350	3350
R^2			0.110	0.061
F-stat. (Instrument)				32.9
Dependent v. mean	0.283	0.283	0.056	0.056

Note: Control variables included are demographics (age, gender and country of birth), educational background variables (final school grade and its interaction with the federal state and the school type, grade repetition, school track background, apprenticeship qualification), parental background (highest qualification and employment), university type, preferred field fixed effects, federal state of school graduation fixed effects, as well as characteristics of the labor market region of school graduation (number of university entrants, number and share of high school graduates with a university entrance examination, population size and density, the unemployment rate, per capita GDP and employment shares by industry). Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8
Taking into account panel attrition.

	Survey Participation		Outcomes from waves 3 or 5			
	Observed in wave 9 or 10		Program dropout		Field change	
	LPM	2SLS	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Preferred field	0.009	−0.086	−0.097***	−0.130*	−0.129***	−0.141**
	(0.014)	(0.082)	(0.014)	(0.077)	(0.012)	(0.059)
N	5748	5748	5190	5190	5190	5190
R^2	0.049	0.042	0.107	0.106	0.084	0.084
F-stat. (Instrument)		170.9		141.9		141.9

Note: Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9
Results on program dropout: including further control variables.

	Parents' opinion		Specific grades		Big Five	
	LPM	2SLS	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Preferred field	−0.066***	−0.167*	−0.080***	−0.214**	−0.080***	−0.204**
	(0.016)	(0.087)	(0.017)	(0.087)	(0.017)	(0.086)
Main controls	Yes	Yes	Yes	Yes	Yes	Yes
Opinion parents/friends	Yes	Yes	No	No	No	No
Federal state FE	No	No	Yes	Yes	No	No
Big Five	No	No	No	No	Yes	Yes
N	3891	3891	3859	3859	3826	3826
R^2	0.126	0.117	0.110	0.099	0.117	0.103
F-stat. (Instrument)		106.2		110.6		109.8

Note: Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10
Results on program dropout: Excluding sub-samples.

	Excl. teaching		Excl. FH		No FT work		Only West	
	LPM	2SLS	LPM	2SLS	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Preferred field	−0.092***	−0.244*	−0.084***	−0.544**	−0.076***	−0.186**	−0.083***	−0.190**
	(0.022)	(0.138)	(0.019)	(0.260)	(0.017)	(0.091)	(0.019)	(0.089)
N	2412	2412	3280	3280	3834	3834	3184	3184
F-stat. (Instrument)		51.1		18.2		104.0		94.3

Note: Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11
Results on program dropout: different instrumental variables.

		Bandwidth		Different	Binary	Excl.	Weighted	Imputed
	Baseline	100 km	200 km	kernel	IV	fields	IV	Pref.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Preferred field	−0.202**	−0.226**	−0.189**	−0.187**	−0.256*	−0.199**	−0.183*	−0.270***
	(0.086)	(0.103)	(0.081)	(0.080)	(0.144)	(0.098)	(0.110)	(0.098)
N	3916	3916	3916	3916	3916	3859	3916	7188
R^2	0.097	0.091	0.099	0.100	0.083	0.099	0.114	0.081
F-stat. (Instrument)	113.2	81.6	126.6	128.2	37.7	90.1	79.7	107.5

Note: Control variables included are as in Table 7. Standard errors are clustered at the level of high school districts, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, *113*(2), 231–263.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arcidiacono, P., Hotz, J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, *166*, 3–16.

Astorne-Figari, C., & Speer, J. (2018). Drop out, switch majors, or persist? The contrasting gender gaps. *Economics Letters*, *164*, 82–85.

Bhuller, M., Dahl, G. B., Lø ken, K. V., & Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, *128*(4), 1269–1324.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband internet: An information superhighway to sex crime? *Review of Economic Studies*, *80*(4), 1237–1266.

Blossfeld, H., Rosbach, H., & von Maurice, J. (2011). Education as a lifelong process: The german national educational panel study (NEPS). *Zeitschrift Für Erziehungswissenschaft*, *14*.

Boudarbat, B., & Montmarquette, C. (2009). Choice of fields of study of university Canadian graduates: The role of gender and their parents education. *Educ. Econ.*, *17*(2), 185–213.

Dahl, G. B., Kostøl, A. R., & Mogstad, M. (2014). Family welfare cultures. *Quarterly Journal of Economics*, *129*(4), 1711–1752.

Daly, M., Jensen, M. F., & le Maire, D. (2022). University admission and the similarity of fields of study: Effects on earnings and skill usage. *Lab. Econ.*, *75*, Article 102118.

Danilowicz-Oesele, K., Lerche, K., Meya, J., & Schwager, R. (2017). Determinants of students' success at university. *Educ. Econ.*, *25*(5), 513–532.

De Paola, M., & Gioia, F. (2012). Risk aversion and field of study choice: The role of individual ability. *Bull. Econ. Res.*, *64*, 0307–3378.

De Paola, M., & Scoppa, V. (2015). Procrastination, academic success and the effectiveness of a remedial program. *J. Econ. Behav. Organ.*, *115*, 217–236.

Denzler, S., & Wolter, S. (2010). Der Einfluss des lokalen Hochschulangebots auf die Studienwahl. *Zeitung Für Erziehungswissenschaft*, *13*, 683–706.

Destatis (2021). *Nichtmonetäre hochschulstatistische Kennzahlen*: *Fachserie 11 Reihe 4.3.1*, Statistisches Bundesamt (Destatis).

Frenette, M. (2006). Too far to go on? Distance to school and university participation. *Educ. Econ.*, *14*(1), 31–58.

Heinesen, E. (2018). Admission to higher education programmes and student educational outcomes and earnings – Evidence from Denmark. *Economics of Education Review*, *63*, 1–19.

Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J., & Woisch, A. (2017). Zwischen studienerwartungen und studienwirklichkeit, ursachen des studienabbruchs, beruflicher verbleib der studienabbrecherinnen und studienabbrecher und entwicklung der studienabbruchquote an deutschen hochschulen. *Forum Hochschule*, *1*.

Hilmer, M. J., & Hilmer, C. E. (2012). On the relationship between student tastes and motivations, higher education decisions, and annual earnings. *Economics of Education Review*, *31*, 66–75.

Imbens, G. W., & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, *64*(4), 555–574.

Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association*, *17*(1), 205–244.

Kirkeboen, L. (2012). Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education. *725*, *Discussion Paper Statistics Norway*.

Kirkeboen, L., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *Quarterly Journal of Economics*, *131*(3), 1057–1111.

Kosfeld, R., & Werner, A. (2012). Deutsche Arbeitsmarktregionen–Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung Und Raumordnung*, *70*(1), 49–64.

Larsen, M. S. g., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R., & Sommersel, H. B. r. y. (2013). Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it? A systematic review. *Technical Report*, Danish Clearinghouse for Educational Research.

Lassibille, G., & Gomez, L. N. (2008). Why do higher education students drop out? Evidence from Spain. *Educ. Econ.*, *16*(1), 89–105.

Modena, F., Rettore, E., & Tanzi, G. (2020). The effect of grants on university dropout rates : Evidence from the Italian case. *J. Hum. Capital*, *14*(3), 343–370.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, *111*(10), 3328–3375.

OECD (2020). *Education at a glance 2020: OECD indicators*. Paris: OECD Publishing.

Ost, B., Pan, W., & Webber, D. (2018). The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies. *Journal of Labor Economics*, *36*(3), 779–805.

Pfeiffer, F., & Stichnoth, H. (2020). Fiscal and individual rates of return to university education with and without graduation. *Applied Economics Letters*, *28*(16), 1432–1435. http://dx.doi.org/10.1080/13504851.2020.1855299.

Spiess, C., & Wrohlich, K. (2010). Does distance determine who attends a university in Germany? *Economics of Education Review*, *29*(3), 470–479.

Stinebrickner, T., & Stinebrickner, R. (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *American Economic Review*, *98*(5), 2163–2184.

Stinebrickner, T., & Stinebrickner, R. (2014). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *J. Labour Econ.*, *32*(3), 602–644.

Stoeber, J., Childs, J. H., Hayward, J. A., & Feast, A. R. (2011). Passion and motivation for studying: Predicting academic engagement and burnout in university students. *Br. J. Educ. Psychol.*, *31*(4), 513–528.

Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E., & Wollscheid, S. (2015). *Dropout and completion in higher education in Europe: Main report*. European Union.

Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identication using an information experiment. *Review of Economic Studies*, *82*(2), 791–824.