// NO.22-017 | 06/2022

DISCUSSION PAPER

// FRANCESCO BERLINGIERI, ANDRÉ DIEGMANN,
AND MARESA SPRIETSMA

Preferred Field of Study and Academic Performance

Leibniz Association

ZEW

Preferred Field of Study and Academic Performance*

Francesco Berlingieri†, André Diegmann‡and Maresa Sprietsma§

May 25, 2022

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 driven by lower academic performance and motivation.

JEL-classification: I21, I23, J24

Keywords: field of study, preferences, academic performance, university dropout

*We thank 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. A significant part of this study has been undertaken by the authors at the ZEW – Leibniz Centre for European Economic Research ZEW Mannheim between 2017 and 2021. 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. The usual disclaimer applies.

†European Commission, Joint Research Centre (JRC), E-mail: francesco.berlingieri1@ec.europa.eu

‡IWH Halle, ZEW Mannheim E-Mail: Andre.Diegmann@iwh-halle.de

§Hochschule der Bundesagentur fuer Arbeit, E-Mail: Maresa.Sprietsma@arbeitsagentur.de

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 and 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 and 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 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 to 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 relative availability of fields 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). Consistent with these estimates, students who study their preferred field are more likely to graduate in the same field. 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.

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. This is further supported by heterogeneous effects with respect to the broader subject area. We show that students who do not study their preferred field but study a field that lies within the broad subject group of preference (e.g. physics vs. chemistry) are also more likely to drop out of university. This suggests that having a comparative advantage in the preferred subject is not the only channel at place, since the skills required to perform well in related fields are likely to be similar. However, we find also some support for the importance of comparative advantage. In fact, those studying their preferred field are less likely to drop out from university because of low performance.

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 effect 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. However, this result only holds when the second choice field belongs to another broad field category than the first-choice field. In line with these results, we find that, in Germany, studying the preferred field leads to lower university dropout. This is also the case when the alternative study field lies in the same broad subject area as the preferred field.

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 and Montmarquette, 2009), risk aversion (De Paola and

Gioia, 2012), and individual taste (Wiswall and Zafar, 2015; Hilmer and Hilmer, 2012) 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 and Gomez, 2008). Students face uncertainty as to their skills and abilities when entering university. According to Stinebrickner and Stinebrickner (2014), learning about own ability in the first years of college accounts for 45% of dropout cases in the first year of college. Credit constraints (Stinebrickner and Stinebrickner, 2008; Modena et al., 2020; Lassibille and Gomez, 2008) and non-cognitive skills (De Paola and 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.

In the first wave, students are asked to report their preferred field of study. 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 first 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 denoting whether students are enrolled in their first preferred field when entering university. Note that preferences are surveyed shortly after university entry and might be influenced by ex-post rationalizations or emotional cues at the time of the interview. Ideally one would observe preferences before university entry. To test for possible adjustments, 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.

Our sample is conditional on observing precise information about the preferred field, which is only available for about 45% of students in the survey. Among the remaining 55% of students, about one-third are not asked about their preferred field of study or do not answer any questions related to field preferences, thus corresponding to true missing values. The other two-thirds of students do provide an answer to the question of whether they are studying their preferred field but do not provide information on the precise field. We do not include these students in our baseline sample because we cannot categorize students according to their first preferences (and potentially second preferences) in the case that they are not studying their preferred field. This is only possible among the students who study their preferred field by imputing the actual field of study as the first preference, suggesting a bias of the sample toward this group.[2] Nevertheless, Appendix

[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.

[2] Within our sample, we also find discrepancies between stated preferences with the question of whether or not students study their preferred field and actual field enrollment. Our final sample therefore consists of students, for whom we observe of a consistent measure of study preferences and mismatch.

Table A.8 shows that the main results are not substantially different when imputing the preferred field.

We further restrict the sample to students who are interviewed during their first semester of university. Moreover, we exclude all individuals with missing information on the main control variables (educational and parental background information). We construct our outcome variables among students who are interviewed four years after university entry.[3] 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.[4] Most university entrants in the sample

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

[4] 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

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).[5] 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.[6] 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.

2.2 Regional offer of field of studies

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 that, on average, only 12 narrow fields out of 35 were offered in the 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 offered at regions in the 75^{th} percentile, only 6 fields are offered in regions of the 25^{th} 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 Figure 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

programs was approximately 8% in 2010.

[5] 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.

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

Table 1: Regional offer of field of studies

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

Note: Summary statistics for the 190 districts (out of 402) and 112 labour 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.

exploits these regional differences that are arguably exogenous conditional on preferences and relevant regional characteristics (see Section 3).

2.3 Actual and preferred fields of study

On average, 67% of students (unweighted) are able to enroll in their first preferred subject in our sample.[7] Figure 1 shows the share of students in their preferred field by broad preferred field classification. Students willing to study fields within law, business, science, engineering and humanities groups are more likely to enroll in these programs compared to the average student. On the contrary, students who want to study medicine, social

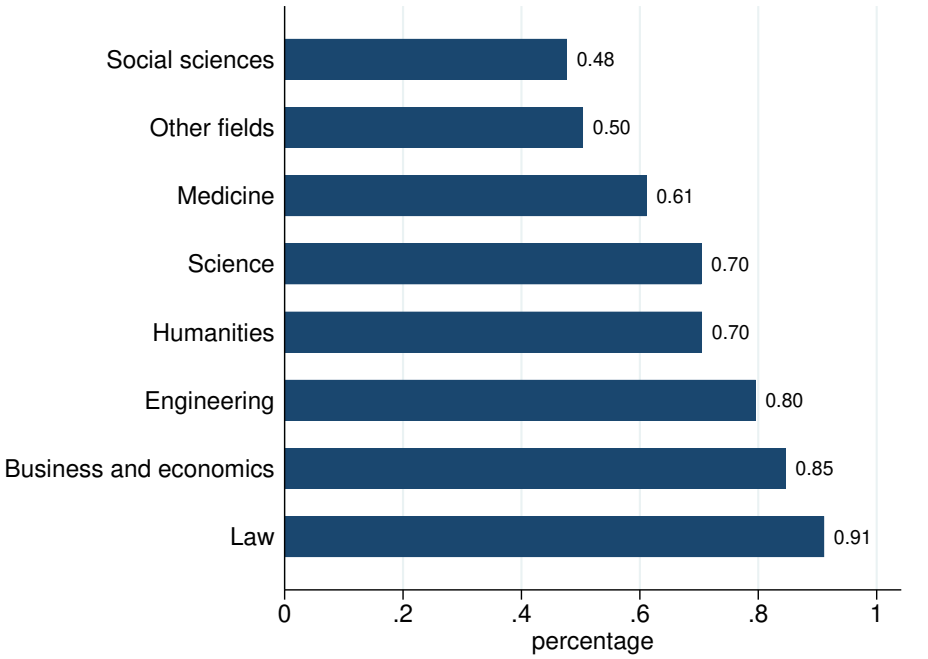


Figure 1: Students in preferred field by broad field preference

sciences and other fields very often end up studying a different field. For medicine, the

[7]When using sampling weights, the share of students enrolled in the first preferred subject is 70%, as reported in footnote 3.

lower than average share is driven by the very strict admission rules. For fields within social sciences and other fields, the lack of adequate supply at universities nearby is likely to be a more important driver of these figures.

Table 2 presents the mean of the main variables in our data for students that enroll in their preferred field and those who do not. It also shows a t-test of difference in means. The table shows that the two groups are different according to important 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 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 fulfilment. 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*).

Figure 2 shows the unconditional means between studying the preferred field and the four key outcome variables. 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

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)
Fachhochschulreife	0.07	0.08	-0.01	(-1.10)
High school: West	0.83	0.81	0.02	(1.65)
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 controls (labour market region of high school graduation)				
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.

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,

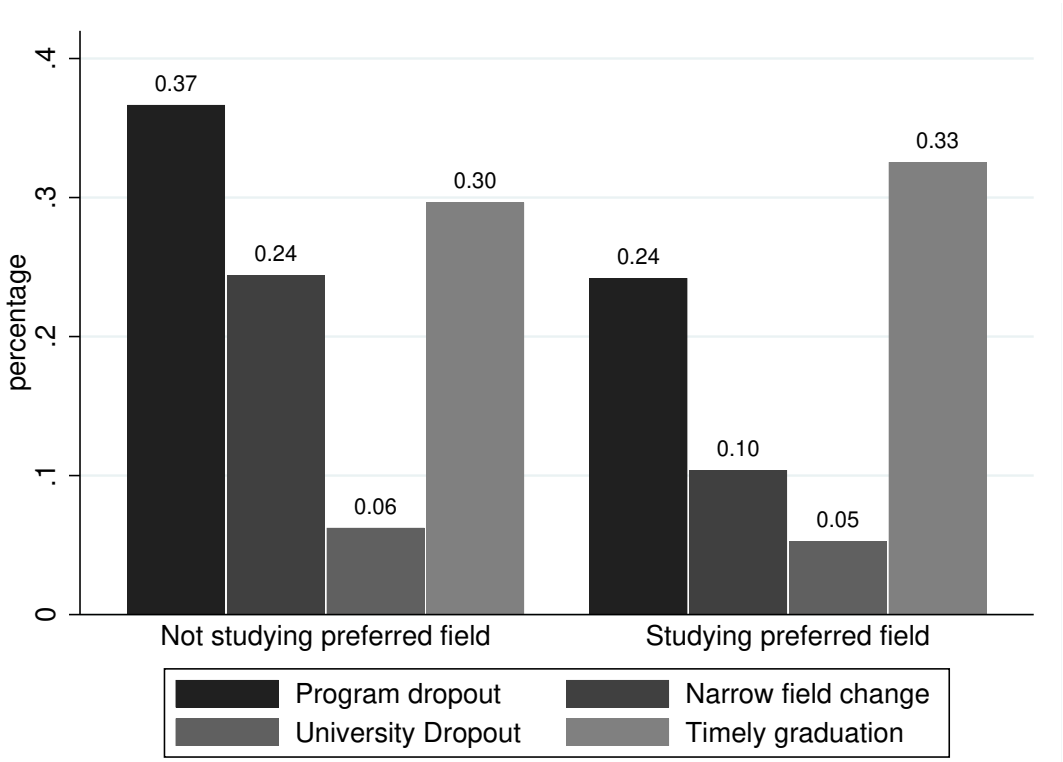


Figure 2: Preferred field and study success

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.[8] 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:

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

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

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

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 bandwith of 150km, 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 100km distance receive a weight of 0.64 and those at 150km distance a weight of 0.37. The weight of universities further away is very small. For instance, for universities at 300km 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 equation. Our results are robust to the choice of the bandwidth (see Table A.8, columns (2)-(3)). We compute the instrument based on the relative supply of study places by field 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.8, column(6)).

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.

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 pre-

ferred 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:

$$y_i = \beta preferred_i + x_i'\gamma + p_i + \epsilon_i \tag{2}$$

where y_i is a binary indicator denoting the outcome variable, i.e. program dropout, field change, timely graduation, university dropout, etc. The variable $preferred_i$ is a binary variable indicating whether individual i studies his/her preferred field 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), university type (regular university and university of applied sciences), 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).

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

$$preferred_i = z_i\gamma + x_i'\pi + p_i + \mu_i \tag{3}$$

where z_i is the instrumental variable that varies at the level of the preferred field 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. Conditional on the vector of control variables, we further introduce preferences-fixed effect, p_i, by controlling for the preferred field of study. With this empirical strategy, we compare 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 relative availability of the preferred field of study. Specifically, we measure field preferences following the 35 narrowly defined fields. This empirical approach allows us to compare students with exactly the same field of study preference graduating from high-school in different regions. For instance, we compare 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.

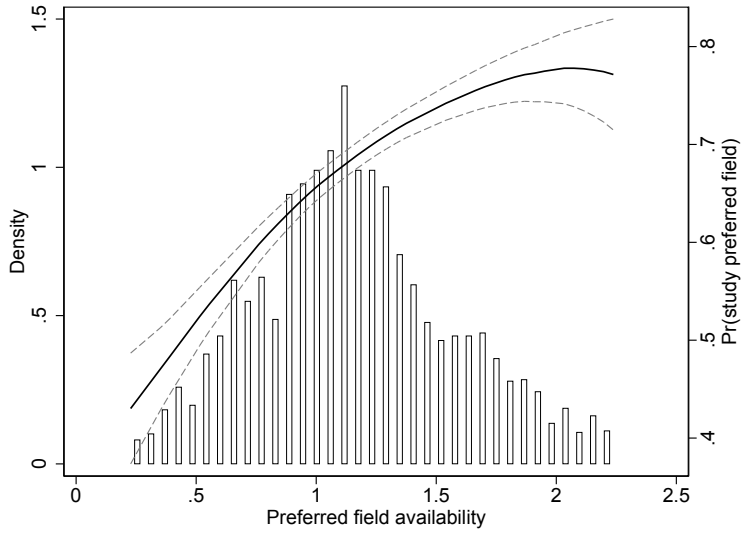
It is important to underline that the estimates based on the instrumental variable approach are local average treatment effects (LATE) and only hold for those individuals who study a given subject because of the regional availability of the subject. Individuals who have a very strong interest in a given 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, these estimates are interesting from a policy perspective, because they refer to the effects of changing the regional distribution of university subjects offered. This is something that could be modified or influenced by policy makers, for instance at the federal state level.

4 Assessment of the instrument

4.1 Instrument relevance and conditional independence

Figure 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 de-

viation 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 equation (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.



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).

Figure 3: First stage

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 Figure 3, the instrumental variable has a significant and positive effect on the probability to studying the preferred field.[9] The coefficient of 0.087 implies that an increase of 1000 slots in the

[9]As shown in the regression tables of the results, the F-statistics on the excluded instruments is 111

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.

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***
	(14.92)	(10.16)	(10.15)	(10.41)	(10.49)	(10.76)	(9.36)
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 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 labour 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, $*p < 0.10, **p < 0.05, ***p < 0.01$.

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 (6) reports average marginal effects from a probit model. The coefficients are very similar to those of the linear probability model.

We test for randomness of our instrument with respect to pre-determined variables by providing balancing results in Table 5. The first column provides the correlation between

in our preferred specification.

the preferred field of study and individual characteristics. The probability of studying the preferred field is particularly correlated with age, gender, and high school performance which indicates that the preferred field of study indicator is potentially endogenous even after conditioning on observables. Column (3) provides the results by regressing the

Table 5: Balancing Test

	Indep. variable: Preferred field		Indep. variable: Instrument	
	Coefficient	st. error	Coefficient	st. error
	(1)	(2)	(3)	(4)
Age (university entrance)	0.009**	0.004	-0.001	0.007
Female	0.040***	0.015	0.041	0.030
Born in Germany	-0.010	0.044	-0.126	0.084
School grade	0.109**	0.046	0.020	0.101
Applied university	-0.180	0.115	0.284	0.284
Repeated grade	-0.014	0.022	0.027	0.041
Middle secondary track	-0.013	0.017	0.009	0.028
High school: abroad	-0.048	0.036	0.034	0.074
Apprenticeship before 2011	0.016	0.028	-0.042	0.062
Mother: foreign born	-0.020	0.031	-0.059	0.054
Mother: employed at age 15	-0.012	0.017	0.038	0.024
Mother: tertiary degree	0.013	0.017	-0.017	0.028
Mother: no upper secondary qual.	-0.057	0.035	0.013	0.050
Father: foreign born	0.056*	0.032	0.073	0.048
Father: employed at age 15	0.010	0.033	-0.059	0.059
Father: tertiary degree	0.018	0.015	0.024	0.025
Father: no upper secondary qual.	-0.014	0.040	-0.091	0.077
		p-value		p-value
Joint F-test	2.40	0.002	1.01	0.442
Observations	3916		3916	

Note: The table shows conditional correlations of individual-level characteristics with the preferred field of study indicator and the instrument. All regressions narrow field preferences, program type, and regional characteristics. Standard errors are clustered at the level of high school districts, $*p < 0.10, **p < 0.05, ***p < 0.01$.

instrument on the observed characteristics. As born out by the table, none of the coefficients are significant which is also supported by the joint F-test at the bottom of the table. Estimates in Table 5 provide, therefore, strong evidence that pre-determined variables are randomly distributed over the instrument.

Appendix Table A.2 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.3 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.

4.3 Monotonicity

In our setting, monotonicity requires that a student who enrolls in their 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 and socio-economic background. Consistent with the monotonicity assumption, each reported coefficient is large, positive, and statistically different from zero.

4.4 Compliance

Ir 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.9 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 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 sub-groups. Our sample

Table 6: Compliance

	$P[X=x]$ (1)	First stage (2)	$P[X=x\|I_{1i}>I_{0i}]$ (3)	$\frac{P[X=x\|I_{1i}>I_{0i}]}{P[X=x]}$ (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
Low SES male	0.209	0.120	0.253	1.211
Low SES female	0.322	0.088	0.286	0.888
High SES male	0.188	0.089	0.169	0.900
High SES female	0.281	0.079	0.224	0.797
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.

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 and 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 interaction of the SES type and gender is provided in the following four rows. It shows that low SES males are quantitatively over-represented among compliers, whereas, in particular, high SES females are underrepresented. 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.

5 Results

5.1 Preferred field and program dropout

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. In all specifications, we control for the main observable variables except for the final school grade. The final school grade reflects individual ability and is a strong indicator of studying the preferred field. We present the OLS and IV results with and without the final school grade to learn about the effects of studying the preferred field on several outcome variables and how these relate to school grades. Table 7 shows the results when controlling for narrow categories of the preferred field of study. Column (1) shows that, in a linear probability model (LPM), studying the preferred subject is negatively correlated with program dropout. The coefficient of interest decreases by 3% points when controlling for the final grade in high school, since individuals with better grades are both more likely to study the preferred field and to graduate in the field of first university enrollment (see column (2)). Columns (3) and (4) present the baseline IV results where we instrument whether individuals study the preferred field with the relative availability of the field. The F-statistic is relatively large, and is greater than 100 in both panels. Therefore, weak identification issues do not apply here. The local average treatment effect (LATE) shows that attending ones first choice field reduces the program dropout probability by about 20% points. In this empirical specification, controlling for the final school grade only affects the coefficient of interest marginally.

Column (5) reports average marginal effects from a probit regression. The estimates are very similar to the LPM model. The estimates from a bivariate probit model including the instrumental variables in the preferred field equation are also fairly similar to the IV estimates (see column 6). This provides confidence that functional form assumptions do not affect our estimates. Therefore, in what follows we only report the results from linear models for simplicity. Columns (7) and (8) show the results for the sample of 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 are larger than those

Table 7: Effects of studying the preferred field on program dropout

	Main sample						1^{st} or 2^{nd} preference	
	LPM (1)	LPM (2)	2SLS (3)	2SLS (4)	Probit (5)	BiProbit (6)	LPM (7)	2SLS (8)
Preferred field	-0.111***	-0.081***	-0.220**	-0.202**	-0.080***	-0.177***	-0.057***	-0.323*
	(0.016)	(0.017)	(0.087)	(0.086)	(0.015)	(0.064)	(0.021)	(0.177)
School final grade		-0.068		-0.056	-0.065	-0.054	-0.065	-0.046
		(0.045)		(0.044)	(0.043)	(0.042)	(0.055)	(0.053)
N	3916	3916	3916	3916	3916	3916	3350	3350
R^2	0.076	0.110	0.065	0.097			0.110	0.061
F-stat. (Instrument)			113.6	113.2				32.9
Dependent v. mean	0.283	0.283	0.283	0.283	0.283	0.283	0.283	0.283

Note: Control variables included are demographics (age, gender and country of birth), educational background variables (final school grade and its interaction with the 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 labour 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, $*p < 0.10, **p < 0.05, ***p < 0.01$.

of the main sample in absolute values. However, there is a loss of precision due to the smaller sample size.

5.2 Other study outcomes

After dropping out of their first university program, students have different options. They may re-enroll in a different program by changing field of study or they may leave higher education without earning a degree. We test for the importance of these different adjustment mechanisms. Table 8 shows that enrolling in the preferred field of study reduces the likelihood of switching fields later on. Results from the LPM model are reported in column (1), while column (2) presents results from the 2SLS specification. All coefficients from the 2SLS specification are negative, larger in magnitude compared to the OLS estimates, 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% points (see column 2). 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 fact, enrolling in the preferred field at university entry is strongly related to the probability of graduating in that field (see columns (3) and (4) of Table 8).

Table 8: Effects on field change, graduation and university dropout

	Field change		Graduation pref. field		Timely graduation		University dropout	
	LPM	2SLS	LPM	2SLS	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Preferred field	-0.126***	-0.204***	0.550***	0.642***	0.037**	0.252**	0.004	-0.113**
	(0.014)	(0.069)	(0.013)	(0.085)	(0.016)	(0.108)	(0.009)	(0.048)
N	3916	3916	3916	3916	3916	3916	3916	3916
R^2	0.088	0.079	0.430	0.424	0.102	0.064	0.090	0.043
F-stat. (Instrument)		113.2		113.2		113.2		113.2
Dependent v. mean	0.150	0.150	0.425	0.425	0.316	0.316	0.056	0.056

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

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 8 presents evidence that studying the preferred field leads to graduating earlier. Columns (5) and (6) 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.

Columns (7) and (8) of Table 8 show that studying the preferred field has an important impact on university dropout, i.e. leaving higher education without acquiring a degree. While the LPM estimates are close to zero, the estimates from the 2SLS are large and statistically significant at the 5% level. These estimates imply that studying the preferred field reduces the probability of university dropout by about 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. The next section puts the magnitude of the 2SLS coefficients into context, by providing estimates of the impacts for the overall sample. 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 overrepresented among the compliers. We now provide evidence on heterogeneous effects with respect to gender and SES type.

Table 9 shows the 2SLS results for the sub-groups on program and university dropout. The table documents sizable differences across sub-groups. As can be seen in columns

Table 9: Results on program and university dropout by socio-economic status and gender

	Program dropout		University dropout		Program dropout		University dropout	
	High SES	Low SES	High SES	Low SES	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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, $*p < 0.10, **p < 0.05, ***p < 0.01$.

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 face a higher risk of university dropout than high SES students overall. 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, whereas we document similar results by gender for university dropout. 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 and 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 10: 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 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 and 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 focus thereby on the impact on program dropout.

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.5 show that studying the preferred field is not related to the participation in the survey in the waves considered. Students with worse grades are less likely to be surveyed in these two waves, showing that there is selective survey participation. Nevertheless, 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).

In the baseline estimates, we condition on an extensive set of individual and regional characteristics. Table A.6 shows that the results are robust to the addition of further control variables. First, since the preferences relative to the field of study may, to a certain extent, reflect parents' opinion, we add information on the parents' judgement about the subject choice in the first university semester. Second, we add mathematics and German final school grades 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. The coefficient of interest changes only little when including these variables.

Table A.7 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 hours 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.

Moreover, Table A.8 shows that the results on program dropout are fairly similar when we employ different instrumental variables. Our instrumental variables uses a bandwidth of 150km, meaning that universities further away than 150km receive very small weights.

[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).

The results are very similar when a smaller bandwith of 100km or a bigger one of 200km 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. 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.

Table A.8 also 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 literature has put forward two possible channels explaining why individuals studying their preferred field have better academic outcomes. First, the psychological literature has stressed that passion and motivation for studying plays an important role for academic achievement (Stoeber et al., 2011). Students may have an intrinsic motivation when pursuing their studies in their preferred field and may find it hard to complete a degree in a different program. Second, students may tend to choose fields in which they have

[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.

a comparative advantage (Kirkeboen et al., 2016). They could thus find it easier to complete the studies and earn better grades in the preferred program. In this section, we investigate the role of these two possible mechanisms.

We start by analysing possible heterogeneous effects depending on whether individuals not enrolled in their preferred programs study a field that lies within or outside of the broad subject group, similarly to Daly et al. (2022). The idea is that if motivation is an important channel, an individual willing to study a specific field, e.g. physics, would also have a worse academic performance when studying a different discipline within the broad subject group (i.e. science), such as philosophy or biology. On the contrary, if we find significant results only when individuals study a different broad subject, this would suggest that comparative advantage is likely to play a more important role.

Column (1) of Table 11 reports the results of a 2SLS regression with university dropout as outcome variable for the subsample of individuals that study a field that lies within the broad subject group of their first preferred field. The coefficient of interest is significant and not smaller than the baseline estimate. Column (2) reports results from a similar regression with individuals studying outside their broad subject preference as a control group. The estimate is also sizable and significant. These results provide some support that passion and motivation for the field of study are important mechanisms behind the results, even if one cannot exclude that comparative advantage plays a role as well.

Table 11: Results on proxies for motivation and performance

	University dropout		Channel proxies			
	Within broad gr. (1)	Outside broad gr. (2)	Study identif. (3)	Study satisf. (4)	Study hours (5)	Dropout reason: low performance (6)
Preferred field	-0.238*	-0.131**	0.486**	1.024**	4.004*	-0.045*
	(0.134)	(0.055)	(0.205)	(0.444)	(2.183)	(0.023)
N	3045	3487	4181	4181	4181	5748
F-stat (Instrument)	21.6	95.7	148.7	148.7	148.7	170.9
Dependent v. mean	0.055	0.055	3.630	6.710	13.88	0.019

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

We further exploit the richness of our data to look at evidence for the importance of motivation and comparative advantage by investigating the impact of studying the pre-

ferred field on four different self-reported variables. First, we look at whether individuals report that they can identify themselves with their degree program and their satisfaction with the studies two or three years after university entry. Columns (3) and (4) of Table 11 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. This is likely to be related to individual motivation. Second, we analyse differences in terms of the number of hours spent studying in a regular week two or three years after university entry. Students who enrol in their preferred field spend on average four hours more per week studying. This higher effort may be related to a a better motivation, but also to a higher interest and passion in the subject. Third, we investigate the stated reasons behind university dropout. Those who study the preferred field are less likely to drop out from their university program because of low performance in the field (see columns (4)). This provides support for both the motivation and comparative advantage channel.

The evidence presented does not allow us to assess which of the two channels plays a more important role in explaining why studying the preferred field leads to a better academic performance. However, we find some suggestive evidence that interest in the subject and intrinsic motivation play a non-negligible role in explaining the higher university dropout rates among those not studying their preferred field. This is in contrast to some previous studies suggesting that studying the preferred program leads to better education and labor market outcomes due to a comparative advantage in the subject studied (Kirkeboen et al., 2016). Nevertheless, our findings cannot rule out that passion and motivation play the only role for study success, but they are less important in explaining differences in labor market outcomes.

7 Conclusion

We investigated 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 preferred field of study, we use an instrumental variables approach in which the probability to enroll in

the preferred field of study depends on the relative 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 are important drivers behind our results. Students who enrol in the 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. This is further supported by heterogeneous effects with respect to the broad field. We show that students who do not study their preferred field, but study a field that lies within the broad subject group, are also more likely to drop out of university. This suggests that having a comparative advantage in the preferred subject is not the only channel at play, since the skills required to perform well in related fields are likely to be similar. 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.

References

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

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

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

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

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

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

Blossfeld, H., Rosbach, H., and von Maurice, J. (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*, 14.

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

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

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

Danilowicz-Oesele, K., Lerche, K., Meya, J., and Schwager, R. (2017). Determinants of students' success at university. *Education Economics*, 25(5):513–532.

De Paola, M. and Gioia, F. (2012). Risk aversion and field of study choice: The role of individual ability. *Bulletin of Economic Research*, 64:0307–3378.

De Paola, M. and Scoppa, V. (2015). Procrastination, academic success and the effectiveness of a remedial program. *Journal of Economic Behaviour and Organisation*, 115:217–236.

Denzler, S. and 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).

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., and 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. and 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. and 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., and 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. *Discussion Paper Statistics Norway*, 725.

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

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

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R., and Sommersel, H. B. (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. and Gomez, L. N. (2008). Why do higher education students drop out? Evidence from Spain. *Education Economics*, 16(1):89–105.

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

OECD (2020). *Education at a Glance 2020: OECD Indicators*. OECD Publishing, Paris.

Ost, B., Pan, W., and 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. and Stichnoth, H. (2020). Fiscal and individual rates of return to university education with and without graduation. *Applied Economics Letters*, 28(16):1432–1435.

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

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

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

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

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

A Additional tables and figures

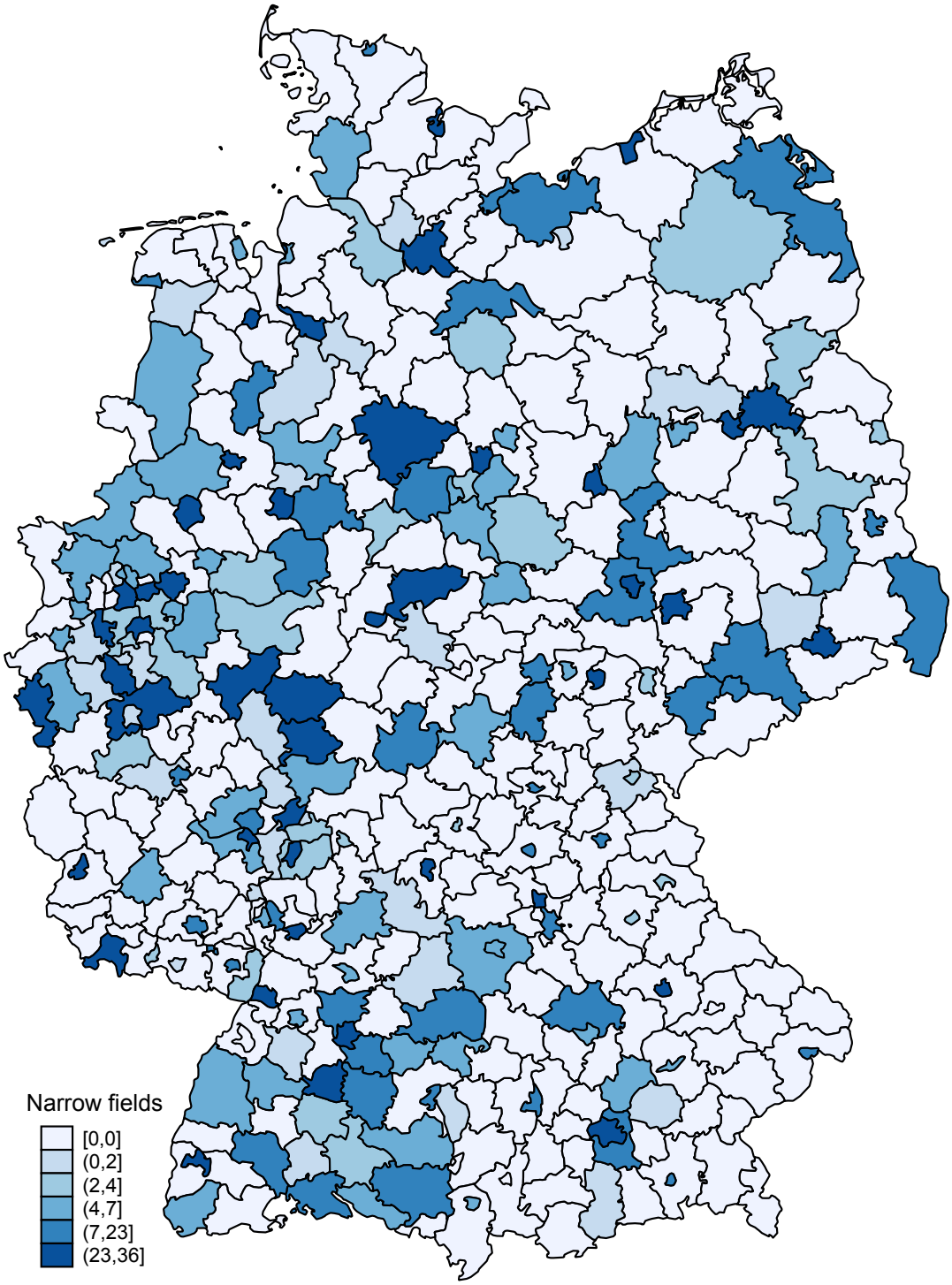


Figure A.1: Number of field of studies offered in German districts

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: 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***
	(14.92)	(10.16)	(10.15)	(10.41)	(10.49)	(10.76)	(9.36)
University of applied science	-0.154***	-0.226***	-0.227***	-0.205***	-0.200***	-0.203***	-0.174***
	(-6.99)	(-9.20)	(-9.29)	(-7.65)	(-7.41)	(-7.39)	(-6.50)
Teaching program	-0.103***	-0.164***	-0.168***	-0.137***	-0.135***	-0.135***	-0.109***
	(-5.35)	(-5.66)	(-5.77)	(-4.62)	(-4.51)	(-4.50)	(-3.98)
Bachelor degree	-0.044**	-0.084***	-0.083***	-0.092***	-0.091***	-0.096***	-0.081***
	(-2.42)	(-3.01)	(-2.98)	(-3.36)	(-3.32)	(-3.50)	(-3.35)
Age (university entrance)			0.004	0.009**	0.009***	0.009***	0.009**
			(1.64)	(2.42)	(2.62)	(2.62)	(2.32)
Female			0.046***	0.036**	0.038**	0.036**	0.034**
			(3.22)	(2.42)	(2.52)	(2.36)	(2.31)
Born in Germany			0.031	-0.015	-0.001	0.003	0.004
			(0.94)	(-0.44)	(-0.03)	(0.07)	(0.09)
School final grade				0.103**	0.100**	0.107**	0.085*
				(2.27)	(2.17)	(2.33)	(1.79)
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 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 labour 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.3: 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.4: First-Stage estimates by different groups

	high SES (1)	low SES (2)	Female (3)	Male (4)
Field availability	0.092***	0.106***	0.090***	0.108***
	(0.014)	(0.013)	(0.012)	(0.015)
Observations	1837	2079	2362	1554
R^2	0.255	0.236	0.254	0.254

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: 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)
School final grade	-0.016	-0.011	-0.076*	-0.074*	-0.014	-0.013
	(0.040)	(0.040)	(0.039)	(0.039)	(0.023)	(0.023)
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.6: 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.7: 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.8: Results on program dropout: different instrumental variables

	Baseline	Bandwith 100km	Bandwith 200km	Different kernel	Binary IV	Excl. fields	Weighted IV	Imputed 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, $*p < 0.10, **p < 0.05, ***p < 0.01$.

Table A.9: 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.

Download ZEW Discussion Papers:

https://www.zew.de/en/publications/zew-discussion-papers

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim
ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de