

Secondary Publication



Patzina, Alexander; Toussaint, Carina

Assessing the Importance of Sample Choice and Selectivity for Sex Segregation in College Majors : A Replication of Ochsenfeld (2016)

Date of secondary publication: 06.05.2024

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-951057

Primary publication

Patzina, Alexander; Toussaint, Carina (2023): „Assessing the Importance of Sample Choice and Selectivity for Sex Segregation in College Majors : A Replication of Ochsenfeld (2016)“. In: Zeitschrift für Soziologie, Vol. 52, Nr. 4, pp. 425-439, Berlin: de Gruyter Oldenbourg, doi: 10.1515/zfsoz-2023-2029.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

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

Alexander Patzina*, Carina Toussaint

Assessing the Importance of Sample Choice and Selectivity for Sex Segregation in College Majors: A Replication of Ochsenfeld (2016)

Die Bedeutung der Stichprobenwahl und -selektivität für Geschlechtersegregation in der Studienfachwahl: Eine Replikation von Ochsenfeld (2016)

<https://doi.org/10.1515/zfsocz-2023-2029>

Abstract: Ochsenfeld (2016) has found that a substantial part of sex segregation in higher education results from differences in vocational interests (i.e., preferences), while constraints (e.g., relative math grades) play only a minor role. We challenge the validity of these findings because earlier work employed a cross-sectional student sample and might therefore suffer from endogenous selection (i.e., post hoc rationalizations due to simultaneous reporting of majors and preferences) and postoutcome collider bias (i.e., conditioning on the outcome). Our replication study uses panel data (National Educational Panel Study, NEPS-SC4) that allow adjustment for the two sources of bias through the application of a pretransition preference measure and inverse probability weighting. Our analyses demonstrate the validity of prior research. Furthermore, our analysis indicates that the explanatory power of the overall model and the role of constraints for sex segregation in majors vary across the propensity of sample inclusion, thereby demonstrating the importance of sample composition for testing sociological theories.

Keywords: Inverse Probability Weighting; Postoutcome Collider; Propensity Score; Replication; Sample Selection.

Zusammenfassung: Ochsenfeld (2016) zeigt, dass ein wesentlicher Teil der Geschlechtersegregation in der Studienfachwahl auf Unterschiede in den beruflichen Interessen (d.h. Präferenzen) zurückzuführen ist, während sog. constraints (z.B. relative Mathematiknoten) nur eine geringe

Rolle spielen. Wir hinterfragen diese Ergebnisse, da die Befunde auf einer Querschnittsstichprobe von Studierenden basieren und folglich durch endogene Selektion (d. h. post-hoc-Rationalisierungen aufgrund gleichzeitiger Angabe von Studienfächern und Präferenzen) und postoutcome collider bias (d.h. Konditionierung auf Studierende) verzerrt sein könnten. Unsere Ergebnisse basieren auf Paneldaten (Nationales Bildungspanel, NEPS-SC4), in welchen Präferenzmaße vor der Studienfachwahl vorliegen und die die Anwendung von inverse probability weighting ermöglichen. Unsere Analysen bestätigen bisherige Forschungsergebnisse und zeigen zusätzlich, dass zentrale Befunde nach Einschlusswahrscheinlichkeit in die Stichprobe variieren. Dies verdeutlicht die Bedeutung der Stichprobensammensetzung bei der Überprüfung soziologischer Theorien.

Schlagworte: Inverse Probability Weighting; Postoutcome Collider; Propensity Score; Replikation; Stichprobenselektivität.

1 Introduction

In recent years, sociological science has identified the need for replication studies to increase the credibility of the field and to overcome its replication crisis (e.g., Auspurg & Brüderl 2022; Freese & Peterson 2017). This study contributes to this agenda and answers the call for replication by Ochsenfeld (2016, *Social Science Research*), who analyzed gender segregation in major choices in Germany. The original study found pronounced gender differences in major choices, which are largely a result of differences in vocational interests (i.e., preferences) between female and male high school graduates. To a much lesser extent, the approval of the choice of major by friends and the relative strength in mathematics also play a role in explaining gender segregation in college major choices. Other investigated mech-

*Korrespondenzautor: Alexander Patzina, Universität Bamberg, Institut für Soziologie; Feldkirchenstr. 21, 96052 Bamberg, E-Mail: alexander.patzina@uni-bamberg.de

Carina Toussaint, Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, Regensburger Straße 104, 90478 Nürnberg, E-Mail: carina.toussaint@iab.de

anisms that fail almost completely in explaining gender differences are job values, anticipated discrimination and parental approval of the choice of major.

These prior findings are important because they indicate that young men and women do not choose different college majors because of environmental factors or given social norms. These findings are in line with a value-based explanation. Therefore, gender differences in major choices are a result of long-lasting socialization processes through early and young childhood that lead to an internalization of gender-specific interests. However, the central problem, which motivates a reanalysis of prior work, is that these findings may be flawed because of two methodological issues the original study (Ochsenfeld 2016) could not address.

First, the original study measured vocational interests and college major choices simultaneously. Therefore, it remains unclear whether major choice influences vocational interests or whether vocational interests explain college major choices. The author of the original study was well aware that measuring preferences (i.e., vocational interests) and major choices simultaneously might introduce bias because “... in the presence of post-hoc rationalization, stated interests could be endogenous to the choice of a major” (Ochsenfeld 2016: 127). Our replication addresses this endogenous selection problem because we rely on a pretransition preference measure that cannot be affected by actual major choices.

Second, the original study relied on a sample of university students. In recent years, sociological research has emphasized the important role sample selection can play in distorting substantial conclusions and theory testing (e.g., Elwert & Winship 2014). This problem might be particularly important for the validity of prior work on gender segregation in higher education if enrolling in university depends on a match between individuals’ vocational interests and college major characteristics and the actual college major choice. If this is the case, prior results suffer from postoutcome collider bias (Elwert & Winship 2014), and a substantial conclusion drawn from earlier work might change. Thus, it remains unclear whether the choice of a sample affects prior findings. In this replication, we address this issue and employ a data set that includes graduates from high schools, which therefore also includes individuals who are not in university education. This feature of the newly employed data set enables us to implement an inverse probability weighting approach and assess the potential role of postoutcome collider bias (i.e., the choice of the sample) in the original study (Breen & Ermisch 2021).

In addition to the replication setup, we further use the propensity score from our inverse probability weighting approach to illustrate the potential influence of the sample

choice for substantial conclusions. This represents an extension of the replication and aims to illustrate the importance of sample selection and how it can distort substantial conclusions regarding sex segregation in higher education. Therefore, this paper comprises two parts. The first part constitutes the replication, which focuses on potential biases in the original study and forms the core of this paper. The second part is conceived as an extension of the replication and effectively demonstrates the importance of considering sample selection in a more general context. Consequently, our replication study asks the following two main research questions: Are we able to replicate previous work on gender segregation in college majors based on a new data set? Are we able to extend this work to assess how the severity of sample selection may affect substantial conclusions?

To answer these questions, our replication study uses prospective panel data from the starting cohort 4 of the German National Educational Panel Study (NEPS) (Blossfeld et al. 2011), which comprises transitions after graduating from high school. In contrast, the original study relied on a sample of university students from the starting cohort 5 of the NEPS. Our model specification follows the Ochsenfeld (2016) specification as closely as possible, and we use Ochsenfeld’s data preparation and analytical files in our replication (Ochsenfeld 2015). Therefore, according to Freese & Pettersen (2017), our study falls into the cluster of more or less “classical” replication studies that aim for the repeatability of prior findings. Thus, although our study uses a different data set with a different temporal structure (i.e., a panel data set, which allows correcting for two sources of collider bias), all operationalizations of empirical constructs and the main analytical approach follow the specification of the original study. In addition, our study includes an extension that goes beyond a classical replication and seeks to further illustrate the central role of sample selection.

The overall aim of our replication study is to improve our understanding of inequality in educational decision making. This is of utmost importance because such decisions early in life have long-lasting consequences for educational careers (e.g., Kerckhoff & Glennie 1999) and the life course development of educational inequality in several life domains (e.g., Kratz & Patzina 2020; Leopold & Leopold 2018; Tamborini et al. 2015). Labor market research indicates that segregation in the choice of major largely contributes to gender wage gaps (e.g., Brown & Corcoran 1997; Machin & Puhani 2003; Paglin & Rufolo 1990) and constitutes a driving force of gender segregation in modern labor markets (e.g., Cech 2013; Reskin 1993), which has hardly changed in recent decades (e.g., Martin-Caughey 2021). Thus, replicating prior research helps to validate our understanding of how social inequalities emerge in modern societies.

2 The Ochsenfeld (2016) Study and Potential Sources of Bias

2.1 Brief Summary of the Ochsenfeld (2016) Study

In 2016, Ochsenfeld published a study that examines the factors explaining the persisting gender-based segregation in college majors. Specifically, the study investigates the explanatory power of preferences (i.e. vocational interests) and constraints (e.g., approval of the choice of major by significant others, the relative strength in mathematics, anticipated working hours) for gendered major choices.

The study employs conditional logit models (McFadden 1974), grounded in the theoretical concept that major choices are essentially a match between individual-level factors (e.g., vocational interests, job values, relative math grades) and college major characteristics (e.g., mathematical content, wage prospects, overwork norms). Thus, the analysis considers both individual variables reflecting the preferences and constraints of college entrees as well as structural variables representing college major attributes of 23 different fields.

Ochsenfeld (2016) obtains the individual-level data from the NEPS SC5 sample of college students. The sample size of the individual data comprises 9,109 cases. Structural data regarding the 23 majors come from the datasets NEPS SC5, Micro Census 2007/2008/2009, and the HIS graduate panel studies 1997/2001.

The core of the analysis is a decomposition analysis based on conditional logit models and designed to assess the explanatory power of each examined mechanism for gender-based segregation. This decomposition reveals that the investigated preferences and constraints can account for approximately 46 % of the gender gap in major choices. 37 % of the overall gap is explained by gender differences in vocational interests. To a lesser extent, factors such as peer approval of major choice (4 %) and relative strength in mathematics (5 %) also contribute to the gender gap in college major choices. Job values, anticipated discrimination, and parental approval of the major choice appear to have minimal to no influence on gender differences.

These highly relevant findings suggest that women and men choose different majors not primarily due to environmental or social factors but as a result of internalizing gendered vocational interests. However, the implementation of this study faces methodological problems that could not be addressed due to data limitations at the time of the Ochsenfeld (2016) study. These problems could potentially introduce biases, emphasizing the need for a replication.

2.2 Potential Sources of Bias

In this study, we refer to the terminology of biases introduced by Morgan & Winship (2007) and Elwert & Winship (2014) to sociology. Typically, every research design faces the challenges of overcoming common confounding bias, overcontrol bias, and endogenous selection bias (e.g., Elwert & Winship 2014; Morgan & Winship 2007). While common confounding bias refers to the omission of important variables that influence both the outcome and the explanatory variable, overcontrol bias refers to the inclusion of mediators, i.e., important variables that explain a relationship between an outcome and an explanatory variable. In this replication study, however, we concentrate on endogenous selection bias, i.e., conditioning on collider variables. Collider bias emerges when a model conditions on "... a variable that is itself caused by two other variables, one that is (or is associated with) the treatment and another that is (or is associated with) the outcome." (Elwert & Winship 2014: 31).

In the following, we show in stylized directed acyclic graphs (DAGs) how collider bias might affect conclusions from previous work. In doing so, we briefly show how (i) measurement error and (ii) sample selection might introduce bias in the estimates presented by Ochsenfeld (2016).

2.2.1 Post Hoc Rationalizations

A possible form of bias displays Figure 1 in a stylized DAG, which does not take unconfounding conditions between the variables into account because we want to focus on collider bias issues and not on common confounding bias. Additionally, we want to stay close to the original study, which also used stylized DAGs. Figure 1 shows that vocational interests are measured at a time when the choice of a major has already taken place—a circumstance that may introduce collider bias (Elwert & Winship 2014) and that was already mentioned by the original study (Ochsenfeld 2016: 129). According to theory, gender (X) (i.e., gender-specific socialization and internalization processes over the early life course) leads to the formation of gender-specific vocational interests (P_o). These interests, which have been developed prior to the actual decision, cause the choice of a specific major (J) that matches these interests at time t_o (prior to the choice of a major). The choice (J) subsequently leads to the observed male- or female-dominated major (Y).

However, as depicted in Figure 1, the NEPS-SC5 data employed by the Ochsenfeld (2016) study measure vocational interests not at t_o but at the same time as the outcome (t_r). Thus, previous findings may suffer from collider bias (i.e., endogenous selection) if the major choice and pretransi-

tion interests influence respondents' answers to preference questions at the time of the survey (t_1). In other words, prior research used the vocational interests measured in t_1 as if they were measured in t_0 .

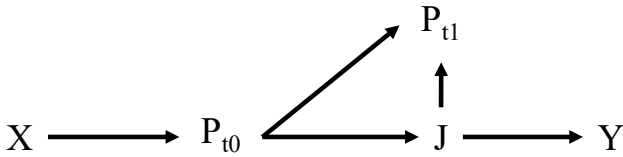


Figure 1: Stylized DAG depicting collider bias due to post hoc rationalizations.

Note: For simplification and demonstration, we do not assume (unobserved and observed) common causes in this DAG. Therefore, we omitted the unconfounding conditions.

The data used in this replication do not suffer from endogenous selection bias that emerges due to the conditioning of a collider variable, which opens the noncausal pathway $P_{t0} \rightarrow P_{t1} \leftarrow J$. For instance, if an individual studies pedagogy who did not have a particularly strong social interest before, he or she may nevertheless report a high social interest after starting the study. This can be explained by post hoc rationalization or an actual shift in interests as a result of studying. However, both of these factors arose after the student began the studies and therefore do not represent a causal explanation for the student's major choice. The new data set measures vocational interests (i.e., preferences) prior to the transition (P_{t0}). Therefore, in our replication, we do not condition on a collider variable, and post hoc rationalizations of respondents cannot bias our results.

2.2.2 Sample Choice

Another form of bias that was not addressed in the original study could emerge from conditioning on a postoutcome collider. Because the Ochsensfeld (2016) study relies on a data set that incorporates only university students, a conditioning on the outcome problem might occur (Elwert & Winship 2014). Figure 2 depicts the central problem in a stylized DAG.

In this stylized DAG, survey participation (S) constitutes a collider variable along the noncausal pathway $M \rightarrow S \leftarrow Y$ (M , match of individuals' preferences and major characteristics; Y , university enrollment and major choice; S , survey participation). This form of postoutcome collider bias is likely to occur because in student data, we only observe individuals enrolled in university and who have a (certain) match between their preferences and major characteristics. However, certain matches between individ-

uals' preferences and major characteristics may divert individuals from studying for a bachelor's degree and may lead to the receipt of vocational training (see Allmendinger 1989 for a classification of the German education system). If this is the case, sample selection (i.e., conditioning on studying) could introduce postoutcome collider bias and affect the validity of prior findings.

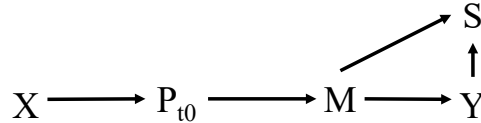


Figure 2: Stylized DAG depicting postoutcome collider bias due to sample selection.

Note: For simplification and demonstration, we do not assume (unobserved and observed) common causes in this DAG. Therefore, we omitted the unconfounding conditions.

Using prospective panel data that incorporate transitions from high school to different postsecondary educational states (i.e., NEPS-SC4 data) allows us to implement an inverse probability weighting (IPW) approach that accounts for sample selection. Breen and Ermisch (2021) have recently demonstrated that in many cases, IPW may reduce postoutcome collider bias.

3 Replication Set Up

3.1 Data Sets, Models and Variables

The main difference between the replication and the Ochsensfeld (2016) study involves the use of prospective panel data from the NEPS starting cohort 4 (NEPS, SC4, version 10.0.0)¹, which includes pretransition information, whereas the original study relied on NEPS starting cohort 5 data (NEPS, SC5, version 4.0.0), which include only university students (see the "Data overview" chapter in the appendix for a brief overview of the data). The NEPS SC4 panel started with 9th graders from all educational tracks in Germany. Consequently, these data survey individuals both during their school years and during the transition into higher education, vocational education and the job market. Because both data sets are provided by the same institute (Leibniz Institute for Educational Trajectories), all employed individual-level survey items used in our replication are the same

¹ Starting Cohort 4–9th grade, doi:10.5157/NEPS:SC4:10.0.0.

as in the original study. To further enhance comparability across our study and the replicated study, our data preparation used the data preparation do-files of the original study (Ochsenfeld 2015). This constitutes an ideal setting to address the outlined collider bias problems with a new data source.²

A key innovation of the original study is the accurate test of the theoretical considerations by means of conditional logit models (McFadden 1974). The key theoretical argument is that major choices reflect a match between individual-level (e.g., vocational interests) and major characteristics (e.g., math intensity). The conditional logit model enables scholars to test this theoretical process. In our replication, we specify the following model thereby using the Ochsenfeld (2016) specification:

$$p_{ik} = \frac{\exp(X_{ik}\beta)}{\sum_{j=1}^J \exp(X_{ij}\beta)}$$

The term p_{ik} indicates the probability for an individual to choose a major k . J represents the number of alternatives (i.e., 23 college majors)³, and X_{ij} is a vector of the independent variables. The major characteristics can be considered predictors in the model. However, characteristics that do not vary between the 23 majors cannot be included. Thus, we need information about the values or costs individuals assign to each major. Since we do not have any direct measurements available for this, we interact the individual characteristics with the major characteristics (Shauman 2006). This ensures, for example, that math-intensive fields of study have a higher value or lower cost for individuals with relative math strength than less math-intensive fields.

The conditional logit considers the individual value for each major, not only for the chosen one (McFadden 1974). Moreover, for the conditional logit models, we need to translate the gender differences in major choice into a structural characteristic (Ochsenfeld 2016). Therefore, we created a variable that reflects the proportion of female students in each major (*prop.female*). This structural variable is interacted with individuals' gender (*female x prop.female*).

The key part of the analyses is the decomposition based on the conditional logit models. The decomposition is a me-

diation analysis with a baseline model to which we add the theoretically identified mechanisms and observe how much the coefficient *female x prop.female* diminishes. The modeling strategy in our replication is exactly the same as in the original study by Ochsenfeld (2016) and differs only in the employed individual-level data (NEPS-SC4 vs. NESP-SC5).

Our replication sets up all variables according to the original Ochsenfeld (2016) study (for an overview of all operationalizations, see Table A2 in the appendix). The crucial variable of vocational interests is measured by RIASEC scores, which are intended to approximate individuals' preferences for different (job) tasks. Vocational interests are classified into six dimensions: realistic, investigative, artistic, social, entrepreneurial, and conventional. It is important to note that all individual-level characteristics stem from the new prospective NEPS- SC4 data and that they are the same constructs as in the original study (see Table A2 in the Appendix).

As outlined above, the theoretical and methodological test requires operationalization of the structural characteristics of majors. To ensure comparability, we use the same data sources as in the original study to create the structural variables.⁴ Our data sources include NEPS-SC5 (version 4.0.0), German Micro Census (2011, 2012, 2013), and the HIS graduate panel (1997, 2001). The operationalization of the structural variables is the same as in the analyses of Ochsenfeld (2016) (for an overview, please refer to the chapter "Operationalization of the constructs used" in the appendix for more details).

Importantly, to retrieve major characteristics, we aggregated (group means) individual data from respondents who studied one of the 23 majors. To keep measurement error low, analogous to Ochsenfeld (2016), we only use majors for which the computation of group means is based on at least 100 observations.

3.2 Analytical Approach of Replication Study

The analytical strategy for our replication is straightforward. As mentioned earlier, we constructed each employed variable as in the Ochsenfeld (2016) study. The key difference is the application of prospective panel data. These data entail the same preference measure as the original study. However, preferences were measured prior to transitions into university education. To elaborate on whether the importance of preferences depends on the timing of meas-

² Another possible source of data would have been the 2009/10 HIS college eligible panel provided by the German Centre of Higher Education Research and Science Studies. However, other individual characteristics included in the Ochsenfeld (2016) specification are missing in this data source. For this reason, the NEPS-SC4 data set appears to be the most suitable panel data set for this replication.

³ The classification in the NEPS data follows the Federal Statistical Office's 2-digit classification scheme.

⁴ Note that we use current versions of the data sources if possible.

urement, we first present our baseline replication results, which we display next to findings from Ochsensfeld (2016).

Second, to investigate whether the importance of preferences and constraints in explaining gender differences in college majors depends on sample selection, we show the results from a model that takes such sample selection into account. A logistic regression model estimates the selection into treatment (i.e., treatment model). To this end, we use the population of all individuals with a standard university entrance qualification⁵ in the data set (N=2,700). The dependent variable of the treatment model is a binary indicator for studying or not after completing high school. To retrieve the propensity score, we regress on the binary outcome gender (dummy indicator), migration background (dummy indicator and a dummy for missing value), parental socioeconomic status (dummy indicator for at least one parent with university education and a dummy variable for missing cases), the six RIASEC scores (composite index with a range from zero to one for each of the six interest dimensions: realistic, investigative, artistic, social, entrepreneurial, and conventional), two metric job value indicators (one approximates a preference for high earnings, while the other proxies a preference for a good work-life balance), and two central school performance measures (metric math and German grades). The results from the matching equation are in Table A3.1 in the appendix. After predicting the conditional probability of studying, we generate the inverse probability weights as $P(T=1 | \text{covariates})$ if $T = 1$ (individual starts studying) and $1 - P(T=1 | \text{covariates})$ if $T = 0$ (individual does not start studying) (Hernán & Robins 2020). The mean of the derived inverse probability weights is 1.99 with a standard deviation of 1.35. The minimum of the weights is 1.05, and the maximum is 11.69. The conditional logit models (i.e., outcome models) are weighted with the inverse of the conditional probability of studying for each individual. The weighting of each individual in the outcome models creates a pseudopopulation that considers individual heterogeneity across the treatment and control groups (Hernán & Robins 2020). Therefore, sample selection based on the included variables in the treatment equation no longer biases the estimation.

Third, to elaborate on whether the importance of preferences and constraints in explaining gender differences in college majors depends on the selectivity of a sample, we use the propensity score to showcase how the gender gap and the explanatory power of the Ochsensfeld (2016) specification vary across its distribution. We provide a model

that splits the sample into above- and below-median sample inclusion probability and a model that splits the sample into four different subsamples corresponding to four quantiles of the propensity score distribution. In so doing, we are able to speculate the extent to which certain skewed samples have the potential to distort substantial conclusions (i.e., the predictive power of sociological theories).

4 Results – Answering the Call for Replication

4.1 The Role of Post Hoc Rationalizations and Sample Choice for the Validity of Prior Findings

Table 1 shows the main results of our replication. The first column includes the original Ochsensfeld (2016) results. The second column shows our replication results with the new vocational preference measure. The third column shows our results additionally accounting for potential sample selection using IPW. The first row of Table 1 indicates that our point estimates are very close to those of the Ochsensfeld (2016) study, already hinting toward strong validity of the earlier findings. When turning to the second row of Table 1, our results indicate that the replication models also explain a substantial part of the gender differences in major choices (52.8 percent in column 2 and 53.4 percent in column 3, respectively).⁶ The explanatory power of the Ochsensfeld (2016) specification even increases slightly when using panel data that account for the two potential sources of collider bias associated with the use of cross-sectional university student data.

Regarding the preference measures employed, our replication results (columns 2 and 3 of Table 1), which use an interest measure that cannot be affected by actual major choices, indicate that the explanatory power of these measures is hardly affected. Gender differences in vocational interests are therefore the major driver of sex segregation in the choice of major in Germany. The constraint part of Table 1 also indicates that our replication findings are more or less in line with the original study. However, one substantial difference emerges. The last row of Table 1 shows negative values for the match between female gender and

⁵ We condition on standard university entrance qualifications to be consistent with the original study.

⁶ The decomposition is a mediation analysis that determines the effect of the mechanisms by adding the corresponding variables to the baseline model. This allows us to observe how much the coefficient of the variable *female x prop.female* decreases.

Table 1: Main results – Addressing post hoc rationalizations and sample choice

	Ochsenfeld (2016)		Replication (SC4)		Replication (SC4) with IPW	
	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹
Baseline model	1.0452	0	1.0424	0	1.0434	0
Full model (all variables)	1.0244	46.0	1.0200	52.8	1.0202	53.4
<i>Preferences</i>						
voc. interests	1.0284	37.2	1.0249	41.2	1.0261	39.8
earning well x wage level	1.0452	0.0	1.0421	0.5	1.0431	0.7
<i>Constraints</i>						
pleasant work hrs. x overwork norms	1.0452	0.0	1.0424	-0.0	1.0434	-0.0
fem. x diff. parents' approval	1.0454	-0.5	1.0419	1.2	1.0432	0.5
fem. x diff. friends' approval	1.0434	4.1	1.0409	3.6	1.0416	4.1
rel. math grade x math int.	1.0428	5.4	1.0395	6.7	1.0406	6.4
fem. x discrimination	1.0449	0.7	1.0443	-4.7	1.0446	-2.9
N	9,109		1,936		1,936	

Note: ¹percentage explained when adding term to baseline model. The replication analysis (shown in the column “Replication (SC4)”) uses, as does the original study, weights provided by the data provider that take into account the complex sampling design and survey attrition rates. The results of the Ochsenfeld (2016) study are presented in Table 4 page 130. The table here does not display the detailed results on all preference measures. Data: NEPS-SC4, NEPS-SC5, Micro Census, HIS graduate panel studies.

majors with high levels of anticipated discrimination. This is somewhat surprising because this result indicates that if women choose majors with high levels of anticipated labor market discrimination (i.e., male-dominated fields), sex segregation increases. One possible explanation could be that the data sets employed to derive the discrimination variable (HIS data from the late 1990s and early 2000s) are outdated. Unfortunately, more recent starting cohorts of the HIS graduate panel no longer include the corresponding variable. Moreover, measuring discrimination as agreement with the statement that the “right gender” is important for finding a job is not optimal. In occupations with a high proportion of women, the female gender could be considered the “right” gender. This would make the measurement ambiguous. Thus, future analyses could search for more up-to-date data and a cleaner measurement of discrimination against women.

4.2 Assessing the Importance of Sample Choice

Table 2 again shows the original Ochsenfeld (2016) results in the first column. The second column shows results based on a sample of individuals with a study propensity below the median, while the third column shows results based on a sample comprising only individuals with a study propensity equal to or above the median. In so doing, we want to

showcase how sample selection may distort substantial conclusions on sex segregation in higher education.

In contrast to our main results (provided in Table 1), Table 2 indicates pronounced differences in the point estimates between the original results presented in column 1 and the replication results in columns 2 and 3 that rely on a sample with a low and high study propensity. Regarding sex segregation, our results indicate a more pronounced gender gap among individuals with a low study propensity, while the gender gap is less pronounced among individuals with a high study propensity.

The influence of potential sample selection on substantial conclusions becomes even more evident when turning to the explanatory parts of Table 2. As row two indicates, the explanatory power of the Ochsenfeld (2016) specification also varies by the study propensity. While the original specification accounts for only 32.6 percent of the variation between the genders among individuals with a low study propensity, it accounts for 67.0 percent of sex differences among individuals with a high study propensity. While this result might also have substantial implications, it reveals the importance of sample selection, which has the potential to substantially change our understanding of sex segregation in major choices.

Regarding preferences, our replication results indicate that among a sample of individuals with a high study propensity, the explanatory power of preferences increases, while it remains roughly the same among a sample of indi-

Table 2: Original results and results for low and high values of sample inclusion probability

	Ochsenfeld (2016)		pscore < median		pscore >= median	
	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹
Baseline model	1.0452	0	1.0495	0	1.0366	0
Full model (all variables)	1.0244	46.0	1.0334	32.6	1.0121	67.0
<i>Preferences</i>						
voc. interests x (math intensity, care)	1.0284	37.2	1.0312	36.9	1.0187	48.8
earning well x wage level	1.0452	0.0	1.0494	0.3	1.0362	1.1
<i>Constraints</i>						
pleasant work hrs. x overwork norms	1.0452	0.0	1.0495	-0.0	1.0366	-0.0
fem. x diff. parents' approval	1.0454	-0.5	1.0483	2.5	1.0365	0.3
fem. x diff. friends' approval	1.0434	4.1	1.0538	-8.6	1.0326	10.9
rel. math grade x math int.	1.0428	5.4	1.0481	2.9	1.0336	8.3
fem. x discrimination	1.0449	0.7	1.0507	-2.5	1.0394	-7.5
N	9,109		968		968	

Note: ¹ percentage explained when adding term to baseline model. The replication analysis uses weights, as does the original study, provided by the data provider that take into account the complex sampling design and survey attrition rates. The results of the Ochsenfeld (2016) study are presented in Table 4 page 130. The table here does not display the detailed results on all preference measures. Data: NEPS-SC4, NEPS-SC5, Micro Census, HIS graduate panel studies.

viduals with a low study propensity (48.8 percent in column 3 and 36.9 percent in column 2, respectively).

The role of constraints changes substantially. Across the two samples, point estimates differ substantially, and the explanatory power also differs. The first pronounced difference emerges when turning to the role of approval by friends. While approval of friends substantially (10.9 percent) reduces the gender gap in a sample of individuals with a high study propensity, friends' approval increases sex segregation in higher education among a sample with individuals with a low study propensity. In addition, in a sample of individuals with a high study propensity, the relative math grade appears to be almost three times as important as in a sample with individuals with a low study propensity (percentage explained in column 3 8.3 in comparison to 2.9 in column 2). Again, the discrimination interaction is negative, which, as mentioned earlier, is a somewhat counterintuitive finding. Nevertheless, the negative influence is more pronounced in the sample with individuals with a high study propensity (see last row of columns 2 and 3). Figure 3 additionally shows that the results tend to differ even more when we further differentiate the likelihood of sample inclusion. To this end, Figure 3 shows how the results of the Ochsenfeld (2016) specification differ when we split the sample into four equal parts representing the quartiles of the sample inclusion probability. Note that these samples become rather small. Thus, the results presented in Figure

3 represent stylized findings that should be interpreted with caution. However, they enable us to illustrate how potential sample selectivity may distort earlier findings.

Figure 3 shows that the gender gap (black solid line) in major choice is lowest among a highly selective sample of individuals with a very high study propensity (above the 75th percentile), whereas it is highest for the group of individuals with a very low study propensity (under the 25th percentile). Furthermore, Figure 3 shows that the explanatory power (black dashed line) of the specification also tends to differ over the study propensity. While the contribution of constraints (gray dashed line) to sex segregation also varies by study propensity, the explanatory power of prefer-

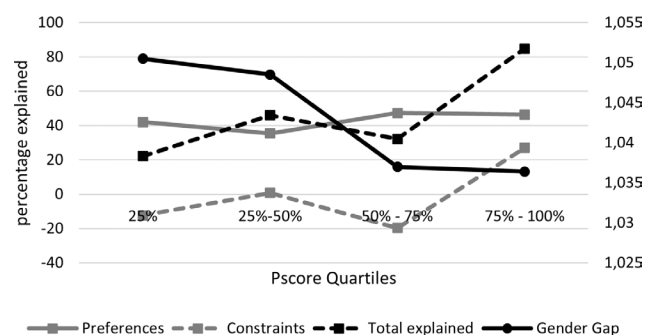


Figure 3. Gender gap and explanatory power of the Ochsenfeld (2016) specification across sample inclusion probability. Data: NEPS-SC4, NEPS-SC5, Micro Census, HIS graduate panel studies.

ences (gray solid line) does not change substantially across potential sample selectivity. Overall, we find evidence that previous findings of sex segregation could be biased by neglecting the heterogeneity of groups with a high/low study propensity. This finding may stimulate future work in the field of educational sociology.

5 Limitations

Although we are convinced that our replication approach makes an important contribution to the scientific discourse, we would like to note a few limitations in the following.

While IPW is an appropriate method to reduce sample selection bias based on observables, this approach fails to account for selection based on unobservables (Breen & Ermisch 2021; Hernán & Robins 2020). Thus, sample selection into studying based on unobserved factors (such as mental health or educational motivation), which also played a role in the original study, could still bias our employed IPW approach.

Another limitation of our replication is that the NEPS-SC4 data set contains significantly fewer cases of university students. While the original analyses by Ochsenfeld (2016) relied on 9,109 individuals, the replication analyses use only 1,936 cases. Given the rich set of mechanisms and the small-scale distinction into 23 different majors, a high case number is advantageous. The lower case number could become problematic, especially when splitting the sample by quartiles (chapter 4.2).

In addition, we want to emphasize that both the original study and our replication only explain approximately 50 percent of the sex segregation in college major choices. Further research is therefore needed to close this gap in the explanation of the differences between women's and men's college major choices.

6 Conclusion and Implications

In this study, we replicated earlier findings on sex segregation in major choices in Germany provided by Ochsenfeld (2016; *Social Science Research*). Our replication study identified two potential sources of collider bias (i.e., post hoc rationalizations due to the simultaneous measure of preferences and college majors and sample selection), which challenged earlier findings. As earlier work relied on a sample of university students (NEPS-SC5), these two forms of bias could not be addressed. In our replication study, we used novel panel data (NEPS-SC4) that enabled us to address both

collider bias problems through the application of a pretransition measure, which could not be affected by post hoc rationalization (i.e., the actual choice of a college major), and inverse probability weighting (IPW), which has the potential to solve postoutcome collider bias (Breen & Ermisch 2021). In a last step, our replication used the retrieved propensity scores from the IPW approach to showcase how substantial conclusions on sex segregation in higher education depend on the selectivity of samples.

The most important result from our replication study is that Ochsenfeld's (2016) findings are not biased by the two forms of collider bias studied here. Therefore, gender differences in preference (i.e., vocational interests) are the most important factors in explaining sex segregation in higher education. The results suggest that vocational interests are already stabilized by the time of transition from school and that post hoc rationalization no longer occurs. In line with this, Low & Rounds (2007) as well show that the stability of interests increases sharply from the age of 16. Our last analytical step (i.e., exploring the Ochsenfeld (2016) specification across the distribution of the propensity score) strengthens prior findings on the role of preferences because their contribution to sex segregation is barely affected by the likelihood of sample inclusion. The fact that the IPW approach does not lead to significantly different results means that individual vocational interests only influence the choice of majors and not the decision whether someone studies or starts a vocational training. Accordingly, sample selection does not cause a problem when using a pure student sample for the general analyses. However, in terms of the explanatory power of the mechanisms, one can see that it is nevertheless important to consider the selection of the sample. This workaround indicated that the gender gap and the explanatory power of the Ochsenfeld (2016) specification vary strongly across the likelihood of sample inclusion, our results showcase the potential of sample selection for testing sociological middle range theories.

With this study, we add a piece to the puzzle to overcome the replication crisis in the sociological sciences (e.g., Auspurg & Brüderl 2022; Freese & Peterson 2017). In addition, our study emphasizes the importance of sample selectivity for testing predictions of middle range theories. This might constitute a somewhat obvious fact within sociology. However, many studies within the social sciences rely on convenience samples (e.g., entry cohorts of U.S. elite universities or high school classes within a city or metropolitan area). Of course, such studies are of high importance because research tailored data sets enable scholars to test theoretical mechanisms. Nevertheless, our results suggest that substantial conclusions may depend on the sample inclusion probabilities of individuals. Therefore, every study

should speculate or, if possible, empirically investigate how far postoutcome collider bias (i.e., sample selection) may distort theory testing. This appears particularly important during phases of societal crises (e.g. COVID–19 pandemic), i.e., during periods in which much scientific knowledge builds on convenience samples.

Replication File: Our replication package is available at the Harvard Dataverse: <https://doi.org/10.7910/DVN/ZXZ6FP>

Acknowledgments: We thank Matthias Collischon, Fabian Ochsenfeld, Tobias Wolbring and the anonymous referees for their helpful comments. A previous version of this article was presented at the fall meeting of the DGS-Methodensektion in 2021. Furthermore, we would like to thank Vanessa Kunze for excellent student assistance.

References

- Allmendinger, J., 1989: Educational systems and labor market outcomes. *European Sociological Review* 5: 231–250.
- Auspurg, K. & J. Brüderl, 2022: How to increase reproducibility and credibility of sociological research. S. 512–527 in: K. Gërkhani, N.D. de Graaf & W. Raub (Hrsg.), *Handbook of Sociological Science: Contributions to Rigorous Sociology*. Cheltenham: Edward Elgar Publishing.
- Blossfeld, H.-P., H.-G. Roßbach & J. von Maurice, 2011: Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft* 14 [Special Issue].
- Breen, R. & J. Ermisch, 2021: Using Inverse Probability Weighting to Address Post-Outcome Collider Bias. *Sociological Methods & Research* 0(0).
- Brown, C. & M. Corcoran, 1997: Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics* 15: 431–465.
- Cech, E. A., 2013: The self-expressive edge of occupational sex segregation. *American Journal of Sociology* 119: 747–789.
- Elwert, F. & C. Winship, 2014: Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology* 40: 31–53.
- Fabian, G. & K.-H. Minks, 2006: *Dokumentation des Scientific Use Files „HIS-Absolventenpanel 1997“*. Hannover: HIS.
- Freese, J. & D. Peterson, 2017: Replication in social science. *Annual Review of Sociology* 43: 147–165.
- Hernán, M. A. & J. M. Robins, 2020: *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Kerckhoff, A. C. & E. Glennie, 1999: The Matthew effect in American education. *Research in Sociology of Education and Socialization* 12: 35–66.
- Kratz, F. & A. Patzina, 2020: Endogenous Selection Bias and Cumulative Inequality over the Life Course: Evidence from Educational Inequality in Subjective Well-Being. *European Sociological Review* 36: 333–350.
- Leopold, L. & T. Leopold, 2018: Education and health across lives and cohorts: a study of cumulative (dis) advantage and its rising importance in Germany. *Journal of Health and Social Behavior* 59: 94–112.
- Low, K.-S. & J. Rounds, 2007: Interest change and continuity from early adolescence to middle adulthood. *International Journal for Educational and Vocational Guidance* 7: 23–36.
- Lüttinger, P. & T. Riede, 1997: Der Mikrozensus: amtliche Daten für die Sozialforschung. *ZUMA Nachrichten* 21: 19–43. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-208344>.
- Machin, S. & P.A. Puhani, 2003: Subject of degree and the gender wage differential: evidence from the UK and Germany. *Economics Letters* 79: 393–400.
- Martin-Caughey, A., 2021: What’s in an occupation? Investigating within-occupation variation and gender segregation using job titles and task descriptions. *American Sociological Review* 86: 960–999.
- McFadden, D., 1974: Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Hrsg.), *Frontiers in Econometrics*. New York: Academic Press.
- Morgan, S. L. & C. Winship, 2007: *Counterfactuals and causal inference*. Cambridge: Cambridge University Press.
- Ochsenfeld, F., 2015: Replication data for: preferences, constraints, and the process of sex segregation in college majors: a choice analysis. Harvard Dataverse, V1. <http://dx.doi.org/10.7910/DVN/HGCRRE>.
- Ochsenfeld, F., 2016: Preferences, constraints, and the process of sex segregation in college majors: A choice analysis. *Social Science Research* 56: 117–132.
- Paglin, M. & A.M. Rufolo, 1990: Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics* 8(1, Part 1): 123–144.
- Reskin, B., 1993: Sex segregation in the workplace. *Annual Review of Sociology* 19: 241–270.
- Schramm, M. & S. Beck, 2010: *Dokumentation des Scientific Use Files „HIS-Absolventenpanel 2001“*. Hannover: HIS.
- Shauman, K.A., 2006: Occupational sex segregation and the earnings of occupations. What causes the link among college-educated workers? *Social Science Research* 35: 577–619.
- Tamborini, C. R., C. Kim & A. Sakamoto, 2015: Education and lifetime earnings in the United States. *Demography* 52: 1383–1407.

Autor:innen

Alexander Patzina
 Universität Bamberg
 Institut für Soziologie
 Feldkirchenstr. 21
 96052 Bamberg
alexander.patzina@uni-bamberg.de

Alexander Patzina, geb. 1988 in Nürnberg. Studium der Wirtschaftswissenschaften und Geografie in Erlangen und Sozialökonomie in Nürnberg. Promotion in Nürnberg. Seit 2014 wissenschaftlicher Mitarbeiter am Institut für Arbeitsmarkt- und Berufsforschung und seit 2022 akademischer Rat a. Z. an der Universität Bamberg. Forschungsschwerpunkte: Ungleichheit im sozialen Vertrauen; Gesundheit(-sverhalten) und Wohlbefinden; Bildungsentscheidungen; Lebensverlaufsforchung. Wichtigste Publikationen: COVID–19 and Gender Differences in Social Trust: Causal Evidence from the First Wave of the Pandemic. *Socius* 8,

2022 (mit M. Collischon), Endogenous selection bias and cumulative inequality over the life course: Evidence from educational inequality in subjective well-being. *European Sociological Review* 36, 2020: 333–350 (mit F. Kratz), Early careers of dropouts from vocational training: Signals, human capital formation, and training firms. *European Sociological Review* 36, 2020: 741–759 (mit G. Wydra-Somaggio).

Carina Toussaint

Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit
Forschungsbereich „Bildung, Qualifizierung und Erwerbsverläufe“
Regensburger Straße 104
90478 Nürnberg
carina.toussaint@iab.de

Carina Toussaint, geb. 1990 in Erlangen. Studium der Sozialökonomik an der Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. Seit 2019 wissenschaftliche Mitarbeiterin am Institut für Arbeitsmarkt- und Berufsforschung (IAB) und seit 2021 Stipendiatin des gemeinsamen Graduiertenprogramms des IAB und des Fachbereichs Wirtschaftswissenschaften der FAU.
Forschungsschwerpunkte: Bildungsintentionen und -entscheidungen; Übergang von der Schule in den Arbeitsmarkt; soziale Ungleichheit im Zugang zu (höherer) Bildung.

Appendix

A1. Data overview

The empirical analyses employ data for individual preferences, aspirations, achievement levels, and major choices from the data of the German National Educational Panel Study. The NEPS data are a longitudinal survey that was collected from 2008 to 2013 as part of the Framework program for the promotion of empirical educational research, which was funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, the Leibniz Institute for Educational Trajectories (LIfBi) has been responsible for the continuation of the NEPS (Blossfeld et al. 2011).

While Ochsensfeld (2016) uses NEPS-SC5 as the central individual data set, which is a representative sample of first-year students at higher education institutions, we refer to the data set of the NEPS-SC4 cohort. The first wave of this cohort starts with ninth-grade school students in fall/winter 2010. The key advantage of these data is that they contain information about preferences, interests, and aspirations prior to choosing a major. As the observation period is now long enough and the transition to higher education can be observed, it is thus possible to use these data for the analyses.

To make the results comparable, we restrict the sample in the same way as the study by Ochsensfeld (2016). We therefore exclude individuals older than 25 years at the start of the study program. In addition, only individuals with a standard leaving certificate ("Abitur") are considered as they have the broad choice set of all study programs at full universities and universities of applied sciences.

A2. Operationalization of the constructs used

Differences in the math intensity of college majors may be a factor in gender segregation across majors. To operationalize the math intensity of majors, we used data from the NEPS-SC5, in which students at the end of their first semester (third wave) were asked to rate the math intensity in their studies on a 4-point Likert scale. The NEPS-SC5 is a representative longitudinal survey of students that starts at the beginning of the first semester (Blossfeld et al. 2011). For the analyses, we form an index for math intensity that ranges from 0 to 100. In addition, we form a dummy variable that takes the value of 100 for majors attributed to the care sector (i.e., medical science, social work, pedagogy, psychology, and teacher).

To represent possible gender differences in the social approval of individuals' choice of majors, we use the ratings of two statements, "1. My parents (2. friends) think that I chose a good major", in the first wave of the NEPS-SC5. We calculate the proportion of women and men who fully agree or somewhat agree with these statements in each major. For the analyses, we use the difference in these approval rates for women and men in college majors.

The majors differ according to their future earning potential and the compatibility of work and family. Since these structural characteristics can also influence the selection of females and males into majors, we consider the average wage level and the average hours worked per week by individuals who graduated from one of the 23 college majors. We form the variables from the pooled data of the Micro Census of the years 2011, 2012, 2013. The Micro Census is a legally mandatory representative survey of households in Germany in which 1 percent of the population participates each year (Lüttinger & Riede 1997). The wage level is estimated by a linear regression on the wages of individuals, including dummy variables for the fields of study and relevant controls.

To map the different anticipated forms of discrimination in the labor market after graduation from certain fields of study, we use data from the HIS graduate panel studies of

1997 and 2001 (Fabian & Minks 2006; Schramm & Beck 2010). These 2- and 3-wave panel surveys include German graduates of higher education institutions. We use the proportion of women who agree with the statement "the right gender is (very) important in order to be successful when looking for a job". This allows us to map anticipated discrimination by major.

Table A2. Operationalization of variables and data source in the original study and replication

<i>Individual characteristics</i>		
	Original Study:	Differences/equivalences in Replication study:
Major Choice (dependent variable)	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Question: What subjects have you enrolled in? Please specify your first subject. - 23 fields of study 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 8 (2014/2015)/wave 9 (2015/2016)/wave 10 (2016/2017) - Same variable, same fields of study
Female	<ul style="list-style-type: none"> - Data set: NEPS-SC5, NEPS-SC5, wave 1 (2010/2011) - Based on variable "gender" - 0/1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 1 (2010) - Same variable
Realistic	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: building or assembling things, preparing smth. according to a plan or to a sketch, working with metal/wood or creating things of metal or wood - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited
Investigative	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: observing and analyzing things in detail, conducting experiments in a test laboratory, analyzing things through a microscope - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited
Artistic	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: drawing pictures, designing smth. artistically, reading or interpreting poetry/literature - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited
Social	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: championing the needs of others, helping sick people, caring for children or adults in need - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited
Enterprising	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: negotiating with other people, appearing in public to back smth., telling other people what they should do - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited

Tab. A2 (continued)

<i>Individual characteristics</i>		
	Original Study:	Differences/equivalences in Replication study:
Conventional	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Composite index based on 3 items of RIASEC with 6-point Likert scale - Items: keeping lists or records of things, counting and sorting things, monitoring adherence to regulations - Range of index: 0–1 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 7 (2013/2014) - Same index composited
Relative math grade	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Subtraction of German grade from math grade - Questions: How many points did you have in German in your final school semester? How many points did you have in mathematics in your final school semester? (points: 0–15) - Range: from –15 to 15 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, last reported German and math grades during school time - Same question, same measure
Earning well	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 3 (2012) - Based on the item, “In the following, we deal with things that might be important to a vocational activity. How important do you personally find these things, independent of your current situation? → how important is good remuneration to you?” - Scale 1 (=very unimportant) to 6 (=very important) 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 3 (2011/2012) - Same item
Pleasant working hours	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 3 (2012) - Based on the item, “In the following, we deal with things that might be important to a vocational activity. How important do you personally find these things, independent of your current situation? → how important are pleasant working hours to you?” - Scale 1 (=very unimportant) to 6 (=very important) 	<ul style="list-style-type: none"> - Data set changed: NEPS-SC4, wave 3 (2011/2012) - Same item
<i>Major characteristics</i>		
Math intensity	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 2 (2011) - Question: To what extent are knowledge and skills in the following areas needed in your studies? – mathematics”, 4-point Likert scale - Aggregated for each major - Range: from 0 to 100 	<ul style="list-style-type: none"> - Identical to original study
Care	<ul style="list-style-type: none"> - own coding - Value 100 for the following fields of study: social work, pedagogy, teacher, psychology, medical science - Value 0 for all other fields of study 	<ul style="list-style-type: none"> - Identical to original study
Difference parents’ approval	<ul style="list-style-type: none"> - Data set: NEPS-SC5, wave 1 (2010/2011) - Proportion of male respondents who responded to the 5-point Likert scale item, “What do your parents and friends think about the fact that you are studying for a degree and about the subject you are studying?” - My parents think I have chosen a good subject with either “applies completely” or “applies”, subtracted from the proportion of female respondents who responded to the same item with either “applies completely” or “applies” 	<ul style="list-style-type: none"> - Identical to original study

Tab. A2 (continued)

<i>Individual characteristics</i>		
	Original Study:	Differences/equivalences in Replication study:
Difference friends' approval	<ul style="list-style-type: none"> - Date set: NEPS-SC5, wave 1 (2010/2011) - Proportion of male respondents who responded to the 5-point Likert scale item, "What do your parents and friends think about the fact that you are studying for a degree and about the subject you are studying?" - My friends think I have chosen a good subject with either "applies completely" or "applies", subtracted from the proportion of female respondents who responded to the same item with either "applies completely" or "applies" 	<ul style="list-style-type: none"> - Identical to original study
Wage level	<ul style="list-style-type: none"> - Data set: German Micro Census 2007, 2008, 2009 (pooled) - Coefficients for major dummy variables from a regression of hourly wages on these dummies and a set of productivity-related characteristics - Productivity-related characteristics: college major, region, experience, age, type of abitur, type of university, PhD, cohort - Scaled to measure the difference in the wage level with economics as reference category, in percentage - Regression sample restricted to individuals with completed college education, age 25-55, working a minimum of 20 h/week 	<ul style="list-style-type: none"> - Data set changed: German Micro Census 2011, 2012, 2013 (pooled) - Same variables, same regression model, same sample restrictions
Overwork norms	<ul style="list-style-type: none"> - Data set: German Micro Census 2007, 2008, 2009 (pooled) - Hours worked per week on average by individuals working at least 35 h/week 	<ul style="list-style-type: none"> - Data set changed: German Micro Census 2011, 2012, 2013 (pooled) - Same variable
Discrimination	<ul style="list-style-type: none"> - HIS graduate panel studies 1997, 2001 (pooled) - Proportion of females who responded to the item, "Which of the following criteria, in your opinion, are important in order to be successful when looking for a job?" - the "right" gender with "very important" or "important" 	<ul style="list-style-type: none"> - Identical to original study

A3. Tables with full regression results

Table A3.1: Results of logit model (basis for propensity score matching and inverse probability weighting)

variable	coefficient	standard error
female	-0.47	0.12
migration background		
no mig. background	ref.	
<i>mig. background</i>	0.25	0.14
<i>missing information</i>	-0.12	0.26
socioec. status		
<i>low</i>	ref.	
<i>high</i>	0.36	0.12
<i>missing information</i>	0.10	0.15
RIASEC – realistic	-1.25	0.25
RIASEC – investigative	0.79	0.24
RIASEC – artistic	0.97	0.22
RIASEC – social	-0.71	0.25
RIASEC – enterprising	0.57	0.28
RIASEC – conventional	-0.33	0.26
job value – earning well	-0.06	0.29
job value – pleasant working hrs	-0.04	0.23
math grade (in points)	0.13	0.02
German grade (in points)	0.06	0.04
constant	-0.17	0.39

N=2,700; Data: NEPS-SC4.

Data basis of the logit models is a sample of N=2,700 that includes all individuals with a standard university entrance qualification from whom we have information on the required variables. Among them, 1,936 individuals start their studies and subsequently form the sample for the main models of our analyses (conditional logit models).

Table A3.2: Results for different p-scores, explanatory power of variables for sex segregation in fields of study, decomposition results from a mediation analysis on the basis of conditional logit

	pscore <= 25 % percentile		pscore 25 % – 50 % percentile		pscore 50 % – 75 % percentile		pscore > 75 % percentile	
	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹	OR for fem x prop.fem	Perc. explained ¹
baseline model	1.0505	0	1.0485	0	1.0370	0	1.0364	0
full model (all variables)	1.0393	22.2	1.0262	46.0	1.0251	32.2	1.0055	84.9
<i>Preferences</i>								
voc. interests x (math intensity, care)	1.0295	41.6	1.0313	35.4	1.0200	46.1	1.0199	45.4
earning well x wage level	1.0503	0.4	1.0485	0.1	1.0366	1.2	1.0361	1.0
<i>Constraints</i>								
pleasant work hrs. x overwork norms	1.0505	0.0	1.0485	-0.0	1.0370	0.0	1.0365	-0.1
fem. x diff. parents' approval	1.0475	5.9	1.0483	0.4	1.0374	-1.1	1.0361	0.9
fem. x diff. friends' approval	1.0575	-13.8	1.0503	-3.7	1.0396	-7.1	1.0274	24.8
rel. math grade x math int.	1.0501	0.9	1.0472	2.8	1.0343	7.4	1.0333	8.5
fem. x discrimination	1.0532	-5.3	1.0479	1.3	1.0404	-9.1	1.0388	-6.6
N	484		484		484		484	

Note: ¹ percentage explained when adding term to baseline model. Figure 3 based on these results.

Data: NEPS-SC4, NEPS-SC5, Micro Census, HIS graduate panel studies.