IAAEU
Institute for Labour Law and Industrial Relations in the European Union

School-starting age and educational mismatch

Theresa Geißler & Sven Hartmann

August 28, 2024

geissler@iaaeu.de
Theresa Geißler
geissler_tr
https://sites.google.com/view/theresageissler

Motivation

Motivation or when should I send my kid to school?


ABOUT EDUCATION

The Fraternities Show Signs of New Strength

Starting School Later

EDUCATION

College Credit for High School Students

CHINESE

The SPANISH INSTITUTE

pratt

Art & Technology

Figure 1: The New York Times, May 21, 1985

Don't send children to school at four, warn experts

Figure 2: The Guardian, February 14, 2009

How can I tell if my child is ready to start school next year?

Figure 3: The Guardian, August 29, 2023

Guidance

Summer born children starting school: advice for parents

Updated 27 April 2023

Figure 4: UK Department for Education, April 27, 2023

OPINION COMMENTARY Follow

Toddlers Don't Have to Go to School

Parents are resisting the pressure to enroll their kids earlier.

Figure 5: The Wall Street Journal, August 5, 2019

Flexible school starting age a step closer in Northern Ireland

Figure 6: BBC, January 20, 2022

Call to raise school start age in Scotland to six

Figure 7: BBC, July 31, 2022

Previous evidence & Contribution

Previous evidence & Contribution

- Extensive research with mixed findings on the relationship between school- (kindergarten-) starting age and educational attainment, cognitive development, and academic performance (Attar and Cohen-Zada, 2018; Balestra et al., 2020; Beatton et al., 2023; Bedard and Dhuey, 2006; Black et al., 2011; Chen and Park, 2021; Cook and Kang, 2016; Cornelissen and Dustmann, 2019; Dhuey et al., 2019; Fletcher and Kim, 2016; Görlitz et al., 2022; Herbst and Strawiński, 2016; Lubotsky and Kaestner, 2016; McEwan and Shapiro, 2008; Mühlenweg and Puhani, 2010; Mühlenweg et al., 2012; Oosterbeek et al., 2021; Peña, 2017; Ponzo and Scoppa, 2014; Puhani and Weber, 2008; Qihui, 2022; Sprietsma, 2010; Tan, 2017; Yamaguchi et al., 2023)
- Limited research on labour market outcomes
  - Mostly related to earnings (Matta et al., 2016; Black et al., 2011; Oosterbeek et al., 2021; Peña, 2017; Fredriksson and Öckert, 2014; Larsen and Solli, 2017)
  - No empirical evidence on the impact of school-starting age on educational mismatch
  - One related study showing no relationship between *relative age* and being matched (Fumarco et al., 2022)

This study

- first evidence on the impact of school-starting age on educational mismatch
- distinguishes over- and undereducation at the extensive and intensive margin

Data & Empirical approach

Data & Dependent variable

- Data from the Socio-Economic Panel (v37) (DIW, 2022)
  - Longitudinal household study on individuals living in Germany
  - Carried out annually since 1984
  - Contains detailed information on individuals life in areas such as education, work and family relations

- Educational mismatch is assessed based on the statistical measure (MEAN) (Verdugo and Verdugo, 1989; Clogg and Shockey, 1984)
  - Required education: average education $\overline{x_{ot}}$ within a reference group o at time t
  - Attained education: years of education x_{it} of individual i at time t
  - Extensive margin:

$$OE_{ito} = 1 \text{ if } x_{it} > \overline{x_{ot}} + \sigma_{ot} \tag{1}$$

$$UE_{ito} = 1 \text{ if } x_{it} < \overline{x_{ot}} - \sigma_{ot} \tag{2}$$

  - Intensive margin:

$$Years_{ito} = \begin{cases} x_{it} - \overline{x_{ot}} - \sigma_{ot} \text{ if } OE_{ito} = 1 \\ x_{it} - \overline{x_{ot}} + \sigma_{ot} \text{ if } UE_{ito} = 1 \end{cases} \tag{3}$$

Identification strategy i

First option: School-entry laws with sharp RDD (Cook and Kang, 2016; Dobkin and Ferreira, 2010)

- In Germany, all children who turn six until the predefined cut-offs should start school in the same year; all those turning six afterward start one year later
- Variation in school-entry cut-off dates on the federal-state level across years Cut-off
- Running variable: $dist_{isy} = birthmonth_i - cutmonth_{sy}$
- Treatment variable: $older_{isy} = \begin{cases} 0 & \text{if } dist_{isy} \leq 0 \\ 1 & \text{if } dist_{isy} > 0 \end{cases}$
- Causal effect of being treated would be identified by:

$$EM_{ito} = \beta_0 + \beta_1 older_{isy} + \beta_2 dist_{isy} + \beta_3 X_i' + \lambda_t + \gamma_b + \delta_s + \epsilon_{it} \tag{4}$$

  - if there was no manipulation around the cut-off (Huang et al., 2020; Kim, 2021) McCrary
  - if there was perfect compliance with the cut-off Compliance

→ Alternative: School-starting age in months (Bahrs and Schumann, 2020)

Preferred option: School-starting age in months (fuzzy RDD) (Mühlenweg and Puhani, 2010)

- SSA **only** provided for individuals who have taken part in the survey when they were still in school (or other education)
- Imputed values for school-starting age in months using information on
  - school-leaving degree, year degree, birth year and month and federal state (Bahrs and Schumann, 2020) and
  - the day of school start (FDR: August 1st, former GDR: September 1st)

Histogram

- Estimating the Local Average Treatment Effect (LATE) of school-starting age in months on educational mismatch (intensive and extensive margin) using the *provided* and *imputed* sample

$$months_i = \alpha_0 + \alpha_1 older_{isy} + \alpha_2 dist_{isy} + \alpha_3 X'_{it} + \lambda_t + \gamma_b + \delta_s + \epsilon_{it} \quad (5)$$

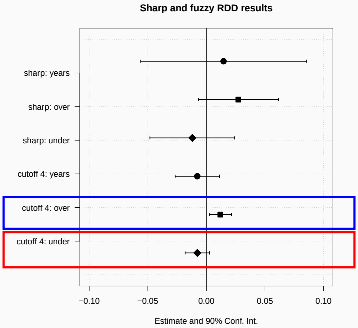
$$EM_{ito} = \beta_0 + \beta_1 \widehat{months}_i + \beta_2 dist_{isy} + \beta_3 X'_{it} + \lambda_t + \gamma_b + \delta_s + \epsilon_{it} \quad (6)$$

X'_{it}: migration background, number of siblings, sex, school education of father and mother; λ_t, γ_b, δ_s: survey year, birth year, and state fixed effects

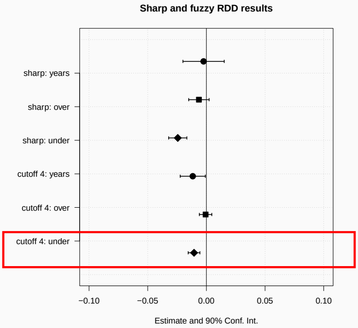
Results

Results

Figure 8: Main results



(a) Provided sample



(b) Imputed sample

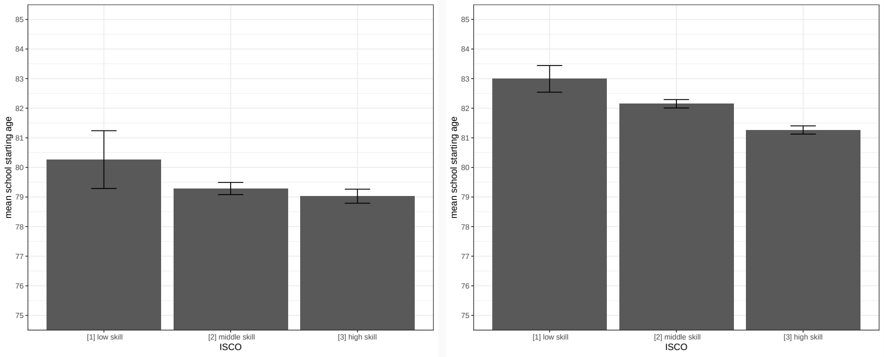
Note: Figures 8a and 8b display results from sharp and fuzzy RDD estimations for the years of educational mismatch (years) as well as the likelihood of overeducation (over) and undereducation (under). The upper three panels present results from sharp RDD estimations, while the lower three stem from fuzzy RDD estimations using $\pm$ 4 months range around the cut-off. The samples are based on the years 1996 to 2020 in the sharp RDD and the years 2001 to 2020 in the fuzzy RDD in the provided sample, while the estimates using the imputed sample are based on the survey years 1991 to 2020.

Sample, Descriptives & Difference in Means
Robustness & Heterogeneity

Channel

Channel: Occupational Choice i

Figure 9: Occupations and school-starting age



(a) Provided sample **(b)** Imputed sample

Note: 9a and 9b display the average school-starting age in months by ISCO skill group. The skill groups cover ISCO-group 9 in "[1] low skill", ISCO-groups 4 to 8 in "[2] middle skill" and ISCO-groups 1 to 3 in "[3] high skill". 90% confidence intervals plotted.

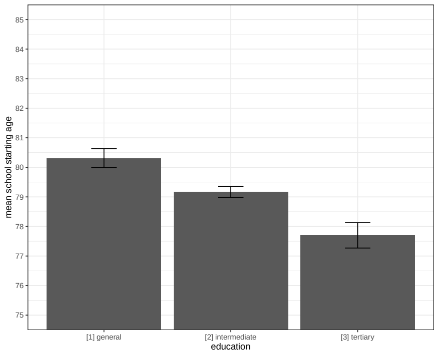
Channel: Occupational Choice ii

Table 1: Channel: Occupational choice

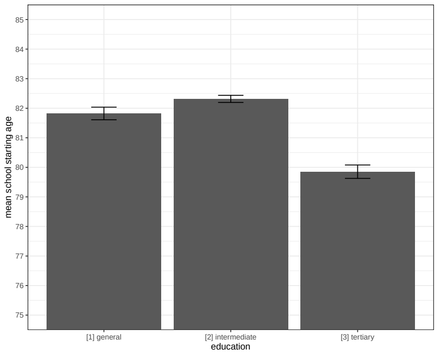
	(1)	(2)	(3)	(4)	(5)
outcome	high skill	overeducation		undereducation	
Provided sample:					
$\hat{months}$	−0.003	0.012*	0.012*	−0.008	−0.007
	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
dist	0.006	−0.010*	−0.010*	0.014**	0.013**
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
high skill			−0.064***		0.127***
			(0.014)		(0.015)
Num.Obs.	2831	2831	2831	2831	2831
Imputed sample:					
$\hat{months}$	−0.019***	−0.001	0.000	−0.011***	−0.009**
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
dist	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
high skill			0.023***		0.082***
			(0.005)		(0.005)
Num.Obs.	36 838	36 838	36 838	36 838	36 838

The estimations are based on data from the SOEP for the years 1996 to 2020 in the provided sample and for the years 1991 to 2020 in the imputed sample. Columns (1) to (5) contain results from fuzzy regression discontinuity estimations using the variable *older* to instrument school starting age in months. Columns (1) to (5) use linear probability models. Column (1) estimates the impact of school starting age on the likelihood of working in a high-skill occupation (ISCO groups 1 to 3), while columns (2) and (3) ((4) to (5)) estimate the impact of school starting age on overeducation (undereducation). Columns (2) and (4) present the baseline specification discussed in 7. Columns (3) and (5) contain the mediation analysis, adding working in a high-skill occupation as an additional covariate. Controls include dummies for having a direct or indirect migration background, the number of siblings, being female, and the level of education obtained by father and mother. Survey year, birth month, and federal state fixed effects are included. Heteroskedasticity-robust standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 10: Education and school-starting age



(a) Provided sample



(b) Imputed sample

Note: Figures 10a and 10b display the average school-starting age in months by educational degree. The educational degrees cover "[1] general", "[2] intermediate" and "[3] tertiary". 90% confidence intervals plotted.

Suppressor: Educational attainment ii

Table 2: Channel: Educational attainment

	(1)	(2)	(3)	(4)	(5)
outcome	tertiary	overeducation		undereducation	
Provided sample:					
$\hat{months}$	−0.013*	0.012*	0.017**	−0.008	−0.009
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
dist	0.011**	−0.010*	−0.015**	0.014**	0.015**
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
tertiary			0.405***		−0.075***
			(0.030)		(0.019)
Num.Obs.	2831	2831	2831	2831	2831
Imputed sample:					
$\hat{months}$	−0.017***	−0.001	0.006+	−0.011***	−0.012***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
dist	0.002+	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
tertiary			0.378***		−0.114***
			(0.010)		(0.009)
Num.Obs.	36 838	36 838	36 838	36 838	36 838

The estimations are based on data from the SOEP for the years 1996 to 2020 in the provided sample and for the years 1991 to 2020 in the imputed sample. Columns (1) to (5) contain results from fuzzy regression discontinuity estimations using the variable *older* to instrument school starting age in months. Columns (1) to (5) use linear probability models. Column (1) estimates the impact of school starting age on the likelihood of having tertiary education, while columns (2) and (3) ((4) to (5)) estimate the impact of school starting age on overeducation (undereducation). Columns (2) and (4) present the baseline specification discussed in 7. Columns (3) and (5) contain the mediation analysis adding tertiary education as an additional covariate. Controls include dummies for having a direct or indirect migration background, the number of siblings, being female, and the level of education obtained by father and mother. Survey year, birth month, and federal state fixed effects are included. Heteroskedasticity-robust standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Summary

Summary

- First evidence on the link between school-starting age and educational mismatch at the intensive and extensive margin
- Results reveal evidence for
  1. a reduction in the likelihood of undereducation (quantitative strength similar in both samples, precisely estimated in the larger sample)
     - consistently driven by open and agreeable individuals
     - partially mediated by a lower likelihood of selecting into high-skilled jobs
     - suppressed by the likelihood of owning a tertiary degree
  2. an increase in the likelihood of overeducation in the smaller sample
     - unaffected by occupational choice
     - suppressed by the likelihood of owning a tertiary degree
       $\rightarrow$ inclusion leading to a positive significant effect in both samples
  3. a negative relationship between school-starting age and the likelihood of working in a high-skill occupation and owning a tertiary degree (Tan, 2017)

School-starting age and educational mismatch

Theresa Geißler & Sven Hartmann

August 28, 2024

geissler@iaaeu.de
Theresa Geißler
geissler_tr
https://sites.google.com/view/theresageissler

Attar, I. and Cohen-Zada, D. (2018). The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. *Journal of Economic Behavior & Organization*, 153:38–57.

Bahrs, M. and Schumann, M. (2020). Unlucky to be young? the long-term effects of school starting age on smoking behavior and health. *Journal of Population Economics*, 33(2):555–600.

Balestra, S., Eugster, B., and Liebert, H. (2020). Summer-born struggle: The effect of school starting age on health, education, and work. *Health economics*, 29(5):591–607.

Beatton, T., Kidd, M. P., Niu, A., and Vella, F. (2023). Age of starting school, academic performance, and the impact of non-compliance: An experiment within an experiment, evidence from australia. *Economic Record*, 99(325):175–206.

Bedard, K. and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4):1437–1472.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? the effects of school starting age. *The review of economics and statistics*, 93(2):455–467.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Chen, J. and Park, A. (2021). School entry age and educational attainment in developing countries: Evidence from china's compulsory education law. *Journal of Comparative Economics*, 49(3):715–732.

Clogg, C. C. and Shockey, J. W. (1984). Mismatch between occupation and schooling: A prevalence measure, recent trends and demographic analysis. *Demography*, 21(2):235–257.

Cook, P. J. and Kang, S. (2016). Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics*, 8(1):33–57.

Cornelissen, T. and Dustmann, C. (2019). Early school exposure, test scores, and noncognitive outcomes. *American Economic Journal: Economic Policy*, 11(2):35–63.

Dhuey, E., Figlio, D., Karbownik, K., and Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management*, 38(3):538–578.

DIW (2022). Sozio-oekonomisches Panel (SOEP)-Version 37, Daten der Jahre 1984-2020, (SOEP-Core, v37, EU Edition). DIW-Deutsches Institut für Wirtschaftsforschung, Berlin. DOI: 10.5684/soep.core.v37eu [Titel anhand dieser DOI in Citavi-Projekt Ã¼bernehmen].

Dobkin, C. and Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of education review*, 29(1):40–54.

Fletcher, J. and Kim, T. (2016). The effects of changes in kindergarten entry age policies on educational achievement. *Economics of education review*, 50:45–62.

Fredriksson, P. and Öckert, B. (2014). Life-cycle effects of age at school start. *The Economic Journal*, 124(579):977–1004.

Fumarco, L., Vandromme, A., Halewyck, L., Moens, E., and Baert, S. (2022). Does relative age affect speed and quality of transition from school to work?

Görlitz, K., Penny, M., and Tamm, M. (2022). The long-term effect of age at school entry on cognitive competencies in adulthood. *Journal of Economic Behavior & Organization*, 194:91–104.

Herbst, M. and Strawiński, P. (2016). Early effects of an early start: Evidence from lowering the school starting age in poland. *Journal of Policy Modeling*, 38(2):256–271.

Huang, C., Zhang, S., and Zhao, Q. (2020). The early bird catches the worm? school entry cutoff and the timing of births. *Journal of Development Economics*, 143:102386.

Kim, T. (2021). Age culture, school-entry cutoff, and the choices of birth month and school-entry timing in south korea. *Journal of Demographic Economics*, 87(1):33–65.

Larsen, E. R. and Solli, I. F. (2017). Born to run behind? persisting birth month effects on earnings. *Labour Economics*, 46:200–210.

Lubotsky, D. and Kaestner, R. (2016). Do skills beget skills'? evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood. *Economics of Education Review*, 53:194–206.

Matta, R., Ribas, R. P., Sampaio, B., and Sampaio, G. R. (2016). The effect of age at school entry on college admission and earnings: a regression-discontinuity approach. *IZA Journal of Labor Economics*, 5:1–25.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

McEwan, P. J. and Shapiro, J. S. (2008). The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates. *Journal of human Resources*, 43(1):1–29.

Mühlenweg, A., Blomeyer, D., Stichnoth, H., and Laucht, M. (2012). Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data. *Economics of Education Review*, 31(3):68–76.

Mühlenweg, A. M. and Puhani, P. A. (2010). The evolution of the school-entry age effect in a school tracking system. *Journal of human Resources*, 45(2):407–438.

Oosterbeek, H., ter Meulen, S., and van Der Klaauw, B. (2021). Long-term effects of school-starting-age rules. *Economics of Education Review*, 84:102144.

Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56:152–176.

Ponzo, M. and Scoppa, V. (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling*, 36(3):578–599.

Puhani, P. and Weber, A. (2008). Does the early bird catch the worm? In Dustmann, C., Fitzenberger, B., and Machin, S., editors, *The Economics of Education and Training*. Physica-Verlaf HD.

Qihui, C. (2022). Impacts of delayed school entry on child learning in rural northwestern china - forced delay versus voluntary delay. *Applied Economics*, 54(21):2453–2472.

Sprietsma, M. (2010). Effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using pisa 2003. *Education Economics*, 18(1):1–32.

Tan, P. L. (2017). The impact of school entry laws on female education and teenage fertility. *Journal of Population Economics*, 30(2):503–536.

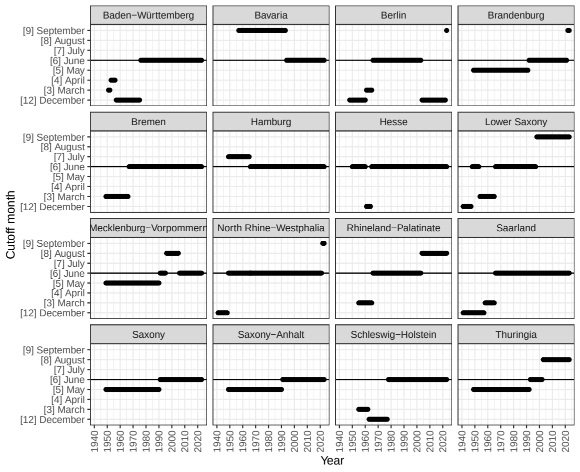
Verdugo, R. R. and Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, 24(4):629–643.

Yamaguchi, S., Ito, H., and Nakamuro, M. (2023). Month-of-birth effects on skills and skill formation. *Labour Economics*, 84:102392.

Appendix

Cut-off dates Back

Figure 11: Cutoff dates by federal state

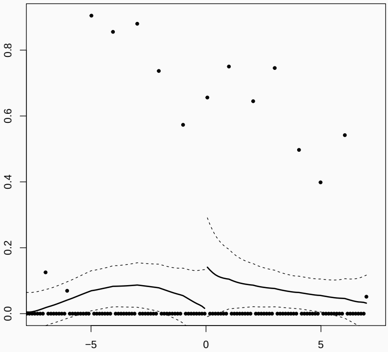


Note: Figure ?? reports the cutoff dates by year and federal state. Data on the cutoff dates was collected by going through the respective legislation such as the BayEUG Art. 37 for Bavaria and the SchulG §42 for Berlin.

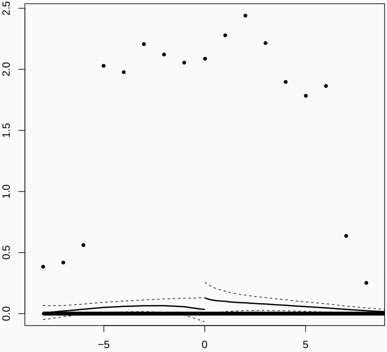
McCrary Test

Back

Figure 12: McCrary manipulation test



(a) Provided sample



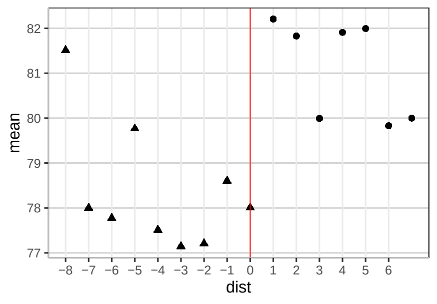
(b) Imputed sample

Note: Figures 12a and 12b provide results from the manipulation test for the running variable *dist* as proposed by McCrary (2008). In Figure 12a, the estimated bandwidth is 3.8, the log difference in heights accounts for 2.427, and the p-value is 0.000. In Figure 12b, the estimated bandwidth is 5.018, the log difference in heights equals 1.321, and the p-value is 0.000.

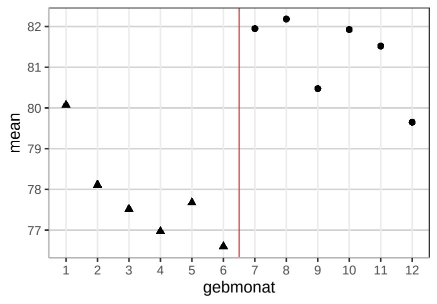
Compliance with cut-off

Back

Figure 13: Discontinuity in school-starting age



(a) By distance to cutoff



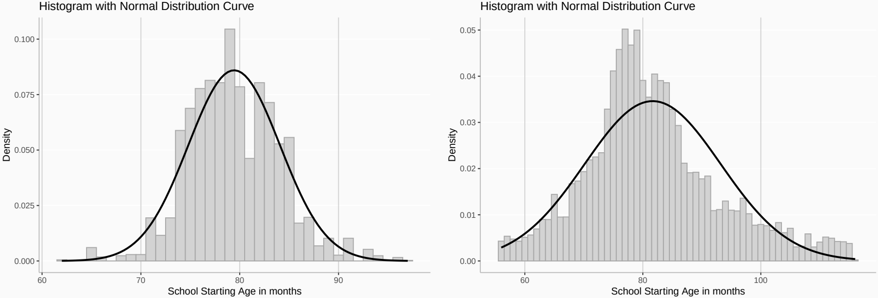
(b) By birth month

Note: Figures 13a and 13b display the discontinuity in the school-starting age variable by distance to the cutoff and by birth month for the provided sample. Plots for the imputed sample are available from the authors.

Histogram and normal distribution of school-starting age variables

Back

Figure 14: School-starting age



(a) Provided sample

(b) Imputed sample

Note: Figures 14a and 14b display the distribution of the school-starting age variable in the provided and the imputed sample, respectively. The sample is based on the survey years 1996 to 2020 in Figure 14a and 1991 to 2020 in Figure 14b.

Sample & Descriptives Back i

Sample restrictions:

- Individuals in paid employment
- Younger than their legal retirement age (65 to 67, see, §235 German Social Security Code)
- Non-missings in all variables

Provided sample

- 3808 (2831) observations
- 721 (530) individuals
- survey waves 1996-2020 (2001-2020)

Imputed sample

- 53562 (36838) observations
- 8954 (6227) individuals
- survey waves 1991-2020

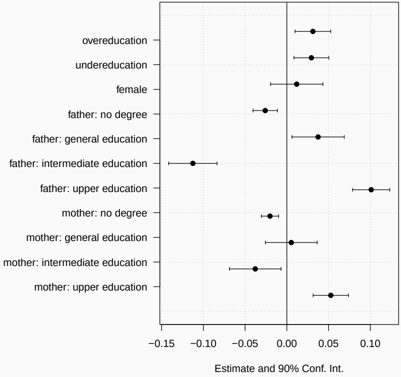
Sample & Descriptives Back ii

Table 3: Summary statistics

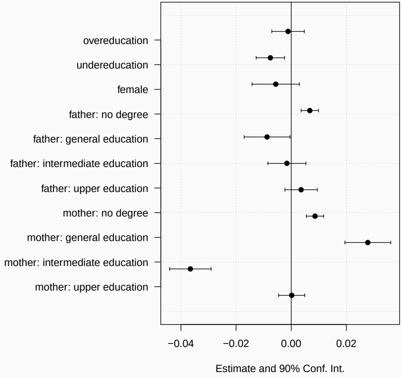
	Provided sample				Imputed sample			
	Mean	SD	Min	Max	Mean	SD	Min	Max
schoolstartingage	79.21	4.98	62.00	97.00	81.79	11.37	56.00	116.00
older	0.42	0.49	0.00	1.00	0.46	0.50	0.00	1.00
dist	-0.29	2.59	-4.00	4.00	0.03	2.54	-4.00	4.00
years of mismatch	0.03	0.63	-5.94	6.34	0.07	0.69	-5.94	6.34
overeducation	0.13	0.34	0.00	1.00	0.14	0.34	0.00	1.00
undereducation	0.13	0.33	0.00	1.00	0.10	0.30	0.00	1.00
number siblings	1.36	1.19	0.00	9.00	1.51	1.35	0.00	14.00
female	0.51	0.50	0.00	1.00	0.52	0.50	0.00	1.00
birth year	1984.76	4.90	1977.00	1998.00	1970.28	11.01	1939.00	1998.00
direct	0.01	0.08	0.00	1.00	0.03	0.17	0.00	1.00
indirect	0.14	0.34	0.00	1.00	0.05	0.21	0.00	1.00
native	0.86	0.35	0.00	1.00	0.92	0.27	0.00	1.00
father rupper	0.15	0.36	0.00	1.00	0.13	0.34	0.00	1.00
father intermediate	0.32	0.47	0.00	1.00	0.20	0.40	0.00	1.00
father general	0.47	0.50	0.00	1.00	0.63	0.48	0.00	1.00
father no	0.06	0.23	0.00	1.00	0.03	0.18	0.00	1.00
mother upper	0.13	0.34	0.00	1.00	0.08	0.28	0.00	1.00
mother intermediate	0.41	0.49	0.00	1.00	0.26	0.44	0.00	1.00
mother general	0.43	0.49	0.00	1.00	0.62	0.48	0.00	1.00
mother no	0.03	0.16	0.00	1.00	0.03	0.18	0.00	1.00
Num.Obs.	2831				36838			

The samples are based on those individuals whose birth month is in a $\pm$ 4-month range around the respective cutoff date, restricting the survey waves to 2001 to 2020 for the provided sample and 1991 to 2020 for the imputed sample. SOEP weights applied.

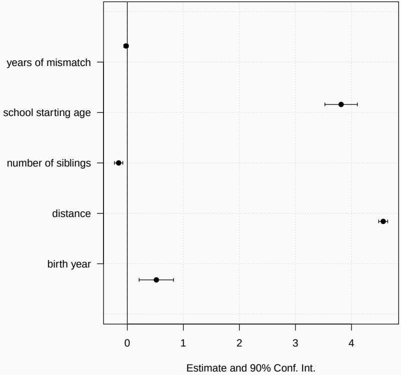
Figure 15: Difference in means between those born before and after the cut-off



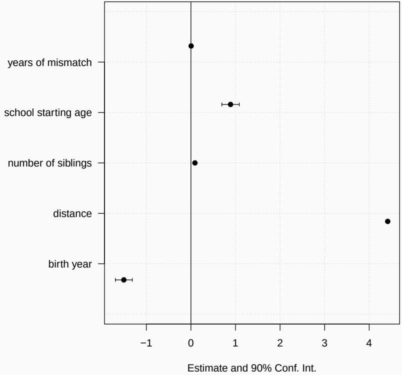
(a) Provided sample: dichotomous



(b) Imputed sample: dichotomous



(c) Provided sample: continuous/count



(d) Imputed sample: continuous/count

Note: Figures 15a and 15b display the differences in means of the dichotomous variables in the provided and the imputed sample, while Figures 15c and 15d show the differences in the means of the continuous or count variables. The samples are based on those individuals whose birth month is in a ± 4-range around the respective cut-off date, restricting the survey years 1996 to 2020 in Figures 15a and 15c and 1991 to 2020 in Figures 15b and 15d. 90% confidence intervals plotted.

Robustness & Heterogeneity

Back

Robustness

Tables

- RDD design
  - Second-order polynomial ✓
  - Data-driven cut-off ✓(Calonico et al., 2020)
  - Separate linear trends (✓)
- Assessment of EM
  - Indirect self-assessment ✓
  - Job analyst (✓)
- Sample composition
  - Outlier (upper/lower 2.5 %) ✓
  - Cross section
- Set of covariates (INKAR)
  - Local unemployment ✓
  - Unemployed by qualification ✓

Heterogeneity

- Demographics Graphs
  - Sex
  - Birth cohort
  - Migration background
  - Region
- Personality Graphs
  - Risk-aversion
  - Openness to experience
  - Agreeableness
  - Extraversion
  - Conscientiousness
  - Neuroticism

✓Robust: + OE in provided sample, and - UE in imputed sample
(✓) Only in one sample

Robustness: Tables Back i

Table 4: Robustness: Provided sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	quadratic		cutoff 3		separate trends		isa		ja		outlier		first		unempl. rate		unempl. qual.	
	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE
$\widehat{months}$	0.011+	−0.006	0.012*	−0.006	0.005	−0.002	0.017**	−0.018**	0.009	−0.009	0.014*	−0.006	−0.002	−0.007	0.013*	−0.008	0.014*	−0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.006)	(0.011)	(0.017)	(0.006)	(0.006)	(0.006)	(0.006)
dist			−0.010+	0.013*	0.001	0.007***	−0.003	0.014**	−0.011*	0.010+	−0.015**	0.012*	−0.002	0.023+	−0.011*	0.014**	−0.013*	0.010+
			(0.006)	(0.006)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.009)	(0.012)	(0.005)	(0.005)	(0.006)	(0.005)
poly(dist, 2)1	−1.325*	1.794**																
	(0.636)	(0.637)																
poly(dist, 2)2	0.227	−0.403																
	(0.331)	(0.346)																
unemployment rate															−0.013*	0.002		
															(0.005)	(0.004)		
share specialist																	0.057	−0.080*
																	(0.049)	(0.037)
share trained																	0.012	0.003
																	(0.008)	(0.006)
share experts																	0.007	0.045+
																	(0.031)	(0.023)
First stage:																		
older	4.667***		4.937***		5.637***		4.386***		4.386***		4.347***		4.102***		4.388***		5.497***	
	(0.393)		(0.444)		(0.313)		(0.390)		(0.390)		(0.360)		(0.997)		(0.390)		(0.642)	
F-stat. (1st stage)	156.7		134.3		131.4		148.5		148.5		225.8		20.6		148.2		152.3	
Num.Obs.	2831	2831	2227	2227	3808	3808	2831	2831	2831	2831	2641	2641	530	530	2831	2831	1800	1800

Komplett anpassen Heteroskedasticity-robust standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Robustness: Imputed sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	quadratic		cutoff2		separate trends		isa		ja		outlier		first		unempl. rate		unempl. qual.	
	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE	OE	UE
$\widehat{months}$	0.000	−0.012***	0.015+	−0.013+	−0.002	−0.008***	0.005	−0.012***	0.012***	−0.011**	−0.003	−0.004*	0.001	−0.008	−0.001	−0.012***	0.009*	−0.012***
	(0.003)	(0.003)	(0.008)	(0.007)	(0.002)	(0.002)	(0.004)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.007)	(0.007)	(0.003)	(0.003)	(0.004)	(0.004)
dist			−0.010*	0.001	0.001*	−0.002***	0.004***	0.000	0.001	0.001	0.001	0.001	0.000	0.003+	0.000	0.000	0.000	0.002
			(0.005)	(0.004)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
poly(dist, 2)1	0.109	0.105																
	(0.374)	(0.365)																
poly(dist, 2)2	−0.368	0.535																
	(0.345)	(0.329)																
unemployment rate															0.001	−0.002+		
															(0.001)	(0.001)		
share specialist																	0.060**	−0.010
																	(0.022)	(0.018)
share trained																	0.000	−0.002
																	(0.004)	(0.003)
share experts																	−0.011	−0.005
																	(0.013)	(0.011)
First stage:																		
older	2.241***		1.313***		3.015***		2.201***		2.201***		3.464***		2.580***		2.157***		2.362***	
	(0.238)		(0.316)		(0.178)		(0.234)		(0.234)		(0.205)		(0.575)		(0.237)		(0.300)	
F-stat. (1st stage)	87.4		17.2		96.2		89.1		89.1		282.8		20.4		83.4		62.6	
Num.Obs.	36 838	36 838	20 985	20 985	53 562	53 562	36 838	36 838	36 838	36 838	34 917	34 917	6195	6195	35 989	35 989	22 374	22 374

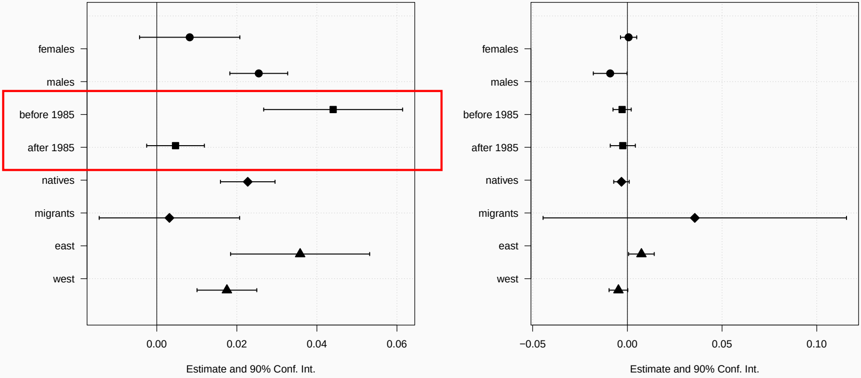
Komplett anpassen Heteroskedasticity-robust standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Heterogeneity: Demographics

Back

i

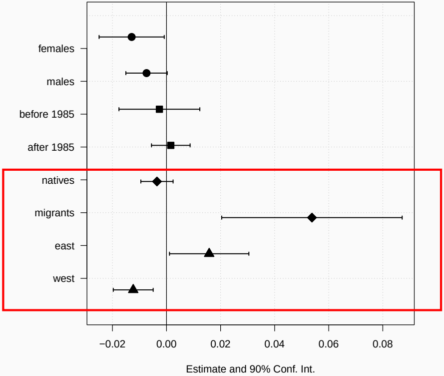
Figure 16: Heterogeneity by demographics



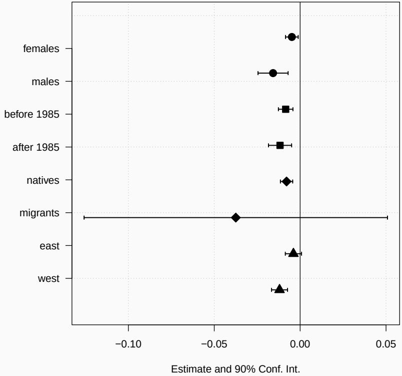
(a) Overeducation: Provided sample

(b) Overeducation: Imputed sample

Heterogeneity: Demographics Back ii



(c) Undereducation: Provided sample



(d) Undereducation: Imputed sample

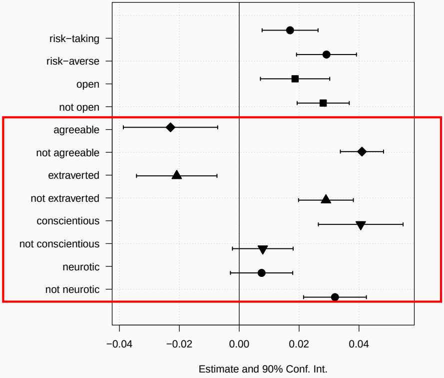
Note: 16a, 16c, 16b and 16d display regression results for the likelihood of over- and undereducation from fuzzy RDD estimations for different subsamples, respectively. 16a and 16c rely on the provided sample, while 16b and 16d use the imputed sample. The subsamples distinguish individuals based on sex, cohort, migration background, and region.

Heterogeneity Personality

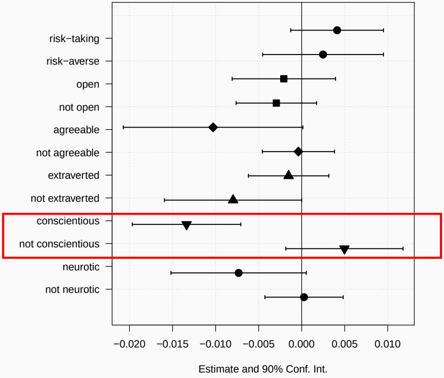
Back

i

Figure 17: Heterogeneity by personality traits

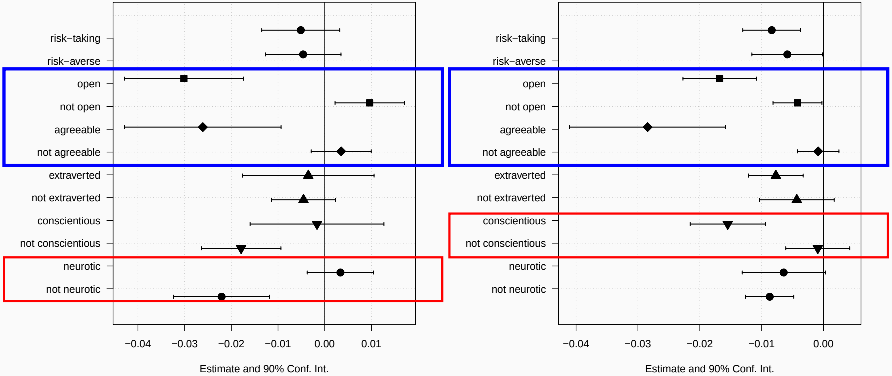


(a) Overeducation: Provided sample



(b) Overeducation: Imputed sample

Heterogeneity Personality Back ii



(c) Undereducation: Provided sample

(d) Undereducation: Imputed sample

Note: 17a, 17c, 17b and 17d display regression results for the likelihood of over- and undereducation from fuzzy RDD estimations for different subsamples, respectively. 17a and 17c rely on the provided sample, while 17b and 17d use the imputed sample. The subsamples distinguish individuals based on their risk-aversion, openness to experience, agreeableness, extraversion, conscientiousness, and neuroticism.