



Growing up in ethnic enclaves: language proficiency and educational attainment of immigrant children

Alexander M. Danzer^{1,2,3,4} · Carsten Feuerbaum¹ · Marc Piopiunik^{4,5} · Ludger Woessmann^{2,4,5,6}

Received: 7 August 2020 / Accepted: 1 February 2022 / Published online: 2 April 2022
© The Author(s) 2022

Abstract

Does the regional concentration of immigrants of the same ethnicity affect immigrant children's acquisition of host country language skills and educational attainment? We exploit the concentration of five ethnic groups in 1985 emanating from the exogenous placement of guest workers across German regions during the 1960s and 1970s. Results from a model with region and ethnicity fixed effects indicate that exposure to a higher own ethnic concentration impairs immigrant children's host country language proficiency and increases school dropout. A key mediating factor for the detrimental language effect is parents' lower speaking proficiency in the host country language, whereas inter-ethnic contacts with natives and economic conditions do not play a role in language proficiency or educational attainment.

Keywords Immigrant children · Ethnic concentration · Language · Education · Guest workers

Responsible editor: Klaus F. Zimmermann

We gratefully acknowledge the helpful comments and suggestions from the editor Klaus F. Zimmermann, four anonymous reviewers, George Borjas, Jesús Fernández-Huertas Moraga, Seth Gershenson, Jeffrey Grogger, John Haisken-DeNew, Martin Halla, Giovanni Mastrobuoni, Elie Murard, Francesco Ortega, Simone Schüller, Andreas Steinmayr, Steven Stillman, Uwe Sunde, Laura Viluma, and Joachim Winter, as well as seminar audiences at EALE in St. Gallen, ESPE in Glasgow, the German Economic Association in Vienna, the 14th IZA Annual Migration Meeting in Bonn, the CESifo area conference on economics of education, the RES Symposium of Junior Researchers in Bristol, the Applied Microeconomics Workshop in Bolzano, the Görres Society in Augsburg, the ZEW Workshop on Assimilation and Integration of Immigrants in Mannheim, the 4th MGSE Colloquium, the EBE Research Strategy Seminar, the ifo CEMIR Seminar, the ifo Junior Economist Workshop on Migration Research, the 22nd BGPE Research Workshop in Munich, and the IZA Summer School in Buch. We thank Stefan Weil from the Statistical Office in Bad Ems for support with respect to the on-site use of the 1987 German Census. Feuerbaum acknowledges funding through the International Doctoral Program Evidence-Based Economics of the Elite Network of Bavaria. Piopiunik acknowledges funding through the Leibniz Association (SAW-2012-ifo-3). The contribution by Woessmann is part of the Deutsche Forschungsgemeinschaft project CRC TRR 190.

Extended author information available on the last page of the article

JEL Classification J15 · I20 · R23 · J61

1 Introduction

Many societies wonder about the best policies to integrate immigrants. One important question is the regional allocation of immigrants. To prevent excessive ethnic concentration, many European countries have dispersal policies that assign refugees to regions (Dustmann et al. 2017). Existing evidence tends to suggest, though, that enclaves may in fact facilitate the labor market integration of immigrants (Schüller 2016), presumably through positive network effects within ethnic groups (Dustmann et al. 2016). The successful integration of immigrants into host country societies in the long run may depend on the intergenerational effects of ethnic concentration on the immigrants' children. Immigrant children's proficiency in the host country language and their educational attainment play a particularly important role in long-term employment opportunities and in cultural and social integration (e.g., Schnepf 2007; Dustmann and Glitz 2011; Chiswick and Miller 2015). On the one hand, children's language acquisition as well as educational and labor market integration may benefit from ethnic enclaves that provide useful information, reduced discrimination, and positive role models, e.g., through ethnic entrepreneurs (Andersson et al. 2021). On the other hand, immigrant children may also potentially be hindered by limited exposure to native children, reduced options for language acquisition, and lower socioeconomic opportunities for their families. In this paper, we study the effect of regional ethnic concentration on the language proficiency and educational attainment of immigrant children.

Our analysis exploits the regional distribution of immigrants in 1975/1985 based on the placement policy of the German guest worker program. Between 1955 and 1973, the German government actively recruited foreign workers to fill low-skilled labor shortages. The guest workers were enlisted in various countries of origin and then quasi-exogenously placed across West German firms. Workers could not choose specific firms and firms could not choose specific workers, and firms had no prior information about workers' specific skills or other characteristics. Once recruited, the program implied a guaranteed immigration permit and immediate placement and job start. Location preferences of guest workers were not considered, and they were not free to change location during a lock-in period. The guest worker population was mainly low-skilled: Only 7% had completed over 12 years of education (compared to 27% of other immigrants in Germany at the time), and more than one-third had left school without any degree. The skill level of guest workers was stable over the entire placement period (Marplan 1982).

The German Socio-Economic Panel (SOEP) allows us to extract a sample of roughly 1000 children whose parents immigrated into Germany from five different countries of origin during the period of the guest worker program. In contrast to administrative datasets, the SOEP household panel provides information on these children's host country language proficiency, as well as their educational attainment. In addition, the SOEP contains rich information on parents' speaking and writing abilities, friendships with Germans, and indicators for parents' social and

labor market integration that allows us to analyze factors that may mediate the effect of ethnic concentration on child outcomes. We merge the SOEP data on individual immigrant children with administrative data on the regional concentration of different ethnicities.

The initial regional assignment of guest workers gives rise to what we view as plausibly exogenous variation in ethnic concentration across regions, reducing the concern of potential bias from endogenous sorting of immigrants into enclaves of co-ethnics. A limitation of our data is that we do not observe guest workers' initially assigned location but only their location in 1985, requiring the identifying assumption that guest workers did not self-select across regions between their arrival in the 1960s/1970s and 1985 in ways that are systematic to our relationship of interest. We argue that this assumption is warranted in our setting by specific features of the German guest worker program: The work permits required guest workers to stay in the initially assigned region for a lock-in period of several years, and immediate integration into the labor market upon arrival minimized incentives to move to other regions. Two pieces of evidence further corroborate the assumption. First, we show that in 1985, demographics of guest worker parents and their children are balanced across regions with low and high ethnic concentration. Second, we show that in the 10 years after 1985, the movement of guest workers across regions was orthogonal to ethnic concentration and to our outcome measures.

To additionally account for *any* type of region-specific or ethnicity-specific differences, our models include region and ethnicity fixed effects. Region fixed effects ensure that any region-specific peculiarities are accounted for to the extent that they are common across guest worker ethnicities. Ethnicity (country-of-origin) fixed effects ensure that any ethnicity-specific differentials in integration are accounted for to the extent that they are common across regions. Thus, we identify the effect of ethnic concentration by observing immigrants who are exposed to differential concentrations of co-ethnics within the same region, thereby alleviating concerns of potential bias from endogenous location choices and from unobserved factors of different ethnic groups such as differing baseline willingness or disposition to integrate.

Our results indicate that growing up in ethnic enclaves significantly reduces immigrant children's proficiency in the host country language and their educational attainment. In particular, a one log-point increase in the size of the own ethnic group in the region—equivalent, e.g., to increasing an ethnicity's share in the regional population from 1.0 to 2.8—leads to a reduction in the German speaking proficiency of the children of the guest worker generation by 19% of a standard deviation, and a reduction in the German writing proficiency by 17% of a standard deviation. In addition, a one log-point increase in exposure to own ethnic concentration increases the likelihood that the immigrant child drops out of school without any degree by 5.6 percentage points, a large effect given the average of 7.1%. Although less robust, there is some indication that ethnic enclaves also reduce the probability of obtaining an intermediate or higher school degree. Effects tend to be larger for those immigrant children who were born abroad, but they do not differ significantly by gender.

The rich background information on children and parents contained in the SOEP allows us to analyze several mediating factors. Potential mechanisms underlying the negative effect of growing up in ethnic enclaves include parents' lower host

country language proficiency, reduced interactions with natives, and lower wages and employment opportunities of immigrant parents. We find that differences in parents' ability to speak the German language—which is strongly related to their children's German language proficiency—can in fact account for much of the effect of growing up in ethnic enclaves. Once parental German speaking abilities are controlled for, the estimated effect of ethnic concentration on children's language proficiency is reduced to close to zero. In this analysis, it proves essential to address measurement error in the self-reported parental language measure by implementing an instrumental variable approach that uses parents' responses on the same survey item from consecutive years (leads and lags) as instruments (Dustmann and van Soest 2002). While measures of parental writing abilities, friendships with German children, visits from Germans at home, parental unemployment, and household income are also significantly related to immigrant children's language proficiency, they do not account for the negative effect of ethnic concentration. None of the investigated mechanisms can account for the negative enclave effect on school dropout.

Our results are robust to a number of sensitivity analyses. We use alternative functional forms for the measure of ethnic concentration, instrument ethnic concentration at the time of observation (1985) by the ethnic concentration observed a decade earlier, use social-security and census data to construct the ethnic concentration measure, measure ethnic concentration at different levels of regional aggregation, and account for interview mode (which may influence self-assessed German language proficiency). We also show that after 1985, neither return migration nor regional migration within Germany were selective with respect to ethnic concentration and that ethnic concentration did not affect family size.

Our paper contributes to several strands of literature. A number of papers address the link between ethnic enclaves and the human capital acquisition of immigrant children. Grönqvist (2006) finds lower university attainment among refugees in enclaves in Sweden. Cortes (2006) detects no test score disadvantages for enclave schools in the USA. Jensen and Rasmussen (2011) find lower test scores among immigrant children in enclaves in Denmark. In the one previous study directly addressing bias from self-selection into ethnic enclaves, Åslund et al. (2011) use a refugee placement policy in Sweden and find that the concentration of highly educated co-ethnics positively affects the achievement of immigrant students in school. We contribute to this literature by providing novel evidence for a different group of immigrants (low-educated guest workers in Germany), by studying effects on language skills, and by finding that in a different setting, ethnic concentration can have a negative effect on immigrant children's language proficiency and educational attainment.

Given the shared emphasis on identification, it is useful to provide a comparison of our study with the Swedish study by Åslund et al. (2011). The main differences relate to the target population and to the results. Guest workers in Germany arrived with a signed labor contract and were quasi-exogenously distributed across regions, whereas Sweden assigned refugees according to municipal integration capacities, which were related to labor market and educational opportunities. Since all guest workers in Germany were employed upon arrival, our study can switch off one

potential channel of how enclaves affect human capital outcomes of children. While our setting focuses on children of a rather homogeneous group of low-skilled labor migrants, an important aspect of Åslund et al. (2011) is to focus on a heterogeneous group of partly well-educated refugees: The sample share who completed over 12 years of education in their setting (39%) is more than five times as large as in ours. Most obviously, our main finding also differs strongly from the Swedish study, indicating that existing evidence from humanitarian immigrants may not generalize to other settings and that any effect of ethnic concentration may strongly depend on the skill and employment levels of co-ethnics in the enclave.¹

A vast literature studies the effects of ethnic enclaves on the *economic* integration of *adult* immigrants (see Schüller 2016, for an overview). Using dispersal policies in Sweden and Denmark, respectively, Edin et al. (2003) and Damm (2009) find positive network effects of ethnic concentration on immigrants' labor market outcomes. By contrast, studying the same setting as in our paper, Danzer and Yaman (2016) and Constant, Schüller, and Zimmermann (2013) find negative effects of ethnic concentration on adult immigrants' proficiency in the host country language and their cultural integration, respectively (see Chiswick 1998, for similar results for Israel). In a different German setting, Battisti et al. (2016) find positive short-term but negative long-term effects of ethnic concentration on labor-market outcomes, with the negative effect being related to lower human capital investments and larger job mismatch. Generally, there is growing awareness that immigrants tend to assimilate as communities, emphasizing the importance of ethnic links (Hatton and Leigh 2011).

Beyond immigrant integration, another large literature studies the effect of spatial segregation and concentration on the economic success of racial minorities, usually finding negative effects (e.g., Cutler and Glaeser 1997; Fryer 2011). More generally, a growing literature studies the effect of exposure to different quality neighborhoods during childhood on children's outcomes in the short and long run (e.g., Chetty et al. 2016; Chetty and Hendren 2018; Gibbons et al. 2013, 2017).

We contribute to this literature by estimating well-identified effects of growing up in low-skilled ethnic enclaves on the language proficiency and educational attainment of immigrant children and by providing a rich analysis of mediating factors. Our findings indicate that parents' limited proficiency in speaking the host country language is a key mediating factor of the negative impact of ethnic enclaves on immigrant children's language proficiency. By contrast, limited interaction with natives and parental economic conditions do not seem to be leading mechanisms. Overall, the opportunity to benefit from large social networks of co-ethnics may be particularly relevant for newly arriving immigrants, but less so for the long-term integration of the children of settled immigrants. In fact, most of the arguments in

¹ Our analysis differs from Åslund et al. (2011) in additional ways. They observe GPA scores and on-time school graduation, whereas we use long-run panel data to assess final school graduation as an economically important outcome. Our survey data allow us to investigate the underlying mechanisms in detail, yielding important insights. In addition, the guaranteed immigration permit, immediate placement and job start, disregard of location preferences, and lock-in period of prohibited residential change imply that the German guest worker placement provides for particularly strict and clean features of the experimental setup in our setting.

favor of ethnic enclaves tend to relate to the labor-market integration of adult immigrants but bear less relevance for integration beyond the labor market. Regarding the cultural and educational integration of the immigrants' second generation, our results suggest that the dispersal policies of several European countries may indeed facilitate integration.

In what follows, Section 2 provides institutional background on the German guest worker program. Section 3 describes the SOEP household data and the administrative data used to compute ethnic concentrations. Section 4 introduces our empirical model and shows balancing of demographic characteristics across regions with low and high ethnic concentration. Section 5 presents our main results on the effect of ethnic concentration on immigrant children's outcomes. Section 6 investigates the relevance of several potential mediating factors. Section 7 provides a number of robustness analyses. Section 8 concludes.

2 Institutional background on the German guest worker program

The German guest worker program was one of the largest guest worker programs worldwide. West Germany (hereafter, Germany) signed bilateral guest worker treaties with Italy in 1955, Greece and Spain in 1960, Turkey in 1961, and Yugoslavia in 1968. During a period of rapid economic growth in the 1960s and early 1970s, increasing demand for low-skilled workers induced a massive inflow of labor migrants to fill the numerous open positions in the economy. Given that all treaties were designed to attract low-skilled and mainly young workers, the guest workers constitute a rather homogeneous immigrant population that is, on average, less educated than the German workers. Due to the severe economic recession triggered by the oil crisis, Germany stopped the recruitment of guest workers in 1973. By that time, 2.6 million foreign workers were employed in Germany, implying that 12% of the labor force were foreigners (Federal Employment Agency 1974).

To take up employment, guest workers were required to hold a valid work permit (*Arbeitserlaubnisbescheinigung*). The formal process of obtaining this permit was initiated at the foreign branches of the German Federal Employment Agency in the guest worker countries, which was similar for all source countries.² Having been recruited through advertisements, e.g., in cinemas and on TV in their home countries, potential guest workers had to show up at these branches. They were screened for basic literacy and underwent medical check-ups.³ Then, guest workers were

² The foreign branches of the German employment agency were called *Deutsche Kommission* in Greece, Italy, and Spain, *Deutsche Verbindungsstelle* in Turkey, and *Deutsche Delegation* in Yugoslavia. Italians could later enter Germany more freely within the European Economic Community (EEC) framework, but were placed by an internal recruitment branch within Germany (*Zentralstelle für Arbeitsvermittlung*). The German embassy in Yugoslavia opened a second track for guest worker applications in 1970 to account for the high number of applicants. For more details, see Dohse (1981) and Federal Employment Agency (1962).

³ At this occasion, applicants also received information on the working and living conditions in Germany. Guest workers were predominantly low-skilled due to the nature of labor demand in the construction, mining, metal, and ferrous industries at that time and because the governments of the sending coun-

matched with German employers. Enrolling in their home countries, the guest workers never met a placement officer in Germany, could not pick a region or firm, and could not turn down a location offer without dropping out of the program.

The German employers submitted recruitment requests together with blank work contracts to their local labor offices, which forwarded them to the foreign branches after initial approval.⁴ The firms received almost no information about their requested workers before arrival and in practice generally could not select workers based on job skills or country of origin (Feuser 1961; Fassbender 1966; and Voelker 1976). As recruited workers were meant to be manual workers, specific skills were not even recorded at application, so that placement in a specific firm was independent of any potential work experience or other skills of the worker.

Successful applicants got a work contract from a specific German company and a 1-year work permit that was only valid for employment at the specific firm (Feuser 1961). Recruited workers were then transferred to Germany in groups.⁵ After having stayed with their initial employer for at least 2 years and in the same occupation (and, in practice, in the same region for most guest workers) for at least 5 years, guest workers could receive an upgrade of their work permit (*Erweiterte Arbeitserlaubnisbescheinigung*) that included free job choice (Dahnen and Kozlowicz 1963).⁶

Given that the initial location in Germany depended on current labor demand, it was exogenous from the perspective of an individual guest worker. Unlike in settings where immigrants can potentially self-select into regions and ethnic enclaves (e.g., Bauer et al. 2005), the guest worker recruitment process generated exogenous variation in the concentration of ethnicities that allows us to estimate the effect of ethnic concentration on immigrant children's outcomes. The regional variation in ethnic concentration differed across ethnicities for at least three reasons. First, regional labor demand fluctuated over time, and guest worker recruitment started in the different countries of origin at different points in time. For instance, the relative size of the guest worker population in North Rhine-Westphalia declined between 1962 and 1973, whereas the share of guest workers in Lower Saxony and Bremen increased (Federal Employment Agency 1964). Second, labor supply in the countries of origin varied over time. More guest workers were recruited from countries that had temporarily abundant labor supply, like Turkey in 1964 (Federal Employment Agency 1965). Third, while guest workers from Spain and Portugal arrived by train in Cologne in the West, the port of entry for the remaining guest worker groups

Footnote 3 (continued)

tries preferred emigration from underdeveloped and disaster-ridden areas (Pennix and Van Renselaar, 1976).

⁴ The local labor office checked whether German workers were available for the open positions, whether housing was available for foreign workers, and whether the request fulfilled all conditions of the bilateral treaty.

⁵ Travel costs were covered by recruiting firms by paying a small flat fee for each recruited worker.

⁶ As an alternative recruitment process, employers were allowed to request guest workers by name if there was a personal relationship to that person, for example, through recommendations by relatives or friends who were already employed at that firm. Recruitment by name became more important as guest workers recommended their spouses. However, for various reasons, a large fraction of individuals who were requested by name were eventually not hired (Federal Employment Agency, 1972).

(Greek, Italian, and Turkish) was Munich in the South, generating additional variation in the ethnic concentration across nationalities.

In 1973, the guest worker recruitment was officially stopped. However, immigration of family members within the family reunification framework ensured high levels of inflows from guest worker countries also afterwards. Those family members immigrated on the basis of the *Aliens Act* of 1965 and were granted a residence permit when joining a guest worker family member.

3 Data

Our analysis uses individual-level information on guest workers and their children from the German Socio-Economic Panel (Section 3.1). We construct our main measure of ethnic concentration from a large employee sample of the Federal Employment Agency's Research Institute (Section 3.2).

3.1 Survey data on guest workers and their children

We use information on guest workers and their children from the German Socio-Economic Panel (SOEP, version 30), a large annual household survey that is representative of the resident population in Germany (Goebel et al. 2019). The first SOEP wave in 1984 strongly oversampled guest workers (by a factor of four). As a consequence, 1393 of the 5921 SOEP households originated from the five guest worker countries, which comprised the largest foreigner populations in Germany at the time (Sample B). For each ethnicity, an independent random sample was drawn to allow for stand-alone analyses (Haisken-DeNew and Frick 2005). The SOEP contains detailed information on individual characteristics, including educational attainment and, for foreigners, self-reported German speaking and writing proficiencies.⁷ The 1985 survey is the first wave that provides sufficient geographic information on the region of residence at the county level. Hence, we identify guest workers and their region of residence based on information in the 1985 wave. Using information from mothers' birth biography and pointers to their partners in 1985, we link parents to their children.⁸ While the SOEP does not contain a direct indicator of guest workers, we identify guest workers by their country of origin, year of immigration, and age at migration. The guest worker immigration differed from previous immigration experiences in Germany in that guest workers were predominantly male, young, and low-educated.⁹

⁷ All questionnaires, in German and partly in English, are available at https://www.diw.de/en/diw_02.c.222729.en/questionnaires.html.

⁸ We use only children whom we can assign mother and father. We assume that matched partners of mothers are the fathers of the children. In 1985, the share of mothers who were married was 98.3%; only three mothers (0.3%) were not married and another three mothers (0.3%) were widowed; the share of mothers with missing marital status is 1.1%.

⁹ We can compare demographic characteristics of guest workers and immigrants from other countries in 1985, based on all first-generation immigrants in the SOEP (details available upon request). Guest workers are, on average, 5 years younger and more likely to be employed in 1985, consistent with their

Our analysis sample consists of 1065 guest worker children with Greek, Italian, Spanish, Turkish, or Yugoslav background. To be included in the sample, children must have at least one parent who was aged 18 or older at immigration and who arrived in Germany during the period when the guest worker program with her/his home country was in place. We restrict the sample to children aged 13 or younger at migration since the focus of our study is to investigate the impact of the region where children grow up.¹⁰ We keep only children with at least one observation for self-reported German language proficiency or one observation for educational attainment.¹¹

We measure children's German language proficiency by two distinct outcomes: speaking proficiency and writing proficiency. Both language outcomes are self-reported and based on the following question: "In your opinion, how well do you speak and write German?" Answers are provided on a five-point scale: very well, well, fairly, poorly, and not at all. Children report their German language proficiency for the first time at the age of 17 or 18, i.e., when they are personally interviewed in the SOEP for the first time (see Appendix Figure 2 for the timing of variable measurements). An advantage of the panel data is that we observe multiple observations of self-reported language proficiency for each child (five observations per child on average), resulting in a large sample of language proficiency observations. An additional advantage of the panel data is that we can address measurement error in *parents'* language proficiency by instrumenting the self-reported language proficiency in a given year with their self-assessments in previous or succeeding years (see Section 6.1). In our sample of language proficiency, each observation is at the child-year level. This sample is based on the SOEP waves 1984–1987, 1989, 1991, 1993, and every 2 years from 1997 to 2005, including about 4900 child-year observations.¹² We standardize each outcome of children's language proficiency to have mean 0 and standard deviation 1.

Children's educational attainment is also measured by two variables. The binary indicator "any school degree" equals 1 if the child obtained any type of school degree and 0 if the child dropped out of school without any degree. The binary indicator "at least intermediate school degree" equals 1 if the child obtained an

Footnote 9 (continued)

demand-driven recruitment. Guest workers have around 2 years less education. Most strikingly, while 36% of guest workers have left school without any type of school degree, this share is only 3% among other immigrants. Only 7% of guest workers have more than 12 years of schooling, compared to 27% of immigrants from other countries.

¹⁰ We present heterogeneity results below for guest worker children born in Germany vs. children born abroad.

¹¹ The main reason for missing values on language proficiency and educational attainment is that households stopped participating in the SOEP survey before the children turned 17 years old and would be personally interviewed for the first time. The share of children with missing values on the outcomes does not differ between regions with low and regions with high co-ethnic concentration (see bottom of Table 1).

¹² Our panel data set for children's language proficiency is unbalanced for two reasons. First, some children were younger than age 17 in 1985 and therefore did not participate in the personal interviews during the first years of our panel data. Second, some children (usually the entire household) left the SOEP survey before 2005.

intermediate school degree (*Realschulabschluss*) or an advanced degree (*Abitur*) and 0 otherwise.¹³ Children's educational attainment is based on the most recent available information in the SOEP.¹⁴

Table 1 reports descriptive statistics of children's outcomes and demographic characteristics of children and their parents, separately for regions with low and high ethnic concentration (split at the ethnicity-specific median of the share of ethnic concentration in 1985). Immigrant children living in regions with a high co-ethnic concentration report lower German speaking proficiency (statistically significant at the 12% level) and lower writing proficiency (significant at the 10% level) than immigrant children living in low co-ethnic concentration regions.¹⁵ Consistent with this finding, immigrant children in regions with high co-ethnic concentration are significantly less likely to obtain a school degree and slightly (and statistically insignificantly) less likely to obtain at least an intermediate school degree.

In terms of ethnicities, 37% of immigrant children in our sample are Turkish, 19% each are Italian and Yugoslav, 15% are Greek, and 10% are Spanish. We identify the ethnicity of the immigrant children primarily based on their first citizenship (94.2% of the children in our sample). In the case of a German citizenship or missing citizenship information, ethnicity is based on the children's country of birth or their parents' nationality (see Appendix Table 10 for definitions of all individual-level variables).¹⁶ A slight majority of immigrant children in the sample (57.1%) were born in Germany. The average year of birth is 1971, and the average age at migration is 2.8 years.

The SOEP also contains a rich set of additional individual characteristics, including the immigration history, educational attainment, and labor-market outcomes of adults.¹⁷ This wealth of information allows us to investigate several potential

¹³ In Germany, there are three types of secondary school degrees: basic (*Hauptschulabschluss*), intermediate (*Realschulabschluss*), and advanced (*Abitur*). The basic degree is acquired at a basic school (*Hauptschule*), the intermediate degree at a so-called *Realschule*, and the advanced degree at a *Gymnasium*. A small share of children in our sample (2.9%) reported to have obtained another type of school-leaving certificate. While we assume that this other type of school-leaving certificate is equivalent to an intermediate school degree, the results do not depend on this assumption.

¹⁴ If the most recent available information indicates dropout or no school degree (yet), we checked for school-leaving degrees reported in previous waves. For only nine children, we adjusted the educational attainment variables based on previously reported school-leaving degrees. For 83% of children, educational attainment is measured at age 22 or younger. Controlling for the age when the child reported her educational attainment yields very similar results (available upon request).

¹⁵ The main variation in levels of German language proficiency is from "fairly" to "very well". On the underlying scale, 55.5% (45.3) of children in our sample report to speak (write) German "very well," 36.2% (37.2) "well," 6.5% (12.3) "fairly," 1.3% (3.8) "poorly," and 0.5% (1.4) "not at all." Levels are lower for those not born in Germany than for those born in Germany.

¹⁶ In the very few instances in which children have a German citizenship or information on citizenship is missing *and* the nationality of mother and father differs, we use mother's nationality or mother's country of birth.

¹⁷ As is typical for surveys, our data on guest workers and their children contain missing values for some variables. Since our set of control variables is large, dropping all children with any missing value would substantially reduce the sample size. We therefore impute missing values by using the mean of each control variable. For binary indicators, imputed means are rounded to the closest integer. To ensure that results are not driven by imputed values, all our estimations include imputation dummies for each variable. The share of imputed values is small for all demographic characteristics (at most 3.6%). The main results are similar when excluding all observations with imputed values (available upon request).

Table 1 Descriptive statistics by degree of ethnic concentration

Variable	Low EC	High EC	Diff	P-Value	Obs
<i>Outcomes (children)</i>					
Speaking proficiency	0.08	-0.07	0.16	0.12	996
Writing proficiency	0.09	-0.08	0.17	0.07	996
Any school degree	0.95	0.91	0.05	0.01	1005
At least intermediate school degree	0.43	0.41	0.02	0.64	1005
<i>Children</i>					
First year of language assessment	1989.29	1988.99	0.30	0.53	996
Male	0.54	0.57	-0.02	0.44	1065
Year of birth	1971.28	1971.03	0.25	0.67	1065
Age at migration	2.55	2.95	-0.40	0.28	1065
Born in Germany	0.58	0.56	0.02	0.55	1065
Greek	0.15	0.15	0.00	0.98	1065
Italian	0.19	0.19	-0.00	0.99	1065
Spanish	0.10	0.09	0.01	0.85	1065
Turkish	0.37	0.38	-0.01	0.92	1065
Yugoslav	0.19	0.18	0.00	0.98	1065
<i>Mothers</i>					
Year of birth	1944.36	1943.90	0.46	0.57	1065
Year of immigration (for the foreign born)	1970.54	1970.29	0.26	0.70	1022
Age at migration (for the foreign born)	26.20	26.49	-0.29	0.70	1022
Born in Germany	0.04	0.04	0.01	0.74	1065
Migrant	0.97	0.98	-0.01	0.59	1065
<i>Education in country of origin</i>					
No schooling	0.23	0.23	-0.00	1.00	1065
Incomplete compulsory schooling	0.41	0.36	0.05	0.38	1065
At least compulsory schooling	0.36	0.41	-0.05	0.40	1065
Years of education	8.29	8.29	-0.00	1.00	1065
Never moved flat since arrival in Germany	0.13	0.13	0.01	0.84	1065
Children	3.67	3.70	-0.02	0.93	1065
Out of the labor force (1984–1986)	0.57	0.51	0.05	0.36	1065
Unemployed (1984–1986)	0.06	0.04	0.02	0.27	1065
<i>Fathers</i>					
Year of birth	1939.78	1940.29	-0.51	0.48	1065
Year of immigration (for the foreign born)	1967.56	1967.71	-0.15	0.76	1056
Age at migration (for the foreign born)	27.73	27.46	0.28	0.64	1056
Born in Germany	0.01	0.01	0.00	0.70	1065
Migrant	0.99	0.99	-0.00	0.70	1065
<i>Education in country of origin</i>					
No schooling	0.09	0.12	-0.03	0.42	1065
Incomplete compulsory schooling	0.27	0.28	-0.01	0.79	1065
At least compulsory schooling	0.63	0.59	0.04	0.46	1065
Years of education	9.15	9.06	0.09	0.66	1065

Table 1 (continued)

Variable	Low EC	High EC	Diff	P-Value	Obs
Never moved flat since arrival in Germany	0.04	0.04	0.00	0.87	1065
Out of the labor force (1984–1986)	0.09	0.08	0.01	0.66	1065
Unemployed (1984–1986)	0.10	0.04	0.06	0.01	1065
Household income (1984–1986)	1700.37	1821.21	–120.84	0.09	1065
<i>For comparison</i>					
Official unemployment rate (1985)	0.11	0.08	0.03	0.00	1065
Information on language proficiency available	0.70	0.70	–0.00	1.00	1429
Information on school degree available	0.71	0.70	0.02	0.66	1429
Children	500	565			1065

Variable means by degree of ethnic concentration (EC). Low vs. high EC split at the ethnicity-specific median of the share of ethnic concentration in 1985. *P* values refer to two-sided tests with standard errors clustered at region-ethnicity level. Speaking/writing proficiency: first reported self-assessed speaking/writing ability in German (from 1 = “not at all” to 5 = “very well”), normalized to mean 0 and standard deviation 1. Any (at least intermediate) school degree: 1 if individual obtained any (at least an intermediate) type of school degree, 0 otherwise. Household income, out of the labor force, and unemployed refer to 3-year means over 1984–1986. Information on language proficiency/school degree available: 1 if information on corresponding outcome is available in the SOEP data in at least one survey year, 0 otherwise. The *F*-statistic of joint significance of a regression of a high-concentration dummy on individual characteristics is 0.22 for children (*p* value 0.992), 0.51 for mothers (0.865), and 3.53 for fathers (0.0005) (which includes household income). Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Federal Employment Agency (2017).

mediating factors that may drive the effects of ethnic concentration. As potential mediating factors, we investigate parents’ speaking and writing proficiencies in German, parents’ employment status, household income, visits from Germans at home, and whether the child’s first friend is German. Parents’ mediating factors are based on the average of mothers’ and fathers’ information.¹⁸

3.2 Ethnic concentration

We compute measures of the concentration of co-ethnics in the region separately for the five guest worker nationalities (Greek, Italian, Spanish, Turkish, and Yugoslav) at the regional level of the so-called *Anpassungsschichten* (hereafter “areas”). Typically comprising several counties, these regions constitute regional labor markets. In West Germany, there were 103 areas in 1985 with an average population of about half a million people. While smaller geographic units may better reflect small-scale ethnic neighborhoods, a higher level of regional aggregation produces more conservative estimates and circumvents potential bias from the typical sorting of immigrants into close-by cities or across city districts. Since compared to the USA, the

¹⁸ Table 1 shows that a few mothers and fathers have been born in Germany. As a robustness check, we have excluded all 40 children (3.7%) with one parent born in Germany. This yields very similar results (available upon request).

degree of ethnic and social segregation is low in Germany (Musterd 2005), the area level reflects the most suitable level of analysis (Danzer and Yaman, 2016, discuss the trade-off between small and large units). In any case, our findings are fully confirmed in robustness analysis at the more fine-grained level of German counties.¹⁹

For the measurement of ethnic concentration, we use the Sample of Integrated Labor Market Biographies (*Stichprobe der Integrierten Arbeitsmarktbiografien*, SIAB) of the Research Institute of the Federal Employment Agency (*Institut für Arbeitsmarkt- und Berufsforschung*, IAB). The SIAB is a 2% random sample of all individuals in Germany who are employed subject to social security, job seeking, or benefit recipients as contained in the Integrated Employment Biographies of the German social security system (Dorner et al. 2011). We use data from 1985, the year when guest workers' region of residence is first observed in the SOEP data.

Ethnic concentration, our key explanatory variable, is measured by the logarithm of the size of the ethnic community in the region of residence in 1985 (see Appendix Table 11 for definitions of regional variables). In our regression analyses, region fixed effects control for the size of the overall population in a region. While it is common to measure ethnic concentration as the log size of the own ethnicity (e.g., Edin et al. 2003; Damm 2009; Åslund et al. 2011), we also show robustness of our results to using the share of the own ethnicity in the total regional population as an alternative measure (e.g., Chiswick 2009; Danzer and Yaman 2013, 2016). We match our measures of ethnic concentration to the individual-level SOEP data at the level of areas and ethnicities.

The extensive demand-driven recruitment of guest workers generated substantial variation in ethnic concentrations across regions. Figure 1 shows the distribution of ethnic concentrations separately for each of the five ethnicities across the 103 West German areas in 1985 (see Appendix Table 12 for descriptive statistics). There are clear differences in the settlement structures between the guest worker ethnicities. For example, while Spanish guest workers tend to be concentrated in central Germany, Italians and Yugoslavs are more concentrated in the southern regions. In our analysis, we exploit the differential concentrations of ethnicities within regions.

For robustness analyses, we also use the 1987 German Census, which includes the entire population in Germany, to compute alternative measures of ethnic concentration. Being based on a 2% employee random sample, the SIAB measure of ethnic concentration may contain classical measurement error, biasing our estimates toward zero. In addition, if the regional share of co-ethnics in the employee sample does not reflect the ethnic concentration in the overall population—for example, because of differential labor-market participation rates—there may be non-classical

¹⁹ A potential source of bias in studies based on potentially arbitrary administrative units stems from the modifiable areal unit problem. We think that this problem is of secondary importance in our application, not least because we exploit variation across different ethnicities within the same geographic units. Importantly, our areal choices are not arbitrary but reflect commuting patterns and, hence underlying economic structures which are plausibly relevant in a labor-based placement scheme. Also, the bias from the modifiable areal unit problem tends to decline with higher levels of aggregation because there are fewer unit borders. In this sense, our areal choice is more conservative than using more disaggregated geographic units.

measurement error. In robustness analyses, we therefore also use an alternative measure of ethnic concentration based on data from the 1987 Census. The depth of the Census data also allows us to perform robustness analyses that define ethnic enclaves at the level of the 328 West German counties. A disadvantage of the 1987 Census is that it does not allow to compute ethnic concentrations for Spanish guest workers (who are included in a residual “other citizenship” category), which reduces the sample size and excludes one of the five guest worker ethnicities. In addition, the ethnicity measure in the Census is based on citizenship information (as country of birth is not observed in the Census), and the 1987 Census measures ethnic concentrations 2 years later than the 1985 SIAB data. Appendix Figs. 3 and 4 depict the distribution of the Census-based measures of ethnic concentration separately for the four ethnicities at the level of areas and counties, respectively.

4 Empirical model

In this section, we discuss the basic setup of our empirical model (Section 4.1) and show the balancing of demographic characteristics of guest workers and their children across regions with low and high concentrations of co-ethnics (Section 4.2).

4.1 Model setup with region and ethnicity fixed effects

We aim to estimate the effect of ethnic enclaves on the language proficiency and educational attainment of immigrant children. Exploiting the regional distribution following the quasi-exogenous placement of guest workers, our basic model setup expresses immigrant children’s outcomes as a function of the concentration of their ethnicity in their region. Conditioning on fixed effects for ethnicities and regions, the model is identified from the concentration of an ethnicity in a particular region compared to the concentration of other guest worker ethnicities in the same region.

When estimating the effect of ethnic enclaves on immigrant children’s host country language proficiency, we make use of the panel structure of the SOEP where immigrant children report their German language proficiency in multiple consecutive years. This allows estimating the following random effects model²⁰:

$$lang_{icrt} = \sigma_c + \delta_r + \tau_t + \beta_1 EC_{cr} + C'_{icr} \beta_2 + P'_{icr} \beta_3 + \mu_{icr} + \epsilon_{icrt} \quad (1)$$

where $lang_{icrt}$ is the German speaking and writing proficiencies, respectively, of child i of ethnicity (country-of-origin) c living in region r in year t . The key explanatory variable is the concentration of child i ’s ethnicity in her region, EC_{cr} .²¹ C_{icr} is a

²⁰ Just like a pooled OLS model, the random effects model effectively uses the average language proficiency across all observed time periods. As shown in Appendix Table 13, pooled OLS regressions yield very similar results. The same is true for cross-sectional regressions with average language skills as dependent variable (results available upon request).

²¹ As described in Section 3.2, ethnic concentration, EC_{cr} , is measured as the (log) size of child i ’s ethnic community in her region of residence in 1985, the first year in which the SOEP provides sufficient geographical information on guest workers. Including the quadratic of EC_{cr} does not indicate any non-

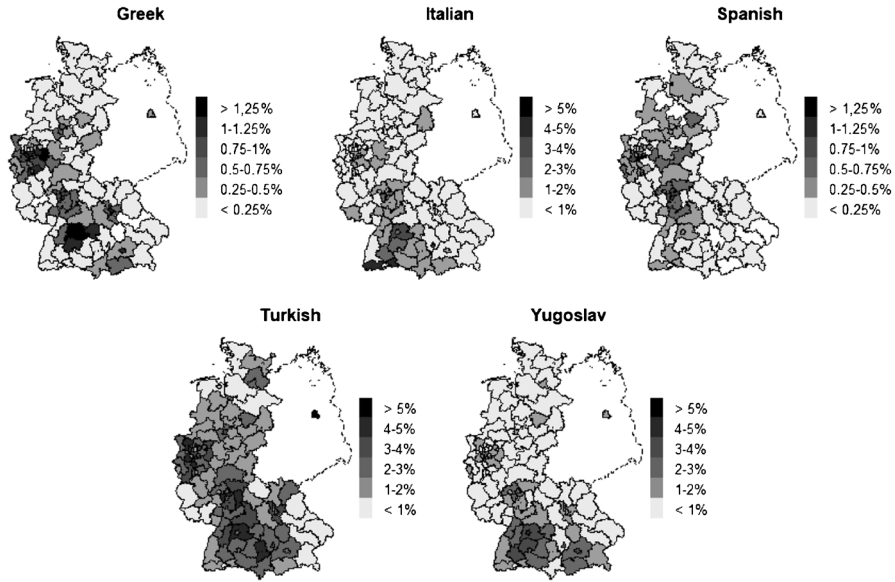


Fig. 1 Ethnic concentrations across West Germany, 1985. Ethnic group's share of the total population in the area of residence, 1985. Source: Institut für Arbeitsmarkt- und Berufsforschung (IAB). Own calculations of ethnic concentrations for 103 areas. Figures based on a historical GIS data file of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011)

vector of child characteristics, including gender, year of birth, and age at migration. P_{icr} is a vector of parent characteristics, including year of birth, year of arrival in Germany, education in country of origin, years of schooling, a migration indicator (which equals 0 for a few spouses who have no migration background),²² and the number of children for mothers. All models include fixed effects for ethnicities, σ_c ; fixed effects for regions, δ_r ; and fixed effects for the year when the child reported her language proficiency, τ_t . The individual-specific effects, μ_{icr} , are assumed to be i.i.d. random variables, and ϵ_{icrt} is an idiosyncratic error term. Throughout, we cluster standard errors at the region-by-ethnicity level, the level at which our measure of ethnic concentration varies.²³

Footnote 21 (continued)

linearities in the effects of ethnic concentration on language proficiency or school dropout, speaking against the existence of crucial thresholds or tipping points.

²² Among the parents in our sample, 2.9% of mothers and 0.8% of fathers are of German nationality without migration background.

²³ Our results are fully preserved when clustering standard errors at the regional level which would account for regional shocks that may affect all ethnicities similarly (not shown). Note that it is unlikely that labor-market shocks affect ethnicities within the same region differently since guest workers from different origin countries had similar qualification levels and worked in similar occupations (details available upon request).

To estimate the effect of ethnic concentration on immigrant children's educational attainment, we estimate the following OLS model using a cross-section of children:

$$educ_{icr} = \sigma_c + \delta_r + \theta_1 EC_{cr} + C'_{icr} \theta_2 + P'_{icr} \theta_3 + \varepsilon_{icr} \quad (2)$$

where $educ_{icr}$ is the educational attainment of child i , measured either by a binary indicator for having obtained any school degree or by a binary indicator for having obtained at least an intermediate school degree. As in Eq. (1), we include controls for child and parent characteristics as well as ethnicity and region fixed effects.

By including ethnicity fixed effects, we account for any differences between ethnicities, such as linguistic distance to the German language, cultural distance, school quality in the country of origin, and general willingness or disposition to integrate into the host country. By including region fixed effects, we exploit only variation in ethnic concentrations within the same region but do not use systematic differences in ethnic concentrations across regions. Thus, we control for any differences across regions, such as unemployment rates, wage levels, overall share of migrants, school quality, and attitudes of the native population. Our model therefore identifies the effect of ethnic concentration on immigrant children's outcomes from the presence of several immigrant groups with differing community sizes within the same region.

4.2 Test of balancing by ethnic concentration

As argued above, the placement policy of the German guest worker program led to quasi-exogenous variation in the regional placement of guest workers. While we observe guest workers only several years after their placement, we can test the plausibility of the exogeneity assumption by comparing observable characteristics of the immigrant children and their parents between regions with low and high ethnic concentration of the respective ethnicity. To do so, we split the sample at the ethnicity-specific median of the share of ethnic concentration in the child's region of residence in 1985. As Table 1 shows, none of the demographic characteristics of immigrant children differs significantly (individually or jointly) across regions with low and high co-ethnic concentration. The same is true for the demographic characteristics of mothers and fathers. Similarly, using the specification of our outcome model, there is no significant relationship between ethnic concentration and background characteristics when regressing the background characteristics on the share of ethnic concentration as well as ethnicity and region fixed effects (Appendix Table 14). Additionally, we perform regression-based balancing tests on the sample of fathers instead of children as in Damm and Dustmann (2014), yielding the same qualitative result (results available upon request). These balancing tests support our assumption that there was no systematic self-selection of guest workers into regions of differing ethnic concentration. Of course, as with any balancing test, we cannot test whether unobserved characteristics, such as parents' willingness to integrate into the German society, are balanced as well. Nonetheless, we run a regression-based balancing test as in Pei et al. (2019) with the candidate confounder—proxied

by father's education—as outcome variable, again finding no evidence for selection (results available upon request).

Beyond demographic backgrounds, the only exceptions where we find a significant difference between regions with low and high ethnic concentration are fathers' unemployment rates and household income. Interestingly, guest workers are *better* off in terms of employment and income in regions with *high* shares of co-ethnic concentration. If anything, this difference should work against finding any negative effect of ethnic concentration on children's outcomes. The unemployment difference observed for guest worker fathers in the SOEP sample is qualitatively in line with the overall unemployment rates in 1985 from the Federal Employment Agency (see bottom of Table 1). Thus, the unemployment difference likely reflects the fact that guest workers were particularly demanded in regions with booming industries, which were still characterized by lower unemployment levels in 1985. Of course, the region fixed effects in our regression models account for any general difference across regions, exploiting only within-regional variation across different ethnicities. Furthermore, as we show below, differences in unemployment and household income cannot account for the effect of ethnic concentration on children's outcomes.

The balancing of guest workers' demographic characteristics across regions with low and high ethnic concentration is particularly reassuring as we observe the location of guest workers in 1985 for the first time. As we do not observe the initial location to which guest workers had been assigned, our analysis is based on the assumption that any movement of guest workers across regions between their arrival in the 1960s/1970s and 1985 is orthogonal to our relationship of interest. Thus, the estimated coefficient on ethnic concentration would be biased downward (upward) if parents with adverse (advantageous) characteristics related to their child's outcomes moved to regions with high ethnic concentrations. The balancing results support our identifying assumption that guest workers in Germany did *not* systematically self-select into regions between their arrival and 1985.²⁴ Further supporting this assumption, we show in Section 7.7 below that the movement of guest workers across regions in the 10 years after 1985, when we can observe them, was in fact orthogonal to ethnic concentration and to our outcome measures.

This is in line with existing work investigating the German guest worker program. Previous studies also did not find any evidence of significant differences in demographic characteristics between guest workers living in regions with high concentrations of co-ethnics and those living in regions with low concentrations (Constant et al. 2013; Danzer and Yaman 2013, 2016). The evidence against endogenous sorting of immigrants into ethnic enclaves in our specific setting (although not necessarily in any immigrant setting) is perfectly consistent with two specific features of the German guest worker program.

²⁴ Alternatively regressing the *continuous* measure of the share of ethnic concentration on each demographic characteristic (including region and ethnicity fixed effects) yields only statistically insignificant coefficients, except for a significant positive coefficient on the variable indicating that parents had no schooling. In contrast, the coefficients on all four outcome variables are negative and statistically significant at about the 1% level (results available upon request).

First, as discussed above, guest workers were restricted in their residential choice as their work permit required them to stay in the initially assigned region for several years (Dahnen and Kozłowicz 1963). Thus, the formal rules of the guest worker program made it hardly possible for guest workers to move across regions during the initial years after their arrival.

Second, guest workers in Germany were well integrated into the labor market immediately upon arrival as they had been recruited specifically for the purpose of filling open positions in the German economy. As a result, the unemployment rate of foreigners in Germany was less than 1.5% in every year between 1968 and 1973 and was even lower than that of natives (Federal Employment Agency 1974). Since guest workers—who migrated to Germany with the aim to work—had been employed immediately upon arrival, the incentive to move to other regions was very low. Accordingly, the current settlement structures of immigrants in Germany have been shown to still reflect the demand for labor in the 1960s and 1970s (Schönwalder and Söhn 2009). Quite generally, ethnic segregation has been reasonably stable across workplaces and residential locations over the entire period from 1975 to 2008 (Glitz 2014).

In sum, the demographic characteristics of guest workers and their children are very similar across regions with low and high ethnic concentration. This finding supports our identification strategy of exploiting the quasi-exogenous placement of guest workers across German regions to estimate the effect of ethnic enclaves on immigrant children's outcomes.²⁵

5 Results

This section presents our main results (Section 5.1) and subgroup analyses (Section 5.2). The subsequent sections investigate mediating factors and provide robustness analyses.

5.1 Main results

Table 2 shows our main results on the effect of ethnic concentration on the host country language proficiency of immigrant children. The results indicate that an increase in co-ethnic concentration significantly reduces immigrant children's proficiency to speak and write German. An increase in the size of the own ethnicity by one log-point is related to a decline in speaking skills by 19% and in writing skills by 17% of a standard deviation. The magnitudes of the estimated coefficients barely change when we include controls for children's and parents' characteristics.

²⁵ Additional supporting evidence for our identification strategy comes from the fact that when we use the natural logarithm of the sum of all guest workers from *other* ethnicities except for the largest one as treatment variable, we do not find significant associations with any of our outcome measures (results available upon request).

To facilitate interpretation of magnitudes, ethnic concentration would increase by one log-point, for example, if a Turkish child moved from the city of Bonn (with a share of Turks of about 1%) to the city of Munich (with a share of about 2.8%).²⁶ This change in the region of residence would, *ceteris paribus*, reduce the child's German speaking proficiency by 19% and her writing proficiency by 17% of a standard deviation, respectively. This is a modest effect, given that the difference between "poor" and "fair" German language proficiency is 1.39 standard deviations for speaking and 1.12 standard deviations for writing.

In line with the negative impact on host-country language proficiency, we also find a negative effect of ethnic concentration on immigrant children's educational attainment (Table 3). Living in an ethnic enclave substantially increases the likelihood of the child to drop out of school without any degree (columns 1 and 2). A one log-point increase in co-ethnic concentration increases the probability of dropping out of school by 5.6 percentage points. Given that the overall dropout rate among immigrant children in our sample is only 7.1%, this is a huge effect. There are several reasons for the impact on school dropout to be larger than on children's German language proficiency. High German language skills are a prerequisite to be able to understand the instructions of teachers (which requires verbal proficiency) as well as to pass the written exams (which requires writing proficiency) in all school subjects. Hence, the two negative effects on speaking proficiency and writing proficiency likely combine and accumulate across all subjects taught in school, leading to worse grades in multiple subjects. This, in turn, may lead to the large negative impact on school dropout. While results also point toward a negative impact on the probability of obtaining at least an intermediate school degree, the coefficient is much less precisely estimated and becomes zero when controlling for child and parent characteristics (columns 3 and 4).²⁷

The negative effects on host country language proficiency and on obtaining any school degree suggest that immigrant children who grew up in regions with high shares of (low-educated) co-ethnics suffer long-term disadvantages in human capital acquisition. The joint effects on language proficiency and educational attainment are in line with the existing evidence of complementarities between language and other forms of human capital (e.g., Chiswick and Miller 2002).

5.2 Subgroup analyses

Beyond the average effects, we investigate effect heterogeneity by country of birth, gender, and ethnicity.²⁸ We start by investigating whether the negative effects of

²⁶ An increase in the size of the ethnic community by one log-point corresponds to an increase by 172%. The difference in average ethnic concentration between low ethnic concentration and high ethnic concentration regions is 1.19 log-points.

²⁷ Similarly, there is no evidence for a significant effect of ethnic concentration on obtaining an advanced school degree (*Abitur*) (not shown).

²⁸ Analysis of effect heterogeneity by parents' educational background does not produce significant differences.

ethnic concentration on children's outcomes differ between children born abroad and children born in Germany. About 42% of the immigrant children in our sample were born abroad, entering Germany through a family reunification scheme. The first two columns of Table 4 suggest that the negative enclave effects on German speaking and writing proficiencies are roughly 30 percent smaller for children who were born in Germany rather than abroad. As children born in Germany start learning the German language already in nursery school and school, co-ethnic concentration may be less important for them compared to children born abroad who typically start learning the German language at an older age. Still, the ethnic-concentration impact is also significant for guest worker children who were born in Germany. Furthermore, the smaller impact on language proficiency does not translate into a smaller disadvantage of children born in Germany in terms of dropping out of school (column 3).

The right panel of Table 4 investigates effect heterogeneity by child gender. Results indicate that the impact of ethnic concentration on children's language proficiency and educational attainment does not differ significantly between boys and girls, although the negative effect on school dropout may be slightly smaller (in absolute terms) for girls.

Subgroup analyses by ethnicity indicate little heterogeneity (Appendix Table 15). Results suggest that the effect of ethnic concentration on German speaking and writing proficiencies and on school dropout does not differ significantly for Greek, Italian, Spanish, Turkish, or Yugoslav guest worker children. There is some indication, however, that ethnic concentration may have a stronger negative effect on the probability of obtaining at least an intermediate school degree for Italian and Turkish children, and a more positive one for Greek and Spanish children.

6 Mediating factors

The effect of ethnic enclaves on immigrant children's outcomes may be mediated through numerous different channels, including parents' language skills, inter-ethnic contacts with natives, and economic conditions. Existing studies that rely on administrative data are usually restricted to looking at the enclave effect as a black box. By contrast, the rich SOEP survey data allow us to investigate several potential mediating factors at the child and parent level.

6.1 Parental proficiency in the host-country language

A first candidate for a mediating factor is parents' host country language skills, as children's human capital accumulation may critically depend on the language proficiency of their parents (e.g., van Ours and Veenman 2003). In fact, Danzer and Yaman (2016) find a strong negative effect of ethnic enclaves on the language skills of first-generation guest workers in Germany. In the SOEP, adult guest workers (i.e., the parents of our children) report their own German language proficiency in speaking and writing. Using the same random effects specification (without child controls) and the same definitions for language proficiency and ethnic concentration as

Table 2 Effect of ethnic concentration on host-country language proficiency

	Speaking proficiency		Writing proficiency	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.189** (0.083)	-0.185** (0.081)	-0.167** (0.075)	-0.173** (0.069)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes
Child characteristics	No	Yes	No	Yes
Parent characteristics	No	Yes	No	Yes
Observations	4932	4932	4922	4922
R ² overall	0.180	0.270	0.188	0.293

Random effects model. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. Child characteristics: dummies for birth cohort (2-year intervals), gender, and age at migration. Parent characteristics: the following variables for father and mother, respectively: year of birth and dummies for arrival cohort (2-year intervals), schooling in country of origin (incomplete compulsory schooling and at least compulsory schooling), years of education in 1985, migrant status, and number of mother's children. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Table 3 Effect of ethnic concentration on educational attainment

	Any school degree		At least intermediate school degree	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.072*** (0.019)	-0.056*** (0.021)	-0.059 (0.052)	0.002 (0.049)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Child characteristics	No	Yes	No	Yes
Parent characteristics	No	Yes	No	Yes
Observations	1005	1005	1005	1005
Adjusted R ²	0.033	0.051	0.086	0.211

OLS regressions. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

in our main model, we find an effect of ethnic enclaves on the speaking proficiency of parents of -0.351 (standard error 0.081), but no significant effect on parents' writing proficiency (-0.072, standard error 0.091).

Table 4 Subgroup analysis

	Country of birth				Child gender			
	Speaking proficiency (1)	Writing proficiency (2)	Any School degree (3)	At least Intermediate degree (4)	Speaking proficiency (5)	Writing proficiency (6)	Any School degree (7)	At least Intermediate degree (8)
Size of ethnic group in 1985	-0.211** (0.082)	-0.200*** (0.070)	-0.060** (0.024)	-0.022 (0.051)	-0.174** (0.082)	-0.154** (0.074)	-0.064*** (0.021)	0.009 (0.051)
Size of ethnic group * born in Germany	0.062** (0.028)	0.063** (0.026)	0.009 (0.018)	0.049** (0.021)				
Size of ethnic group * female					-0.027 (0.031)	-0.049 (0.038)	0.019* (0.010)	-0.016 (0.027)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a	n.a	Yes	Yes	n.a	n.a
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005	4932	4922	1005	1005
R ² overall	0.271	0.295	0.050	0.215	0.271	0.295	0.052	0.210
Adjusted R ²								

Columns 1–2 and 5–6: Random effects model. Columns 3–4 and 7–8: OLS regressions. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

In a standard descriptive analysis of potential mechanisms, we add different potential mediating factors as control variables to our main models.²⁹ As indicated in column 2 of Table 5, parents' German speaking proficiency is significantly positively related to their children's German speaking proficiency.³⁰ Controlling for parents' German speaking proficiency reduces the effect of ethnic concentration and renders it statistically insignificant, although the negative point estimate remains quite sizeable. However, self-assessed language proficiency is likely measured with error. To circumvent downward bias in the estimated effect of parents' language proficiency, we follow the approach of Dustmann and van Soest (2002) and exploit the panel dimension of the SOEP to instrument parents' speaking proficiency reported in a given year with their speaking proficiency reported in preceding (lag) and subsequent (lead) years.³¹

When accounting for random measurement error in this instrumental variable (IV) approach, parents' German speaking proficiency can fully account for the effect of ethnic concentration on children's speaking proficiency. The IV estimate on parents' speaking proficiency (column 3) is three times as large as the OLS estimate, indicating that the latter suffers from substantial attenuation bias.³² Intriguingly, once the independent-over-time measurement error is accounted for, the point estimate of the effect of ethnic concentration on guest worker children's German speaking proficiency is reduced to close to zero. This suggests that poor parental host country language skills in ethnic enclaves are a main driver of the enclave effect on children's host-country language proficiency.

Columns 4 and 5 present equivalent analyses for parents' writing proficiency in German. While parents' German writing skills are also significantly related to their

²⁹ Causal interpretation of the mechanisms rests on two key assumptions of sequential ignorability (Imai et al. 2010, 2011). The first one is the standard exogeneity assumption: the treatment (ethnic concentration) is assumed to be ignorable given the pre-treatment covariates; that is, it is statistically independent of potential outcomes and potential mediators. The second assumption is that the mediator variable is ignorable given ethnic concentration and pre-treatment covariates. This implies that conditional on ethnic concentration and pre-treatment covariates, the mediator is statistically independent of potential outcomes. This assumption would be violated, for example, if parents' underlying attitude toward education affected both their own language proficiency and their children's language proficiency. When we implement the nonparametric version of the causal mediating analysis proposed by Imai et al. (2010) using the Stata module "mediation," we obtain the same qualitative results as in Tables 5, 6, and 7. For example, parents' speaking proficiency accounts for about 80% of the impact of ethnic concentration on children's language proficiency, whereas all other mediating factors explain less than 15% (not shown).

³⁰ Missing data on the self-reported language proficiency of parents reduce the sample size by 16%, but this does not qualitatively affect the estimate of our main effect (see column 1).

³¹ If one of the two instruments is missing, the missing value is imputed with the other instrument. We add an imputation dummy taking on the value of one for observations with imputed values, zero otherwise. The same applies to parents' writing proficiency. Excluding observations with imputed leads and lags of parents' language proficiency yields qualitatively similar results. Note that the IV approach solves the issue of idiosyncratic (i.e., year-specific) measurement error but does not address the issue that immigrants may systematically over- or underrate their host country language proficiency (Dustmann and van Soest 2002).

³² The *F*-statistic of the excluded instruments in the first stage is 40.7, and it is at least 10 in all following IV specifications (results available upon request).

Table 5 Mediating factors—effect of ethnic concentration on host-country speaking proficiency

	Mediating factors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.178** (0.085)	-0.123 (0.084)	-0.007 (0.092)	-0.173** (0.086)	-0.135 (0.111)	-0.181** (0.083)	-0.169** (0.084)	-0.182** (0.085)
Speaking abilities, parents		0.165*** (0.019)						
Speaking abilities, parents, IV lead+ lag			0.531*** (0.106)					
Writing abilities, parents				0.073*** (0.024)				
Writing abilities, parents, IV lead+ lag					0.635*** (0.244)			
First friend German						0.226*** (0.059)		
Visits from Germans, parents							0.089** (0.045)	
Unemployed (1984–1986), parents								-0.685** (0.279)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	4125	4125	4125	4125	4125	4125	4125	4125
R ² overall	0.270	0.291	0.271	0.275	0.209	0.281	0.273	0.276

Random effects model. Columns 3 and 5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

children's German speaking proficiency, controlling for them does not reduce the estimated effect of ethnic concentration by much.

Table 6 shows the same analyses for children's writing rather than speaking proficiency. We find similar associations of parents' German language proficiency with their children's writing proficiency as we found for children's speaking proficiency. Intriguingly, it is again only parents' *speaking* proficiency (column 3), rather than their writing proficiency (column 5), that reduces the estimated enclave effect on children's *writing* proficiency to close to zero. Thus, it appears that reduced speaking proficiency in the host-country language (and therefore likely reduced speaking of the host-country language at home), rather than limited writing proficiency in the host-country language, is a leading mechanism by which ethnic enclaves inhibit the language proficiency of immigrant children. One potential reason for the greater importance of parents' speaking proficiency is the oral communication of guest workers and their children at home.³³

6.2 Inter-ethnic contacts with natives and economic conditions

Limited contacts to German natives may constitute a further mediating factor of the negative effect of co-ethnic concentration on children's host country language proficiency. Prior research shows that guest workers in Germany who were placed in ethnic enclaves tend to interact less with natives (Danzer and Yaman 2013), and reduced contact with natives may in turn affect the human capital acquisition of their children. As columns 6 and 7 of Tables 5 and 6 show, having personal contacts with natives—either measured by whether the child's first friend is German or whether parents regularly receive visits from Germans—is indeed significantly positively associated with the child's German speaking and writing proficiencies.³⁴ Yet, controlling for the reduced contacts with natives does not significantly change the negative estimate of ethnic enclaves on children's host-country language skills.

Differences in economic conditions such as parental unemployment or household income might also explain the negative effect of ethnic enclaves on immigrant children's language proficiency. As column 8 of Tables 5 and 6 shows, parents' unemployment status is significantly associated with their children's host country language proficiency in the expected way, but controlling for parental unemployment and household income does not affect the estimated effect of ethnic concentration on children's language proficiency at all.

³³ Furthermore, linguists point out that while speaking requires inter-personal communication, writing is based on structural learning (Blanche and Merino 1989). The lack of language courses for guest workers may thus have suppressed the development of writing skills more than of speaking skills. As guest worker parents not only have a lower average level of writing ability, but also lack the competency to self-assess this ability, the self-assessed measure of writing skills may suffer from greater measurement error (Danzer and Yaman 2016).

³⁴ The two respective SOEP questions read as follows: "What is the nationality of the first person befriended?" [German national, other national] (answered by the children) and "Have you received German visitors in your home in the last 12 months?" [yes, no] (answered by the parents).

Table 6 Mediating factors—effect of ethnic concentration on host-country writing proficiency

	Baseline		Mediating factors					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.134* (0.079)	-0.095 (0.078)	0.033 (0.095)	-0.127 (0.080)	-0.090 (0.105)	-0.138* (0.077)	-0.126 (0.079)	-0.138* (0.078)
Speaking abilities, parents		0.121*** (0.021)						
Speaking abilities, parents, IV lead+lag			0.574*** (0.120)					
Writing abilities, parents				0.103*** (0.024)				
Writing abilities, parents, IV lead+lag					0.680** (0.289)			
First friend German						0.252*** (0.054)	0.082* (0.046)	
Visits from Germans, parents								
Unemployed (1984–1986), parents								-0.545* (0.281)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	4120	4120	4120	4120	4120	4120	4120	4120
R ² overall	0.296	0.310	0.269	0.303	0.226	0.308	0.299	0.299

Random effects model. Columns 3 and 5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Table 7 shows that none of the mediating factors analyzed here can account for the effect of ethnic enclaves on children's schooling outcomes. Parents' speaking ability is the only factor that is significantly associated with their children's probability to obtain a school degree. Still, controlling for parents' speaking ability does not reduce the estimated effect of ethnic concentration on whether children obtain a school degree.³⁵

While parents' German language proficiency mediates the effect of ethnic enclaves on children's language proficiency, it does not account for the effect on children's school dropout. This indicates that parents' German language proficiency differs in its importance for these outcomes. In line with prior evidence on the intergenerational transmission of language skills (e.g., Bleakly and Chin 2008; Casey and Dustmann 2008), parents' (instrumented) German speaking and writing proficiencies are strongly correlated with their children's German language proficiency (columns 3 and 5 in Tables 5 and 6). In contrast, parents' German language proficiency is only weakly correlated with the probability that their child leaves school without any degree (Table 7). This weaker correlation implies a much lower potential to mediate the effect of ethnic enclaves. A potential explanation for this finding is that children's host-country language proficiency is strongly affected by their parents' host-country language skills due to frequent interactions at home. This is particularly true for immigrant children who often do not attend kindergarten, which provide opportunities to learn the host country language. While language skills can be strongly mediated in the parental home, obtaining a school degree depends on many factors outside the family, such as the quality of school peers or teachers, which affect, among others, children's motivation to learn. Also, school dropout rates in Germany in the observation period were modest (5%), possibly reducing the scope of parental language skills as a mediating factor.

Some additional potential mediating factors are hard to explicitly address in our setup. For example, some schools may have fewer and lower-achieving native Germans due to "native flight" at the disaggregated level of school catchment areas (Cascio and Lewis 2012; De la Croix and Doepke 2009).³⁶ In most parts of Germany, children normally attend the only primary school of their residential school district. The extent of school segregation at the primary school level is, hence,

³⁵ Similar analyses for obtaining at least an intermediate school degree as the child outcome do not indicate any significant enclave effects; only instrumented parental writing abilities and having a German as the first friend are significantly associated with this outcome (not shown).

³⁶ Note that there are no historical segregation measures at such disaggregated levels (see, e.g., Alesina, Murard, and Rapoport, 2021). Segregation implies that natives and immigrants are separated in daily interaction, either in residential neighborhoods or at school. While we have no school-level data to study school segregation directly, we investigate the potential issue of native flight at the regional level by checking whether the impact of ethnic concentration systematically varies with the size of the native population. We interact ethnic concentration with a dummy variable indicating a low share of the native population in the region (below the median share of the native population in 1985). Significant negative interaction terms would imply that child outcomes are *more* negative in areas characterized by higher ethnic concentrations in combination with fewer contact possibilities to natives. In the regressions with the four child outcomes, only one out of the four interaction terms is marginally significant (results available upon request), suggesting that native flight is unlikely to be a main mechanism.

Table 7 Mediating factors—effect of ethnic concentration on obtaining any school degree

	Baseline		Mediating factors					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.062*** (0.022)	-0.062*** (0.022)	-0.065*** (0.020)	-0.062*** (0.022)	-0.064*** (0.020)	-0.056** (0.022)	-0.062*** (0.022)	-0.065*** (0.022)
Speaking abilities, parents		0.002 (0.010)						
Speaking abilities, parents, IV lead + lag			0.042** (0.019)					
Writing abilities, parents				0.002 (0.012)				
Writing abilities, parents, IV lead + lag					0.007 (0.022)			
First friend German						0.027 (0.019)		
Visits from Germans, parents							0.007 (0.032)	
Unemployed (1984–1986), parents								-0.072 (0.095)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household income	No	No	No	No	No	No	No	Yes
Observations	943	943	943	943	943	943	943	943
Adjusted R ²	0.057	0.056	0.045	0.056	0.056	0.066	0.055	0.059

OLS regressions. Columns 3 and 5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

mostly determined by residential patterns, since school choice is limited unlike, for instance, in the USA, UK, or Netherlands (Kristen 2005). Private primary schools play a negligible role with only 1.3% of enrolled primary school students (Statistisches Bundesamt 2002). School segregation in secondary education is mostly determined by early ability tracking in Germany.³⁷

Another potential mediating factor is the existence of network effects combined with self-selection that becomes increasingly negative over time. Migration networks reduce migration costs and will normally increase a low-skilled person's propensity to move, find a job, or find accommodation (McKenzie and Rapoport 2010). In our setting, networks did not play a role in the initial placement of guest workers who received jobs and accommodation upon arrival and were unable to influence their placement. Networks may have influenced individual decisions to sign up for migration in the first place, not least since educational attainment in the home countries increased somewhat during the 1950s to 1970s (Lee and Lee 2016). Post-migration personal networks may have influenced the propensity of relocation after the end of the lock-in period, potentially influencing the incentives of parents to learn German, the nature of native flight, and the quality of children's peers at school.

In sum, the negative effect of ethnic enclaves on immigrant children's host country language proficiency can be fully accounted for by parents' lower host-country speaking proficiency. Parents' writing proficiency accounts for only a small part of the negative enclave effect. Limited contacts to natives and economic factors do not appear to be relevant mediating factors of the negative enclave effects. None of the investigated mediating factors—parents' language skills, inter-ethnic contact, and economic conditions—can account for the detrimental effect of ethnic enclaves on the schooling success of immigrant children.³⁸

7 Robustness analyses

In this section, we show that our results are robust to measuring ethnic concentration by ethnic shares (Section 7.1), instrumenting ethnic concentration in 1985 by its 1975 value (Section 7.2), measuring ethnic concentration with Census data (Section 7.3), measuring ethnic concentration at the county level (Section 7.4), and accounting for the interview mode (Section 7.5). We also investigate return

³⁷ Given a certain level of ethnic concentration, diversity in schools or residential segregation could be another potential mechanism. For example, previous research has shown that higher neighborhood diversity is associated with lower employment probability, which may directly or indirectly affect schooling outcomes of children (Hémet and Malgouyres 2018). However, ethnic enclaves of guest workers in Germany are associated with *more* employment (Table 1). If higher immigrant shares are associated with more diversity, then neighborhood diversity will tend to be higher in regions with high ethnic concentration. Since we find that higher ethnic concentration leads to worse schooling outcomes of children, this suggests that our estimates might be conservative. Furthermore, while Frattini and Meschi (2019) find a negative effect on math scores in classrooms with a high immigration share, they show that ethnic diversity in school does not play a significant role.

³⁸ This result is robust to including all mediating factors in the regression model simultaneously (not shown).

migration (Section 7.6), regional migration within Germany (Section 7.7), and family size (Section 7.8).

7.1 Measuring ethnic concentration by ethnic shares

There is no strong a priori argument for any specific functional form of the ethnic concentration measure. At least two different specific measures of ethnic concentration have been used in the literature. In our analyses so far, we followed Edin et al. (2003), Damm (2009), and Åslund et al. (2011) in using the logarithm of the size of the own ethnicity. In contrast, Chiswick (2009) and Danzer and Yaman (2013, 2016) measure ethnic concentration as the share of the own ethnicity in the total regional population.

When using the share of the own ethnicity in the regional population as an alternative measure of ethnic concentration, results on guest worker children's German speaking and writing proficiencies and on school dropout are qualitatively similar to our main models (Table 8). Interestingly, the alternative concentration measure also produces significant results on the probability that guest worker children obtain at least an intermediate school degree. Specifically, the point estimate suggests that a one percentage point increase in the share of own ethnics in the regional population reduces the likelihood of obtaining at least an intermediate school degree by 5.1%.

7.2 Instrumenting ethnic concentration in 1985 by ethnic concentration in 1975

In another robustness check, we restrict the variation in our treatment variable to variation in ethnic concentration that was given already towards the end of the guest worker program. The measure of ethnic concentration in our main analysis refers to 1985, the year in which we first observe guest workers and their region of residence (see Section 4.2). While the balancing tests indicate no evidence of self-selection of guest workers across regions with different ethnic concentrations, the extent of ethnic concentration may have changed between the end of the German guest worker program in 1973 and the observed ethnic concentration in 1985. To account for potential endogeneity of our main explanatory variable, we can instrument a region's ethnic concentration in 1985 by the region's ethnic concentration in 1975, i.e., towards the end of the German guest worker recruitment program (Danzer and Yaman 2013). 1975 is the first year of the SIAB data. This IV model rules out any bias from changes in ethnic concentrations in a given region during the decade before we first observe guest workers' region of residence. For instance, if economic conditions improved between immigration and 1985 in the initial placement region of a guest worker relative to other regions, we may expect an increase in ethnic concentration in this region owing to economically motivated in-migration. In this case, the ethnic concentration observed in 1985 differs from the one at quasi-exogenous placement.

Ethnic concentration in 1975 is a very strong instrument for ethnic concentration in 1985. The F statistic on the excluded instrument in the first stage is 236 in the regressions for language outcomes and 321 in the regressions for schooling

outcomes.³⁹ In line with Schönwalder and Söhn (2009), this suggests that there is strong persistence in the settlement structures of guest workers between the end of the guest worker program and 1985.⁴⁰

Table 9 presents the results of the IV model that uses only that part of the variation in ethnic concentration in 1985 that can be traced back to variation in ethnic concentration that already existed in 1975. For both speaking and writing proficiencies, the enclave effect is somewhat stronger when instrumenting 1985 with 1975 ethnic concentration compared to the baseline model. The effect on school dropout does not change, and the coefficient for obtaining at least an intermediate school degree remains insignificant. Similarly, all results on mediating factors are very similar in the IV model compared to the baseline model (not shown).

In sum, our baseline estimates are not biased by any change in ethnic concentration that occurred between 1975 and 1985. If anything, restricting the analysis to variation in ethnic concentration that already existed in 1975 leads to slightly larger estimates of the detrimental effect of ethnic enclaves on immigrant children's outcomes.

7.3 Measuring ethnic concentration with census data

Measuring the size of the immigrant population based on a 2% random sample of employees like the SIAB can lead to attenuation bias in estimating effects of immigration measures (Aydemir and Borjas 2011). To address potential measurement error in our preferred measure of ethnic concentration, we use data from the 1987 German Census, which includes the entire population in Germany. As the 1987 Census data do not allow identifying Spanish citizens, the Census analysis is restricted to the other four ethnicities. For each ethnicity, the correlation coefficient between our preferred 1985 SIAB measure and the 1987 Census measures of the (log) size of the ethnic community exceeds 0.96.

Replacing the 1985 SIAB measure of ethnic concentration with the 1987 Census measure yields very similar results to our main specifications (results available upon request). Furthermore, IV models that instrument the 1987 Census measure of ethnic concentration with the concentration of guest workers in the mid-1970s using the SIAB 1975 data—which simultaneously account for measurement error and changes in regional ethnic concentration after the end of the guest worker program—are also quite similar to the baseline results. Again, the IV estimates are somewhat larger than the non-instrumented estimates. The results on mediating factors are also unaffected when using the 1987 Census data to compute measures of ethnic concentration, both in the non-instrumented and in the instrumented models

³⁹ The first-stage coefficient on the size of the ethnic community in 1975 is 0.85 ($p=0.000$) in the language sample and 0.84 ($p=0.000$) in the schooling sample.

⁴⁰ In fact, there is data on the *placement* of guest workers by sending country at the broad regional level of *Landesarbeitsamtsbezirke* (roughly equivalent to the *Länder*) for 1 year, 1969. A regression of the log stock of guest workers in 1985 (in logs) on the log placement count in 1969 (omitting Berlin) yields an R^2 of 0.95.

(not shown). In sum, we do not find evidence that measurement error in our ethnic concentration measure has a substantial effect on our results.

7.4 Measuring ethnic concentration at the county level

Our preferred regional level for measuring ethnic concentration are the areas (*Anpassungsschichten*), as they comprise sufficiently large regions in order to circumvent bias from commuting within regional labor markets. While the much smaller regional entity of counties may more precisely measure immigrant children's exposure to co-ethnics, they also increase concerns of bias due to commuting and moving across county borders. Still, using the 1987 Census, which includes the entire population, we can test for robustness of our results to measuring ethnic concentration at the level of 328 counties rather than 103 areas.⁴¹

When measuring ethnic concentration and including fixed effects at the county level, the effects of ethnic concentration on children's speaking and writing proficiencies are very similar to the estimates when measuring ethnic concentration at the area level (results available upon request). By contrast, the effect on obtaining any school degree becomes smaller and loses statistical significance. Besides the fact that Spanish guest worker children are missing in the analysis, statistical power in the county-level analysis may be impaired by the fact that enclave effects are identified from fewer guest worker children observed within the same region in the SOEP data. This likely affects in particular the analysis of school dropout, which on average is already rather low (7.1%). In fact, incidents of school dropout by guest worker children are observed in only 42 of the 114 counties with guest worker children in the SOEP. This suggests that models with county fixed effects exploit only very limited variation in school dropout.

7.5 Accounting for interview mode

We also show that immigrants' self-reported language proficiency is not affected by the specific interview mode used in the SOEP, such as oral face-to-face interview or written interview by mail. Adding a control for the interview mode used when guest worker children report their levels of German language proficiency does not affect the estimated enclave effects on children's proficiency in speaking or writing German (results available upon request).

7.6 Investigating return migration

Acquiring host country language skills and education is an investment decision that may depend on whether immigrants intend to stay in the host country or return to their home country (Dustmann and Glitz 2011). To account for this possibility, we

⁴¹ The guest worker children observed in the SOEP data live in only 114 of the 328 counties, limiting the variation used in the analysis.

can include a binary indicator that equals 1 if guest worker parents see their future in Germany (0 otherwise).⁴² Adding this control variable does not affect our baseline estimates (results available upon request). Parents' intention to stay in Germany is positively associated with the children's outcomes, albeit statistically significantly only in the case of obtaining a school degree.

We also investigate to what extent immigrant children in our sample actually returned to their home country during the first 10 years of our analysis, and in particular, whether return migration was related to ethnic concentration. Following Dustmann (2003), we use the information that individuals provide when leaving the SOEP survey. We construct a binary indicator, $return_{icr}$, which equals 1 if an immigrant child left the SOEP between 1985 and 1995 reporting a move abroad, and 0 otherwise.⁴³ To predict return migration, we use the same explanatory variables as in our main model, in particular ethnic concentration in 1985:

$$return_{icr} = \sigma_c + \delta_r + \gamma_1 EC_{cr} + C'_{icr} \gamma_2 + \varepsilon_{icr} \quad (3)$$

During the observation period 1985–1995, 94 immigrant children (8.8% of our sample) left the SOEP and moved abroad. However, moving abroad is unrelated to ethnic concentration (results available upon request). The result that there is no selective return migration with respect to ethnic concentration strengthens our identifying assumptions. Furthermore, attrition due to return migration likely generates an upward bias if return migrants are negatively selected on host-country skills, which seems likely. This would imply that our results underestimate the true effect of ethnic enclaves.

7.7 Investigating regional migration within Germany

Our analysis requires the validity of the assumption that there was no systematic sorting of guest workers across regions between their initial placement in Germany and 1985, when we observe them for the first time in the SOEP data. While lack of data prevents us from investigating cross-regional mobility before 1985, we can analyze moving patterns after 1985, a period when overall regional mobility was higher. We find that only 6.9% of the 749 immigrant children who remained in the SOEP sample moved across regions between 1985 and 1995. Similarly, the moving rate was modest during the 20-year period between 1985 and 2005 (15.6%). This rather low mobility is consistent with Glitz (2014), who shows that regional ethnic concentrations were stable in Germany between 1975 and 2008.

Most importantly, ethnic concentration in 1985 does not predict whether guest worker families moved across regions over the next 10 years (results available upon request). In sum, the low mobility and the unsystematic moving patterns with

⁴² The respective SOEP question reads as follows: "How long do you want to remain in Germany?" [up to 12 months, a few years, stay in Germany].

⁴³ It is possible that children report other reasons for leaving the SOEP survey and later on return to their home country, which we would not observe.

Table 8 Measuring ethnic concentration by share of own ethnicity in regional population

	Speaking proficiency	Writing proficiency	Any school degree	Intermedi- ate school degree
	(1)	(2)	(3)	(4)
Share of own ethnicity in 1985	-0.080** (0.034)	-0.080*** (0.029)	-0.025** (0.011)	-0.051** (0.020)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a	n.a
Child characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005
R^2 overall	0.269	0.292		
Adjusted R^2			0.051	0.216

Columns 1–2: Random effects model. Columns 3–4: OLS regressions. Share of own ethnicity in 1985: share of own ethnicity in the population of the area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Table 9 Instrumental-variable estimates using ethnic concentration in 1975

	Speaking proficiency	Writing proficiency	Any school degree	Intermedi- ate school degree
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.234** (0.103)	-0.183** (0.075)	-0.056*** (0.019)	0.032 (0.049)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	n.a	n.a
Child characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
Observations	4932	4922	1005	1005
R^2 overall	0.269	0.293		
Adjusted R^2			0.051	0.210
First-stage F-statistic	236.63	236.29	321.31	321.31

Columns 1–2: Random effects model. Columns 3–4: OLS regressions. Size of ethnic group in 1985 is instrumented by size of ethnic group in 1975 (both variables in logs). Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

respect to ethnic concentration support our identifying assumption that guest workers did not systematically sort across regions before 1985.

7.8 Investigating family size

If ethnic concentration affected the number of children in a household, our analysis sample could be subject to selection. For example, ethnic concentration might affect fathers' labor market success, which in turn might affect their decision to have children or to bring their existing families from their home countries to Germany. However, results show that this is not the case: ethnic concentration is neither related to the probability of having children in the household nor to the number of children living in the household (results available upon request). These results are also consistent with the fact that mothers do not have significantly different numbers of children in 1985 in regions with low and high ethnic concentration (Table 1). Therefore, our analysis seems to be unaffected by endogenous fertility and family reunification.

8 Conclusion

This paper exploits the quasi-exogenous placement of guest workers across Germany during the 1960s and 1970s to estimate the effect of growing up in ethnic enclaves on the language proficiency and educational outcomes of immigrant children. We find that growing up in regions with higher own ethnic concentration significantly reduces immigrant children's proficiency in the host country language and their educational attainment. For schooling outcomes, the effect is concentrated at the lower end of the educational distribution, although there is some indication that more academic school degrees may be affected as well. The enclave effects tend to be larger for immigrant children who were born abroad.

The rich information contained in the German Socio-Economic Panel, most importantly on parents' host country language proficiency, allows us to investigate several factors that might mediate the effect of ethnic concentration on child outcomes. We find that parents' German speaking proficiency can completely account for the negative effect of ethnic enclaves on their children's German language proficiency. Parents' writing abilities account for only little, and contacts to natives and parents' economic conditions cannot account for the negative effect of ethnic enclaves on immigrant children's outcomes at all.

These findings imply that even children of immigrants who are well integrated into the labor market may suffer from worse human capital outcomes—host country language proficiency and educational attainment—when growing up in regions with many, mainly low-educated immigrants of their own ethnicity. The important role of parents' host country language skills in the mediation analysis suggests that host country language training for adult immigrants may have important positive spillover effects on their children. More generally, the results indicate that the long-run cultural and social integration of immigrants, particularly for the next generation, may be more successful when ethnic concentration is limited.

Two caveats of our analysis are worth mentioning. First, our analysis is limited in that we can only observe guest workers' location in 1985, not their initially assigned place of residence. The results therefore rely on the identifying assumption that guest workers did not systematically move across regions between their arrival in the 1960s/1970s and 1985. While we argue—in line with the prior literature—that this assumption is plausible owing to specific features of the German guest worker program, we cannot finally rule out some cross-regional or return migration based on unobservables. Second, concerning external validity, our study results cannot capture all sorts of migration processes. Recruited guest workers had an education level substantially below the native population. Such migratory movements of less skilled workers to relatively richer countries take place around the globe. However, the guest workers had no difficulties in finding employment. Given that workplace integration may support other forms of integration, we expect that findings based on guest workers are more conservative than in settings without employment guarantees.

Appendix

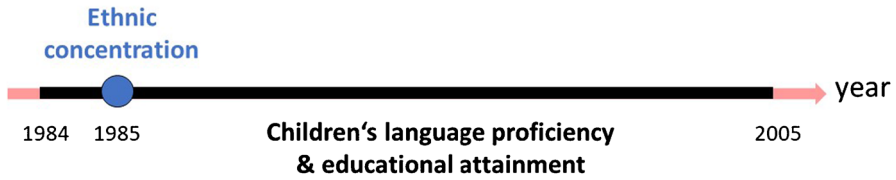


Fig. 2 Timeline of measurements. The educational attainment of very few children (less than 1 percent) was measured between 2005 and 2010

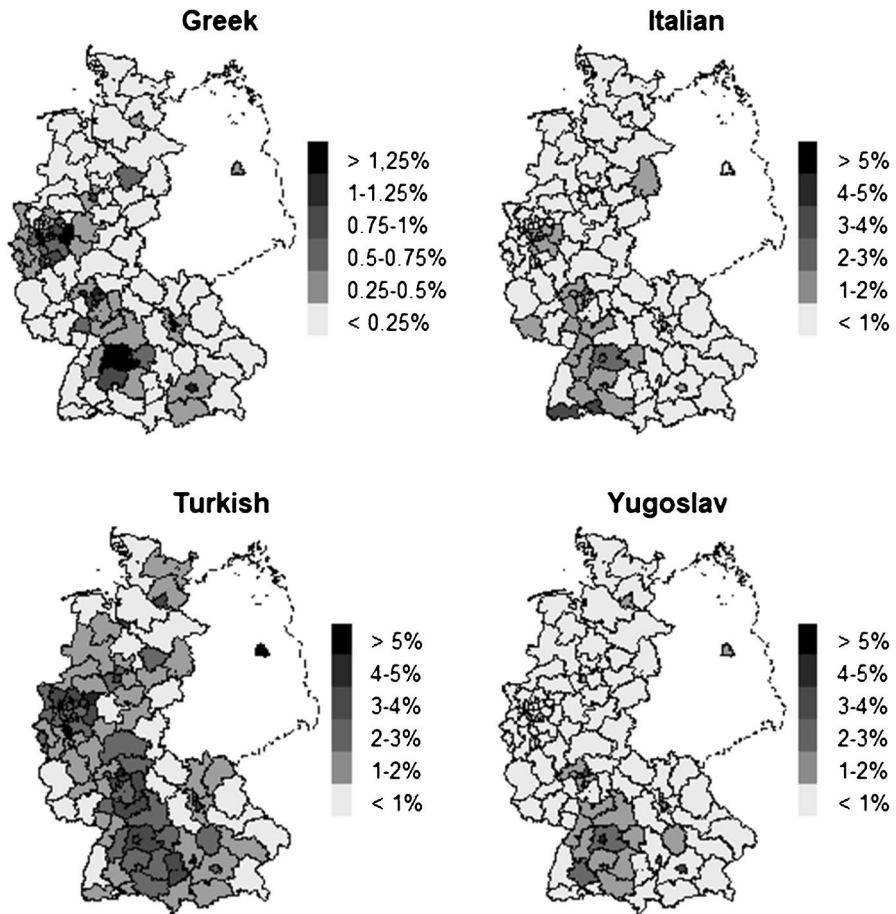


Fig. 3 Ethnic concentrations across West Germany: Census 1987. Ethnic group's share of the total population in the area of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 103 areas. Figures based on a historical GIS data file of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011)

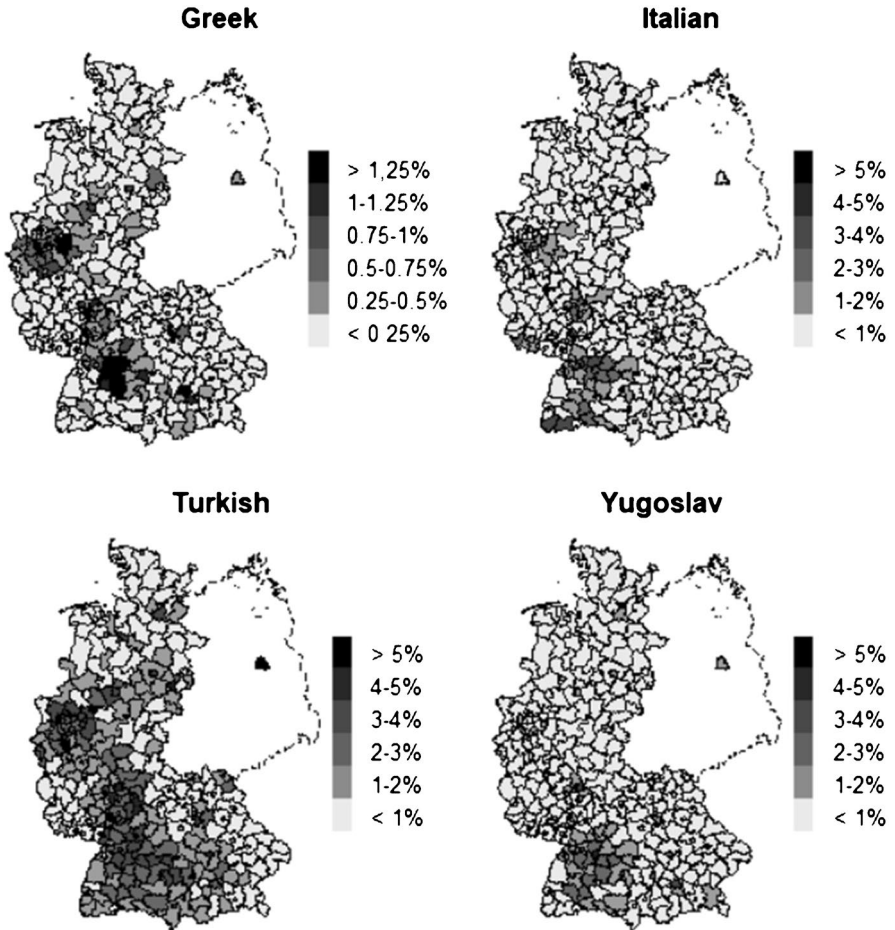


Fig. 4 County-level ethnic concentrations across West Germany: Census 1987. Ethnic group's share of the total population in the county of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 328 counties. Figures based on a historical GIS data file of the Federal Republic of Germany from the Max Planck Institute for Demographic Research and the Chair for Geodesy and Geoinformatics, University of Rostock (2011) and Bundesamt für Kartographie und Geodäsie (2011)

Table 10 Individual-level variables

Variable	Description
<i>Outcomes (Children)</i>	
Speaking proficiency	Generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5), normalized to have mean 0 and standard deviation 1, <i>Random Effects Model</i> : each observation is a child-year observation based on self-reported language proficiency in the years 1984–1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005
Writing proficiency	Generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5), normalized to have mean 0 and standard deviation 1, <i>Random Effects Model</i> : each observation is a child-year observation based on self-reported language proficiency in the years 1984–1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005
Any school degree	Binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise. Based on most recent available educational level. If the most recent available information is <i>dropout, no school degree or no school degree yet</i> , we checked for school-leaving degrees reported in previous years. In nine cases, we adjusted the educational attainment variables based on previously reported school-leaving degrees
At least intermediate school degree	Binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 otherwise. Based on the most recent available educational level. If the most recent available information is <i>dropout, no school degree or no school degree yet</i> , we checked for school-leaving degrees reported in previous years. In nine cases, we adjusted the educational attainment variables based on previously reported school-leaving degrees
<i>Demographics of Children</i>	
Ethnicity dummies (Greek, Italian, Spanish, Turkish, Yugoslav)	Binary indicators primarily based on children's first citizenship (94.2%). In case of German citizenship or no available citizenship information, these indicators are based on parents' joint nationality (1.0%) or on children's country of origin (0.3%). If children's ethnicity is not yet available and one parent is intermarried to a German or to a foreigner with different or missing nationality, we use the citizenship of the parent with the guest-worker background as a proxy for children's ethnicity (4.4%). Regarding rare cases, if children's ethnicity is not available and both parents are migrants but their country of origin differs, we use the mother's nationality or country of origin. In the few cases of remaining missing children's ethnicities, we base children's ethnicity on father's country of origin or nationality. For more than 98.5% of the children in our analysis sample, children's ethnicity corresponds to the father's country of origin. We could assign an ethnicity to all children in our sample
Age at migration	Age at migration (in years). If a child is born in Germany, the variable is coded as zero
<i>Demographics of Parents</i>	
Migrant	Binary indicator that equals 1 if individual has a migrant background and 0 otherwise. Based on variable "migback" of the SOEP Person-related meta-dataset
Variables on education in country of origin	Three binary indicators for <i>No schooling, Incomplete compulsory schooling, and At least compulsory schooling</i> , based on survey question "Obtained School Degree Outside Germany" in survey year 1985
Years of education	Amount of education or training (in years), generated variable by SOEP. Based on survey year 1985
Never moved flat since arrival in Germany	Binary indicator equal to 1 if the individual's year of immigration is either equal to the year in which the household moved into the dwelling or is later than the year in which the household moved; 0 otherwise. Based on survey year 1985
Children	Number of mother's children. Based on variable <i>Bumkids</i> " from the SOEP Birth Biography of Female Respondents
Household income (1984–1986)	Mean of parents' adjusted household income over three years. Based on survey years 1984–1986
Out of the labor force (1984–1986)	Mean of an indicator for being out of the labor force during the survey years 1984–1986
Unemployed (1984–1986)	Mean of an unemployment dummy during the survey years 1984–1986

Table 10 (continued)

Variable	Description
<i>Mediating Factors</i>	
Speaking abilities, parents	Parents' speaking ability, generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported speaking ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available speaking proficiency, largely based on the second survey year of the SOEP in 1985 (99%)
Speaking abilities, parents, IV lead + lag	Parents' speaking ability is instrumented with the corresponding lead and lag to reduce measurement error (see Dustmann and van Soest 2002), missing leads (lags) of parents' current language proficiency are imputed with available lags (leads), all regressions include imputation dummies for imputed leads or lags of parents' language proficiency, generated from self-assessed speaking ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported speaking ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available speaking proficiency, largely based on the second survey year of the SOEP in 1985 (99%)
Writing abilities, parents	Parents' writing ability, generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported writing ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available writing proficiency, largely based on the second survey year of the SOEP in 1985 (99%)
Writing abilities, parents, IV lead + lag	Parents' writing ability is instrumented with the corresponding lead and lag to reduce measurement error (see Dustmann and van Soest 2001), missing leads (lags) of parents' current language proficiency are imputed with available lags (leads), all regressions include imputation dummies for imputed leads or lags of parents' language proficiency, generated from self-assessed writing ability in German (not at all = 1, poorly = 2, fairly = 3, good = 4, very well = 5). Based on the average of self-reported writing ability of the mother and the father, normalized to have a mean 0 and standard deviation 1. For language proficiency estimations: measured at the time of children's reported language proficiency. For educational attainment estimations: measured as parents' second available writing proficiency, largely based on the second survey year of the SOEP in 1985 (99%)
First friend German	Binary indicator equal to 1 if a child's first friend is German. Based on the first available variable on the nationality of the first-named friend. Based on survey years 1988, 1990, 1992, 1994, 1996, 2001, 2006, and 2011
Visits from Germans, parents	Average of the following variable of mother and father: a binary indicator equal to 1 if mother or father received visits from Germans at home during the previous 12 months. For language proficiency estimations: refers to year of children's reported language proficiency. For educational attainment estimations: refers to average of the years 1985 and 1986
Unemployed (1984–1986), parents	Average of the following variable of mother and father: Mean of an unemployment dummy over three years. Based on survey years 1984–1986
Household income (1984–1986), parents	Mean of parents' household income over three years (in logs). Based on survey years 1984–1986
<i>Robustness Checks</i>	
Stay in Germany, parents	Average of the following variable of mother and father: A binary dummy indicating the intent to stay in Germany. Based on the following answer categories: "I intend to stay in Germany forever" (= 1), "I intend to stay in Germany for several years" (= 0), "I intend to leave Germany within 12 months" (= 0). Based on survey year 1985

Table 10 (continued)

Variable	Description
Interview mode	Dummies based on a variable indicating the interview mode in the corresponding survey years of self-reported language proficiency (years 1984–1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005). Based on the following answer categories: "Oral Interview", "Written Questionnaire Interviewer", "Mixed Type", "Written Questionnaire No Interviewer", "Oral And Written", "Proxy", "Third Person Present", "No Third Person Present", "Computer Assisted Personal Interviewing", "Telephone Assistance", "Written, By Mail", and "Telephone Interview"

Source (for all variables): German Socio-Economic Panel (SOEP)

Table 11 Regional variables

Variable	Description
Size of ethnic group in 1985	Log size of ethnic community (individuals of same ethnicity) in area (<i>Anpassungsschicht</i>) of residence, 1985. Log of 1 used in rare case of zero co-ethnics in the region; all regressions include a corresponding imputation dummy. See variable <i>Region of residence</i> for details on the assignment of children to 1985 regions. Based on a two percent sample of the German employee population (incl. recipients of social transfers) from the Institut für Arbeitsmarkt- und Berufsforschung (IAB)
Size of ethnic group in 1987	Log size of ethnic community (individuals of same ethnicity) in region of residence, based on German Census 1987, regional level: area (results available upon request) or county (results available upon request). Measure not available for immigrant children with Spanish ethnicity
Official unemployment rate 1985	Unemployment rate in the year 1985, regional level: area, based on county-level data from Federal Employment Agency (2017)
Region of residence (used to construct ethnic concentration measures)	The region of residence (area or county) is primarily based on children's 1985 region of residence (94.7%). If children's household IDs for the year 1985 are not available, the ethnic concentration measures are based on parents' 1985 region of residence for the following scenarios: children were born after 1985 (2.1%), children had the same household ID as their parents in 1984 (1.5%), children migrated to Germany after 1985 (0.2%), or children joined the SOEP in a later wave than 1985 for other reasons (1.6%). All children in our sample could be assigned to a 1985 region

Table 12 Ethnic concentration by ethnicity

	Mean	SD	Min	Max
Greek	0.95	0.63	0.06	2.25
Italian	1.51	0.95	0.16	4.03
Spanish	0.39	0.24	0.00	0.97
Turkish	3.05	1.20	0.29	6.19
Yugoslav	2.06	1.21	0.48	4.14
Total	2.00	1.39	0.00	6.19

Share of ethnicity in the total population of the area of residence, 1985 (based on full sample of guest-worker children in SOEP). Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Table 13 Effect of ethnic concentration on host-country language proficiency: pooled OLS model

	Speaking proficiency		Writing proficiency	
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.153*	-0.188**	-0.129*	-0.197***
	(0.080)	(0.078)	(0.073)	(0.070)
Region fixed effects	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes
Child characteristics	No	Yes	No	Yes
Parent characteristics	No	Yes	No	Yes
Observations	4932	4932	4922	4922

OLS model. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Year of assessment: dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Table 14 Balancing test using continuous share of ethnic concentration

Variable	Coefficient	Standard Error	P-Value	Obs
<i>Outcomes (Children)</i>				
Speaking proficiency	-0.15	0.04	0.00	996
Writing proficiency	-0.11	0.04	0.00	996
Any school degree	-0.03	0.01	0.00	1005
At least intermediate school degree	-0.06	0.02	0.00	1005
<i>Children</i>				
First year of language assessment	-0.02	0.24	0.92	996
Male	0.02	0.02	0.26	1065
Year of birth	-0.08	0.33	0.80	1065
Age at migration	0.23	0.24	0.34	1065
Born in Germany	-0.01	0.03	0.76	1065
<i>Mothers</i>				
Year of birth	-0.24	0.41	0.56	1065
Year of immigration for the foreign born	-0.33	0.30	0.27	1022
Age at migration	-0.19	0.48	0.70	1022
Born in Germany	-0.02	0.01	0.20	1065
Migrant	0.01	0.01	0.18	1065
No schooling	-0.02	0.03	0.59	1065
Incomplete compulsory schooling	0.01	0.03	0.86	1065
At least compulsory schooling	0.01	0.04	0.80	1065
Years of education	0.06	0.09	0.48	1065
Never moved flat since arrival in Germany	-0.01	0.02	0.39	1065
Children	0.05	0.13	0.72	1065
Out of the labor force (1984–1986)	0.00	0.04	0.91	1065
Unemployed (1984–1986)	-0.00	0.01	0.75	1065
<i>Fathers</i>				
Year of birth	0.37	0.48	0.45	1065
Year of immigration for the foreign born	0.11	0.22	0.63	1056
Age at migration	-0.27	0.36	0.46	1056
Born in Germany	0.00	0.00	0.53	1065
Migrant	-0.00	0.00	0.53	1065
No schooling	0.05	0.01	0.00	1065
Incomplete compulsory schooling	0.00	0.04	0.92	1065
At least compulsory schooling	-0.05	0.04	0.15	1065
Years of education	-0.16	0.14	0.23	1065
Never moved flat since arrival in Germany	-0.00	0.01	0.76	1065
Out of the labor force (1984–1986)	0.02	0.01	0.10	1065
Unemployed (1984–1986)	-0.01	0.01	0.36	1065
Household income (1984–1986)	-15.04	36.89	0.68	1065
<i>For Comparison</i>				
Information on language proficiency available	0.00	0.01	0.86	1065
Information on school degree available	-0.00	0.01	0.71	1065

Table 14 (continued)

Each row reports the results of a separate regression where the respective variable is regressed on the continuous share of ethnic concentration. All regressions include region and ethnicity fixed effects. Speaking/writing proficiency: first reported self-assessed speaking/writing ability in German (from 1 = “not at all” to 5 = “very well”), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Household income, out of the labor force, and unemployed refer to 3-year means over 1984–1986. Information on language proficiency/school degree available: 1 if information on corresponding outcome is available in the SOEP data in at least one survey year, 0 otherwise. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Federal Employment Agency (2017)

Table 15 Subgroup analysis by ethnicity

Interacted ethnicity	Greek	Italian	Spanish	Turkish	Yugoslav
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: speaking proficiency</i>					
Size of ethnic group in 1985	-0.176** (0.079)	-0.184** (0.081)	-0.200** (0.082)	-0.204** (0.085)	-0.167** (0.077)
Size of ethnic group * ethnicity	-0.032 (0.063)	-0.005 (0.077)	0.122 (0.113)	0.109 (0.110)	-0.074 (0.081)
<i>Panel B: writing proficiency</i>					
Size of ethnic group in 1985	-0.177** (0.070)	-0.173** (0.073)	-0.187*** (0.072)	-0.187** (0.075)	-0.151** (0.063)
Size of ethnic group * ethnicity	0.015 (0.063)	-0.002 (0.084)	0.108 (0.114)	0.076 (0.088)	-0.094 (0.083)
<i>Panel C: any school degree</i>					
Size of ethnic group in 1985	-0.051** (0.022)	-0.055*** (0.021)	-0.059*** (0.021)	-0.055** (0.022)	-0.060*** (0.021)
Size of ethnic group * ethnicity	-0.018 (0.024)	-0.007 (0.025)	0.030 (0.026)	-0.002 (0.020)	0.016 (0.026)
<i>Panel D: at least intermediate school degree</i>					
Size of ethnic group in 1985	-0.048 (0.044)	0.021 (0.048)	-0.017 (0.050)	0.023 (0.046)	0.011 (0.054)
Size of ethnic group * ethnicity	0.177*** (0.031)	-0.115*** (0.042)	0.157** (0.076)	-0.125*** (0.042)	-0.035 (0.041)
Region fixed effects	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes

Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in area of residence, 1985. Panels A and B additionally include dummies for year of language assessment. See notes of Table 2 for lists of child and parent characteristics. Standard errors clustered at the region-ethnicity level in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Acknowledgements Feuerbaum acknowledges funding through the International Doctoral Program Evidence-Based Economics of the Elite Network of Bavaria. Piopiunik acknowledges funding through the Leibniz Association (SAW-2012-ifo-3). The contribution by Woessmann is part of the Deutsche Forschungsgemeinschaft project CRC TRR 190.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alesina A, Murard E, Rapoport H (2021) Immigration and preferences for redistribution in Europe. *J Econ Geogr* 21(6):925–954
- Andersson M, Larsson JP, Öner Ö (2021) Ethnic enclaves and self-employment among Middle Eastern immigrants in Sweden: ethnic capital or enclave size? *Reg Stud* 55(4):590–604
- Åslund O, Edin P-A, Fredriksson P, Grönqvist H (2011) Peers, neighborhoods, and immigrant student achievement: evidence from a placement policy. *Am Econ J Appl Econ* 3(2):67–95
- Aydemir A, Borjas GJ (2011) Attenuation bias in measuring the wage impact of immigration. *J Law Econ* 29(1):69–112
- Battisti M, Peri G, Romiti A (2016) Dynamic effects of co-ethnic networks on immigrants' economic success. NBER Working Paper 22389, National Bureau of Economic Research, Cambridge, MA
- Bauer T, Epstein GS, Gang IN (2005) Enclaves, language, and the location choice of migrants. *J Popul Econ* 18(4):649–662
- Blanche P, Merino BJ (1989) Self-assessment of foreign-language skills: implications for teachers and researchers. *Language Learning. A Journal of Research in Language Studies* 39(3):313–338
- Bleakly H, Chin A (2008) What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *Journal of Human Resources* 43(2):267–298
- Bundesamt für Kartographie und Geodäsie (2011) VG 2500 Verwaltungsgebiete (Ebenen) 1:2.500.000. Stand 01.01.2009. Frankfurt am Main
- Cascio EU, Lewis EG (2012) Cracks in the melting pot: immigration, school choice, and segregation. *Am Econ J Econ Pol* 4(3):91–117
- Casey T, Dustmann C (2008) Intergenerational transmission of language capital and economic outcomes. *Journal of Human Resources* 43(3):660–687
- Chetty R, Hendren N (2018) The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *Quart J Econ* 133(3):1107–1162
- Chetty R, Hendren N, Katz LF (2016) The effects of exposure to better neighborhoods on children: new evidence from the Moving to Opportunity experiment. *American Economic Review* 106(4):855–902
- Chiswick BR (1998) Hebrew language usage: determinants and effects on earnings among immigrants in Israel. *J Popul Econ* 11(2):253–271
- Chiswick BR (2009) The economics of language. In *The Education of Language Minority Immigrants in the United States*, edited by Wiley T, Sook Lee J, Rumberger R *Multilingual Matters*, Bristol:72–91
- Chiswick BR, Miller PW (2002) Immigrant earnings: language skills, linguistic concentrations and the business cycle. *J Popul Econ* 15(1):31–57

- Chiswick BR, Miller PW (2015) International migration and the economics of language. In *Handbook of the Economics of International Migration*, edited by Chiswick BR, Miller PW. North-Holland, Amsterdam:211–269
- Constant AF, Schüller S, Zimmermann KF (2013) Ethnic spatial dispersion and immigrant identity. IZA Discussion Paper 7868. Bonn: Institute for the Study of Labor
- Cortes KE (2006) The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Econ Educ Rev* 25(2):121–132
- Cutler DM, Glaeser EL (1997) Are ghettos good or bad? *Quart J Econ* 112(3):827–872
- Dahnen J, Kozłowicz W (1963) *Ausländische Arbeitnehmer in der Bundesrepublik. Sozialpolitik in Deutschland*. W. Kohlhammer Verlag, Stuttgart
- Damm AP (2009) Ethnic enclaves and immigrant labor market outcomes: quasi-experimental evidence. *J Law Econ* 27(2):281–314
- Damm AP, Dustmann C (2014) Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review* 104(6):1806–1832
- Danzer AM, Yaman F (2013) Do ethnic enclaves impede immigrants' integration? Evidence from a quasi-experimental social-interaction approach. *Rev Int Econ* 21:311–325
- Danzer AM, Yaman F (2016) Ethnic concentration and language fluency of immigrants: evidence from the guest-worker placement in Germany. *J Econ Behav Organ* 131:151–165
- De la Croix D, Doepke M (2009) To segregate or to integrate: education politics and democracy. *Rev Econ Stud* 76(2):597–628
- Dohse K (1981) *Ausländische Arbeiter und bürgerlicher Staat: Genese und Funktion von staatlicher Ausländerpolitik und Ausländerrecht: vom Kaiserreich bis zur Bundesrepublik Deutschland*. Hain, Königstein
- Dorner M, König M, Seth S (2011) Stichprobe der Integrierten Arbeitsmarktbiografien. *Regionalfiler 1975–2008 (SIAB-R 7508)*. FDZ-Datenreport 07/2011. Nürnberg
- Dustmann C (2003) Return migration, wage differentials, and the optimal migration duration. *Eur Econ Rev* 47(2):353–369
- Dustmann C, van Soest A (2001) Language fluency and earnings: estimation with misclassified language indicators. *The Review of Economics and Statistics* 83(4):663–674
- Dustmann C, Fasani F, Frattini T, Minali L, Schönberg U (2017) On the economics and politics of refugee migration. *Economic Policy* 32(91):497–550
- Dustmann C, Glitz A (2011) Migration and education. In *Handbook of the Economics of Education*, edited by Hanushek EA, Machin S, Woessmann L, vol 4. North Holland, Amsterdam:327–439
- Dustmann C, Glitz A, Schönberg U, Brücker H (2016) Referral-based job search networks. *Rev Econ Stud* 83(2):514–546
- Dustmann C, van Soest A (2002) Language and the earnings of immigrants. *Ind Labor Relat Rev* 55(3):473–492
- Edin P-A, Fredriksson P, Åslund O (2003) Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment. *Quart J Econ* 118(1):329–357
- Fassbender S (1966) Unternehmerische und betriebliche Probleme. In *Probleme der ausländischen Arbeitskräfte in der Bundesrepublik, Beihefte der Konjunkturpolitik* 13:50–55
- Federal Employment Agency (Bundesanstalt für Arbeitsvermittlung und Arbeitslosenversicherung) (1962) *Anwerbung und Vermittlung ausländischer Arbeitnehmer - Erfahrungsbericht 1961*. Nürnberg
- Federal Employment Agency (Bundesanstalt für Arbeitsvermittlung und Arbeitslosenversicherung) (1964) *Anwerbung und Vermittlung ausländischer Arbeitnehmer - Erfahrungsbericht 1963*. Nürnberg
- Federal Employment Agency (Bundesanstalt für Arbeitsvermittlung und Arbeitslosenversicherung) (1965) *Ausländische Arbeitnehmer: Beschäftigung, Anwerbung, Vermittlung - Erfahrungsbericht 1964*. Nürnberg
- Federal Employment Agency (Bundesanstalt für Arbeit) (1972) *Ausländische Arbeitnehmer: Beschäftigung, Anwerbung, Vermittlung - Erfahrungsbericht 1971*. Nürnberg
- Federal Employment Agency (Bundesanstalt für Arbeit) (1974) *Ausländische Arbeitnehmer: Beschäftigung, Anwerbung, Vermittlung - Erfahrungsbericht 1972/73*. Nürnberg
- Federal Employment Agency (Bundesagentur für Arbeit) (2017) *Zeitreihe für Kreise und kreisfreie Städte (Historische Daten von 1985 bis 2004)*. Produkt-ID: 211C016034
- Feuser G (1961) *Ausländische Mitarbeiter im Betrieb*. Verlag Moderne Industrie, München
- Frattini T, Meschi E (2019) The effect of immigrant peers in vocational schools. *Eur Econ Rev* 113(C):1–22
- Fryer RG Jr (2011) The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial achievement gap. In *Handbook of Social Economics*, edited by Benhabib J, Bisin A, Jackson MO. North-Holland, Amsterdam:1165–1191
- Gibbons S, Silva O, Weinhardt F (2013) Everybody needs good neighbours? Evidence from students' outcomes in England. *Economic Journal* 123(571):831–874

- Gibbons S, Silva O, Weinhardt F (2017) Neighbourhood turnover and teenage attainment. *J Eur Econ Assoc* 15(4):746–783
- Glitz A (2014) Ethnic segregation in Germany. *Labour Econ* 29:28–40
- Goebel J, Grabka MM, Liebig S, Kroh M, Richter D, Schröder C, Schupp J (2019) The German Socio-Economic Panel (SOEP). *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)* 239(2):345–360
- Grönqvist H (2006) Ethnic enclaves and the attainments of immigrant children. *Eur Sociol Rev* 22(4):369–382
- Haisken-DeNew JP, Frick JR (2005) Desktop Companion to the German Socio-Economic Panel Study (SOEP). Version 8.0. https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dc.409713.pdf. Accessed 14 May 2018
- Hatton TJ, Leigh A (2011) Immigrants assimilate as communities, not just as individuals. *J Popul Econ* 24(2):389–419
- Hémet C, Malgouyres C (2018) Diversity and employment prospects: Neighbors matter! *Journal of Human Resources* 53(3):825–858
- Imai K, Keele L, Yamamoto T (2010) Identification, inference and sensitivity analysis for causal mediation effects. *Stat Sci* 25(1):51–71
- Imai K, Keele L, Tingley D, Yamamoto T (2011) Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105(4):765–789
- Jensen P, Rasmussen AW (2011) The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Econ Educ Rev* 30(6):1503–1515
- Kristen C (2005) School choice and ethnic school segregation: Primary school selection in Germany. Waxmann: Münster/New York/München/Berlin
- Lee J-W, Lee H (2016) Human capital in the long run. *J Dev Econ* 122:147–169
- Marplan O (1982) Ausländische Arbeitnehmer in der Bundesrepublik (Datenpool 1970–1982). GESIS Datenarchiv, Köln. ZA1358 Datenfile Version 1.0.0, <https://doi.org/10.4232/1.1358>
- Max Planck Institute for Demographic Research, and Chair for Geodesy and Geoinformatics, University of Rostock (2011) MPIDR Population History GIS Collection (partly based on Bundesamt für Kartographie und Geodäsie (2011). Rostock
- McKenzie D, Rapoport H (2010) Self-selection patterns in Mexico-U.S. migration: The role of migration networks. *Rev Econ Stat* 92(4):811–821
- Musterd S (2005) Social and ethnic segregation in Europe: levels, causes, and effects. *Journal of Urban Affairs* 27(3):331–348. <https://doi.org/10.1111/j.0735-2166.2005.00239.x>
- Pei Z, Pischke J-S, Schwandt H (2019) Poorly measured confounders are more useful on the left than on the right. *Journal of Business & Economic Statistics* 37(2):205–216
- Penninx, R, Van Renselaar, H (1976) Evolution of Turkish migration before and during the current European recession. In *Migration and Development: A Study of the Effects of International Labor Migration on Bogazliyan District*, edited by Abadan-Unat N, Keles R, Penninx R, Van Velzen L, Yenisey L. Ajans-Turk Press, Ankara
- Schnepf SV (2007) Immigrants' educational disadvantage: an examination across ten countries and three surveys. *J Popul Econ* 20(3):527–545
- Schönwalder K, Sohn J (2009) Immigrant settlement structures in Germany: general patterns and urban levels of concentration of major groups. *Urban Studies* 46(7):1439–1460
- Schüller S (2016) Ethnic enclaves and immigrant economic integration. *IZA World of Labor* 2016:287. <https://doi.org/10.15185/izawol.287>
- Statistisches Bundesamt (2002) Allgemeinbildende Schulen 2000/01, Reihe 1, Fachserie 11, Bildung und Kultur. Wiesbaden: 78
- van Ours JC, Veenman J (2003) The educational attainment of second-generation immigrants in the Netherlands. *J Popul Econ* 16(4):739–753
- Voelker GE (1976) More foreign workers - Germany's labour problem No. 1. In *Turkish Workers in Europe 1960–1975*, edited by Abadan-Unat N. E. J. Brill, Leiden

Authors and Affiliations

Alexander M. Danzer^{1,2,3,4} · **Carsten Feuerbaum**¹ · **Marc Piopiunik**^{4,5} · **Ludger Woessmann**^{2,4,5,6}

✉ Ludger Woessmann
woessmann@ifo.de

Alexander M. Danzer
alexander.danzer@ku.de

Carsten Feuerbaum
carsten.feuerbaum@ku.de

Marc Piopiunik
piopiunik@ifo.de

¹ KU Eichstätt-Ingolstadt, Ingolstadt, Germany

² IZA, Bonn, Germany

³ CReAM, London, United Kingdom

⁴ CESifo, Munich, Germany

⁵ ifo Institute, Munich, Germany

⁶ University of Munich, Munich, Germany