Danzer, Alexander; Feuerbaum, Carsten; Piopiunik, Marc; Woessmann, Ludger

Working Paper

Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children

CReAM Discussion Paper Series, No. 09/18

Provided in Cooperation with:
Rockwool Foundation Berlin (RFBerlin)

Suggested Citation: Danzer, Alexander; Feuerbaum, Carsten; Piopiunik, Marc; Woessmann, Ludger (2018) : Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children, CReAM Discussion Paper Series, No. 09/18, Centre for Research & Analysis of Migration (CReAM), Department of Economics, University College London, London

This Version is available at:
https://hdl.handle.net/10419/295582

CReAM

Centre for Research & Analysis of Migration

Discussion Paper Series

CPD 09/18

Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children

Alexander M. Danzer, Carsten Feuerbaum, Marc Piopiunik, and Ludger Woessmann

Centre for Research and Analysis of Migration
Department of Economics, University College London
Drayton House, 30 Gordon Street, London WC1H 0AX

www.cream-migration.org

Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children*

Alexander M. Danzer, Carsten Feuerbaum, Marc Piopiunik, and Ludger Woessmann†

Abstract

Does a high regional concentration of immigrants of the same ethnicity affect immigrant children's acquisition of host-country language skills and educational attainment? We exploit the exogenous placement of guest workers from five ethnicities across German regions during the 1960s and 1970s in a model with region and ethnicity fixed effects. Our results 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 this 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.

Keywords: immigrant children, ethnic concentration, language, education, guest workers

JEL classification: J15, I20, R23, J61

June 4, 2018

* We gratefully acknowledge helpful comments and suggestions from George Borjas, Jesús Fernández-Huertas Moraga, Seth Gershenson, Jeffrey Grogger, John Haisken-DeNew, Martin Halla, Giovanni Mastrobuoni, Elie Murard, Francesco Ortega, 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 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).

† Danzer: KU Eichstätt-Ingolstadt, IZA, CReAM, and CESifo, alexander.danzer@ku.de; Feuerbaum: KU Eichstätt-Ingolstadt, carsten.feuerbaum@ku.de; Piopiunik: ifo Institute at the University of Munich and CESifo, piopiunik@ifo.de; Woessmann: University of Munich, ifo Institute, IZA, and CESifo, woessmann@ifo.de.

1. Introduction

With the recent arrival of large numbers of refugees in Europe, many societies wonder about the best policies to integrate immigrants. One central issue is the regional allocation of immigrants. To prevent ethnic ghettoization, many European countries adopted dispersal policies that assign refugees across 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). However, for the successful integration of immigrants into host-country societies in the long run, the intergenerational effects of ethnic concentration on the immigrants' children seem even more important. To that extent, immigrant children's proficiency in the host-country language and their educational attainment play a particular role for long-term employment opportunities and for cultural and social integration (Dustmann and Glitz, 2011; Chiswick and Miller, 2015). On the one hand, children's language acquisition and educational integration may benefit from ethnic enclaves that provide useful information, reduced discrimination, and positive role models. On the other hand, immigrant children may also be hindered by limited exposure to native children, reduced options for language acquisition, lower socioeconomic opportunities of families, and negative role models. 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 placement policy of the German guest-worker program. Between 1955 and 1973, the German government actively recruited (mainly low-skilled) foreign workers to fill labor shortages. The guest workers were enlisted in various countries of origin and then quasi-exogenously placed across West German firms. The German Socio-Economic Panel (SOEP) allows us to extract a sample of roughly 1,000 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 provides us with plausibly exogenous variation in ethnic concentration across regions, circumventing bias from endogenous sorting of immigrants into enclaves of co-ethnics. We show that demographics of guest-worker parents and their children are balanced across regions with low and high ethnic concentration. To account for any type of region-specific or ethnicity-specific differences, our models additionally 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 on immigrant children's host-country language proficiency and educational attainment by observing different (exogenously placed) immigrant groups who are exposed to differential concentrations of co-ethnics within the same region, thereby circumventing bias from endogenous location choices of immigrants and from unobserved factors such as differing baseline willingness or disposition to integrate of different ethnic groups.

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 percent to 2.8 percent – leads to a reduction in the German speaking proficiency of the children of the guest-worker generation by 19 percent of a standard deviation and a reduction in the German writing proficiency by 17 percent 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 (compared to an average of 7.1 percent). Although less robust, there is some indication that ethnic enclaves also reduce the probability of obtaining an intermediate or higher school degree. Concerning effect heterogeneities, we find that effects tend to be larger for those immigrant children who were born abroad, whereas there are no significant gender differences.

Importantly, 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. In particular, 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. For this analysis, it proves essential to address measurement error in the self-reported parental language measure by implementing an instrumental variable (IV) 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. Furthermore, none of the investigated mechanisms can explain the negative enclave effect on school dropout.

Our results are robust to a number of sensitivity analyses. In particular, we use alternative functional forms for the measure of ethnic concentration, instrument ethnic concentration at the time of observation by the ethnic concentration observed a decade earlier, use social-security as well as census data to construct the ethnic concentration measure, measure ethnic concentration at different levels of regional aggregation, and account for interview mode and intentions to return to the home country.

Our paper contributes to several strands of literature. Closest to our analysis is Åslund et al. (2011), who use a refugee placement policy in Sweden and find that the concentration of high-educated co-ethnics *positively* affects the achievement of immigrant students in school.[1] A crucial difference to their setting is that the guest-worker population in Germany is relatively low-educated, indicating that any effect of ethnic concentration may strongly depend on the skill level of co-ethnics in the enclave. In addition, the effect of enclaves on immigrant children's language proficiency may differ from the effect on how their achievement is evaluated by their teachers.

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, Fredriksson, and Åslund (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,

[1] Åslund et al. (2011) provide references to additional studies on ethnic concentration and immigrant children's outcomes that put less emphasis on addressing potential bias from non-random location decisions of immigrants.

respectively. In a different German setting, Battisti, Peri, and Romiti (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.

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, Hendren, and Katz, 2016; Chetty and Hendren, 2018; Gibbons, Silva, and Weinhardt, 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. More generally, most of the arguments in 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 second generation of immigrants, our results suggest that the fear of ghettoization that underlies the dispersal policies of several European countries may not be totally misplaced.

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 percent 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] Potential workers were screened for basic literacy and underwent medical check-ups.[3] Then, guest workers were matched with German employers. The employers could submit recruitment requests together with blank work contracts to their local labor offices, which forwarded them to the foreign branches after initial approval.[4]

German 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). Successful applicants got a work contract from a specific German company and a one-year work permit that was only valid for employment at the specific firm (Feuser, 1961). Recruited workers were then transferred to

[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 countries 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.

Germany in groups.[5] After having stayed with their initial employer for at least two years and in the same occupation (and, in practice, in the same region for most guest workers) for at least five 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, the initial location was exogenous from the perspective of an individual guest worker. Most importantly, the guest-worker recruitment process generated exogenous variation in ethnic concentrations that allows us to estimate the causal effect of ethnic concentration on immigrant children's outcomes.

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 Research Institute of the Federal Employment Agency (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), a large annual household survey that is representative of the resident population in Germany. The first SOEP wave in 1984 strongly oversampled guest workers (by a factor of four). As a consequence, 1,393 of the 5,921 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 in order to allow for stand-alone analyses (Haisken-DeNew and Frick, 2005). The SOEP contains detailed information on individual characteristics, including educational

[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).

attainment and, for foreigners, self-reported German speaking and writing proficiency.[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.

Our analysis sample consists of 1,065 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.[9] We keep only children with at least one observation for self-reported German language proficiency or one observation for educational attainment.[10]

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. 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

[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 for whom both mother and father could be identified.

[9] We present heterogeneity results below for guest-worker children born in Germany vs. children born abroad.

[10] 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).

every two years from 1997 to 2005, including about 4,900 child-year observations.[11] 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 intermediate school degree (*Realschulabschluss*) or a higher secondary school degree and 0 otherwise.[12] Children's educational attainment is based on the most recent available information in the SOEP.[13]

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 (statistical significance at 12 percent) and lower writing proficiency (significant at the 10 percent level) than immigrant children living in low co-ethnic concentration regions. 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 percent of immigrant children in our sample are Turkish, 19 percent each are Italian and Yugoslav, 15 percent are Greek, and 10 percent are Spanish. We identify the ethnicity of the immigrant children primarily based on their first citizenship (94.2 percent 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 A1 for definitions of all individual-level variables).[14] A slight majority of immigrant children in the sample (57.1 percent) were born in Germany. The average year of birth is 1971, and the average age at migration is 2.8 years.

[11] 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.

[12] In Germany, there are three types of secondary school degrees: basic (*Hauptschulabschluss*), intermediate (*Realschulabschluss*), and advanced (*Abitur*). A small share of children in our sample (2.9 percent) 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.

[13] 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.

[14] 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.

The SOEP also contains a rich set of additional individual characteristics, including the immigration history, educational attainment, and labor-market outcomes of adults.[15] This wealth of information allows us to investigate several potential mediating factors that may drive the effects of ethnic concentration. As potential mediating factors, we investigate parents' speaking and writing proficiency 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.

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*. Typically, these regions comprise several counties and constitute a regional labor market. In West Germany (incl. West Berlin), there were 103 *Anpassungsschichten* in 1985 with an average population of about half a million people. Allowing for sorting within large regions, this level of regional aggregation produces conservative estimates and circumvents potential bias from the typical sorting of immigrants into close-by cities or across city districts (Danzer and Yaman, 2016).

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 percent 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, König, and Seth, 2011). We use data from 1985, the year when guest workers' region of residence is observed for the first time 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 A2 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, Fredriksson, and Åslund, 2003;

[15] 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.

Damm, 2009; Åslund et al., 2011), below we also report the 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 regions (*Anpassungsschichten*) 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 regions (*Anpassungsschichten*) in 1985 (see Appendix Table A3 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. We exploit the differential concentrations of ethnicities across regions in our analyses by using only differences in ethnic concentrations within the same region.

For robustness analyses, we also use the 1987 German Census to compute alternative measures of ethnic concentration. Being based on a 2 percent 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 measurement error. In robustness analyses, we therefore also use an alternative measure of ethnic concentration based on data from the 1987 Census. An advantage of this alternative measure is that the 1987 Census includes the entire population in Germany. 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 major disadvantage of the 1987 Census is that it does not allow to compute ethnic concentrations for Spanish guest workers, which reduces the sample size and excludes one out of the five guest-worker ethnicities.[16] 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 two years later than the 1985 SIAB data. Appendix Figures A1 and A2 depict the distribution of the Census-based measures of ethnic concentration separately for the four ethnicities at the level of *Anpassungsschichten* and counties, respectively.

[16] Individuals with Spanish citizenship are included in the category "other citizenship." In the SOEP data, Spanish guest-worker children make up about 10 percent of the analysis sample.

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 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 ethnicitiy 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 model:

$$lang_{i,t} = \beta_0 + \beta_1 EC_i + \boldsymbol{C_i'}\beta_2 + \boldsymbol{P_i'}\beta_3 + \delta_r + \sigma_c + \tau_t + \mu_i + \epsilon_{i,t} \quad (1)$$

where $lang_{i,t}$ is the German speaking and writing proficiency, respectively, of child i in year t. The key explanatory variable is the concentration of child i's ethnicity in her region, EC_i.[17] $\boldsymbol{C_i}$ is a vector of child characteristics, including gender, year of birth, and age at migration. $\boldsymbol{P_i}$ 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),[18] and the number of children for mothers. All models include fixed effects for regions, δ_r, fixed effects for ethnicities (countries of origin), σ_c, and fixed effects for the year when the child reported her language proficiency, τ_t. The individual-specific effects, μ_i, are assumed to be i.i.d. random variables, and $\epsilon_{i,t}$ is an

[17] As described in Section 3.2, ethnic concentration, EC_i, 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.

[18] Among the parents in our sample, 2.9 percent of mothers and 0.8 percent of fathers are of German nationality without migration background.

idiosyncratic error term. Throughout, we cluster standard errors at the region-by-ethnicity level, the level at which our measure of ethnic concentration varies.

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:

$$educ_i = \theta_0 + \theta_1 EC_i + \boldsymbol{C_i'}\theta_2 + \boldsymbol{P_i'}\theta_3 + \delta_r + \sigma_c + \varepsilon_i \quad (2)$$

where $educ_i$ 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 equation (1), we include controls for child and parent characteristics as well as region and ethnicity 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 Balancing Test by Degree of 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. We can test this 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 indicated by Table 1, 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. These balancing tests support our assumption that there was no systematic self-selection of guest workers into regions of differing ethnic concentration.

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 do not explain 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, we have to assume 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.

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, Schüller, and Zimmermann, 2013; Danzer and Yaman, 2013, 2016). In contrast to the settings studied in some other papers (such as refugees in Sweden in Åslund et al., 2011), the evidence against endogenous sorting of immigrants into ethnic enclaves in our setting is perfectly consistent with two specific features of the German guest-worker program.

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 Kozlowicz, 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 to fill open positions in the German economy. As a result, the unemployment rate of foreigners in Germany was less than 1.5 percent 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 West German regions to estimate the effect of ethnic enclaves on immigrant children's outcomes.

5. The Effect of Ethnic Concentration on Immigrant Children's Language Proficiency and Educational Attainment

This section presents our main results (Section 5.1) and subgroup analyses (Section 5.2). In the subsequent sections, we provide investigations of mediating factors and 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 speaking and writing proficiency in German. An increase in the size of the own ethnicity by one log-point is related to a decline in speaking skills by 19 percent and in writing skills by 17 percent of a standard deviation. The magnitudes of the estimated coefficients barely change when we include controls for children's and parents' characteristics.

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 percent) to the city of Munich (with a share of about 2.8 percent).[19] This change in the region of residence would, ceteris paribus, reduce the child's German speaking proficiency by 19 percent and her writing proficiency by 17 percent 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 drop-out rate among immigrant children in our sample is only 7.1 percent, this is a huge effect. 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).[20]

Both findings – the negative effect on host-country language proficiency and the negative effect 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.

5.2 Subgroup Analysis

Next, we investigate effect heterogeneity by country of birth, gender, and ethnicity. We start by investigating whether the negative effects of ethnic concentration on children's outcomes differ between children born abroad and children born in Germany. About 42 percent 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 proficiency are roughly 30 percent smaller

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

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

for children who were born in Germany rather than abroad. As children born in Germany start learning the German language already in kindergarten 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 negative impact on the host-country language proficiency of children born in Germany does not translate into a smaller disadvantage 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 A4). Results suggest that the effect of ethnic concentration on German speaking and writing proficieny 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 more 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 constrast, 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. 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 report their German language proficiency in speaking and writing similar to their children. Using the same random effects specification (without child controls) and the same definitions for language proficiency and ethnic concentration as 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 5 adds different potential mediating factors as control variables to our main model for children's German speaking proficiency.[21] As indicated in column 2, parents' German speaking proficiency is significantly positively related to their children's German speaking proficiency. 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.[22]

After accounting for random measurement error by instrumenting parents' speaking proficiency with their reported proficiency in the preceding and subsequent years, 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. 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 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

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

[22] 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. 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).

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.

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 proficiency.[23] 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.

Furthermore, differences in economic conditions such as parental unemployment or household income might 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.

Similar analyses indicate that none of the mediating factors analyzed here can account for the effect of ethnic enclaves on children's schooling outcomes. As indicated in Table 7, parents' speaking ability is the only analyzed factor that is significantly associated with their children's probability to obtain a school degree. Still, controlling for parents' speaking ability

[23] 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).

does not reduce the estimated effect of ethnic concentration on whether children obtain a school degree.[24]

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 explains the negative enclave effect only to a small extent. By contrast, 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.[25]

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 ethnic concentration in 1975 (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 interview mode and for intentions to return to the home country (Section 7.5).

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, Fredriksson, and Åslund (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 proficiency 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

[24] 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).

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

own-ethnics in the regional population reduces the likelihood of obtaining at least an intermediate school degree by 5.1 percent.

7.2 Instrumenting Ethnic Concentration in 1985 by Ethnic Concentration in 1975

As discussed in Section 4.2, we do not observe guest workers and their region of residence before 1985. While the balancing tests indicated 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 can rule 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 example, due to improving or deteriorating economic conditions.

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 248.9 in the regressions for language outcomes and 321.3 in the regressions for schooling outcomes.[26] 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.

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 proficiency, 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.

[26] 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.

7.3 Measuring Ethnic Concentration with Census Data

Measuring the size of the immigrant population based on a 2 percent 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.

As the odd-numbered columns of Appendix Table A5 indicate, replacing the 1985 SIAB measure of ethnic concentration with the 1987 Census measure yields very similar results to our main specifications. Furthermore, the even-numbered columns show 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. These IV estimates, 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 model (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 *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 *Anpassungsschichten*. However, the guest-worker children observed in the SOEP data live in only 114 different counties, reducing the variation used in the analysis.

When measuring ethnic concentration at the county level, the effects of ethnic concentration on children's speaking and writing proficiency are very similar to the estimates

when measuring ethnic concentration at the *Anpassungsschicht* level (Appendix Table A6). 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 percent). 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 and Intentions to Return Home

Finally, we show that our results are not driven by two potential alternative explanations for the estimated ethnic enclave effects. First, immigrants' self-reports of their language proficiency may be affected by the specific interview mode used in the SOEP, such as oral face-to-face interview or written interview by mail. Therefore, the first two columns of Appendix Table A7 control for the interview mode used when guest-worker children report their levels of German language proficiency. Adding this control does not affect the estimated enclave effects on children's proficiency in speaking or writing German.

Second, 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, columns 3-6 of Appendix Table A7 include a binary indicator that equals 1 if guest-worker parents see their future in Germany (0 otherwise).[27] Adding this control variable does not affect our baseline estimates. 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.

8. Conclusion

We exploit 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

[27] 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].

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 investigating several factors that might mediate the effect of ethnic concentration on child outcomes. We find that parents' German speaking proficiency completely explains the negative effect of ethnic enclaves on their children's German language proficiency. Parents' writing abilities explain 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. Since the enclave effect on children's language proficiency is completely explained by parents' lower host-country language skills, our findings suggest that host-country language training for adult immigrants might have important positive spillover effects on their children. Language training for adult immigrants would complement current policies in Germany that emphasize language training for immigrant children themselves, which includes compulsory German language tests before starting school.

More generally, our results indicate that the long-run cultural and social integration of immigrants, including the next generation, may be more successful when immigrants do not live in ethnic enclaves. Concerning current policy debates about how to disperse refugees across regions, our findings suggest that avoiding the emergence of ethnic enclaves might help refugee children to learn the host-country language and to avoid school dropout.

References

Åslund, Olof, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist. 2011. "Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy." *American Economic Journal: Applied Economics* 3 (2): 67-95.

Aydemir, Abdurrahman, and George J. Borjas. 2011. "Attenuation bias in measuring the wage impact of immigration." *Journal of Labor Economics* 29 (1): 69-112.

Battisti, Michele, Giovanni Peri, and Agnese Romiti. 2016. "Dynamic effects of co-ethnic networks on immigrants' economic success." NBER Working Paper 22389. Cambridge, MA: National Bureau of Economic Research.

Bundesamt für Kartographie und Geodäsie. 2011. "VG 2500 Verwaltungsgebiete (Ebenen) 1:2.500.000." Stand 01.01.2009. Frankfurt am Main.

Chetty, Raj, and Nathaniel Hendren. 2018. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *Quarterly Journal of Economics*, forthcoming.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 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, Barry R. 2009. "The economics of language." In *The Education of Language Minority Immigrants in the United States*, edited by Terrance Wiley, Jin Sook Lee, and Russell Rumberger, 72-91. Bristol: Multilingual Matters.

Chiswick, Barry R., and Paul W. Miller. 2015. "International migration and the economics of language." In *Handbook of the Economics of International Migration*, edited by Barry R. Chiswick and Paul W. Miller, 211-269. Amsterdam: North-Holland.

Constant, Amelie F., Simone Schüller, and Klaus F. Zimmermann. 2013. "Ethnic spatial dispersion and immigrant identity." IZA Discussion Paper 7868. Bonn: Institute for the Study of Labor.

Cutler, David M., and Edward L. Glaeser. 1997. "Are ghettos good or bad?" *Quarterly Journal of Economics* 112 (3): 827-872.

Dahnen, Josef, and Werner Kozlowicz. 1963. *Ausländische Arbeitnehmer in der Bundesrepublik. Sozialpolitik in Deutschland*. Stuttgart: W. Kohlhammer Verlag.

Damm, Anna Piil. 2009. "Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence." *Journal of Labor Economics* 27 (2): 281-314.

Danzer, Alexander M., and Firat Yaman. 2013. "Do ethnic enclaves impede immigrants' integration? Evidence from a quasi-experimental social-interaction approach." *Review of International Economics* 21: 311-325.

Danzer, Alexander M., and Firat Yaman. 2016. "Ethnic concentration and language fluency of immigrants: Evidence from the guest-worker placement in Germany." *Journal of Economic Behavior & Organization* 131: 151-165.

Dohse, Knuth. 1981. *Ausländische Arbeiter und bürgerlicher Staat: Genese und Funktion von staatlicher Ausländerpolitik und Ausländerrecht: vom Kaiserreich bis zur Bundesrepublik Deutschland*. Königstein: Hain.

Dorner, Matthias, Marion König, and Stefan Seth. 2011. "Stichprobe der Integrierten Arbeitsmarktbiografien. Regionalfile 1975-2008 (SIAB-R 7508)." FDZ-Datenreport 07/2011. Nürnberg.

Dustmann, Christian, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg. 2017. "On the economics and politics of refugee migration." *Economic Policy* 32 (91): 497-550.

Dustmann, Christian, and Albrecht Glitz. 2011. "Migration and education." In *Handbook of the Economics of Education, Vol. 4*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 327-439. Amsterdam: North Holland.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. 2016. "Referral-based job search networks." *Review of Economic Studies* 83 (2): 514-546.

Dustmann, Christian, and Arthur van Soest. 2002. "Language and the earnings of immigrants." *Industrial and Labor Relations Review* 55 (3): 473-492.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment." *Quarterly Journal of Economics* 118 (1): 329-357.

Fassbender, Siegfried. 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 Arbeitslosen-versicherung). 1962. *Anwerbung und Vermittlung ausländischer Arbeitnehmer - Erfahrungsbericht 1961*. 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, Günther. 1961. *Ausländische Mitarbeiter im Betrieb*. München: Verlag Moderne Industrie.

Fryer, Roland G., 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 Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1165-1191. Amsterdam: North-Holland.

Gibbons, Stephen, Olmo Silva, and Felix Weinhardt. 2013. "Everybody needs good neighbours? Evidence from students' outcomes in England." *Economic Journal* 123 (571): 831-874.

Gibbons, Stephen, Olmo Silva, and Felix Weinhardt. 2017. "Neighbourhood turnover and teenage attainment." *Journal of the European Economic Association* 15 (4): 746-783.

Glitz, Albrecht. 2014. "Ethnic segregation in Germany." *Labour Economics* 29: 28-40.

Haisken-DeNew, John P., and Joachim R. Frick. 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/dtc.409713.pdf (Accessed 14 May 2018).

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.

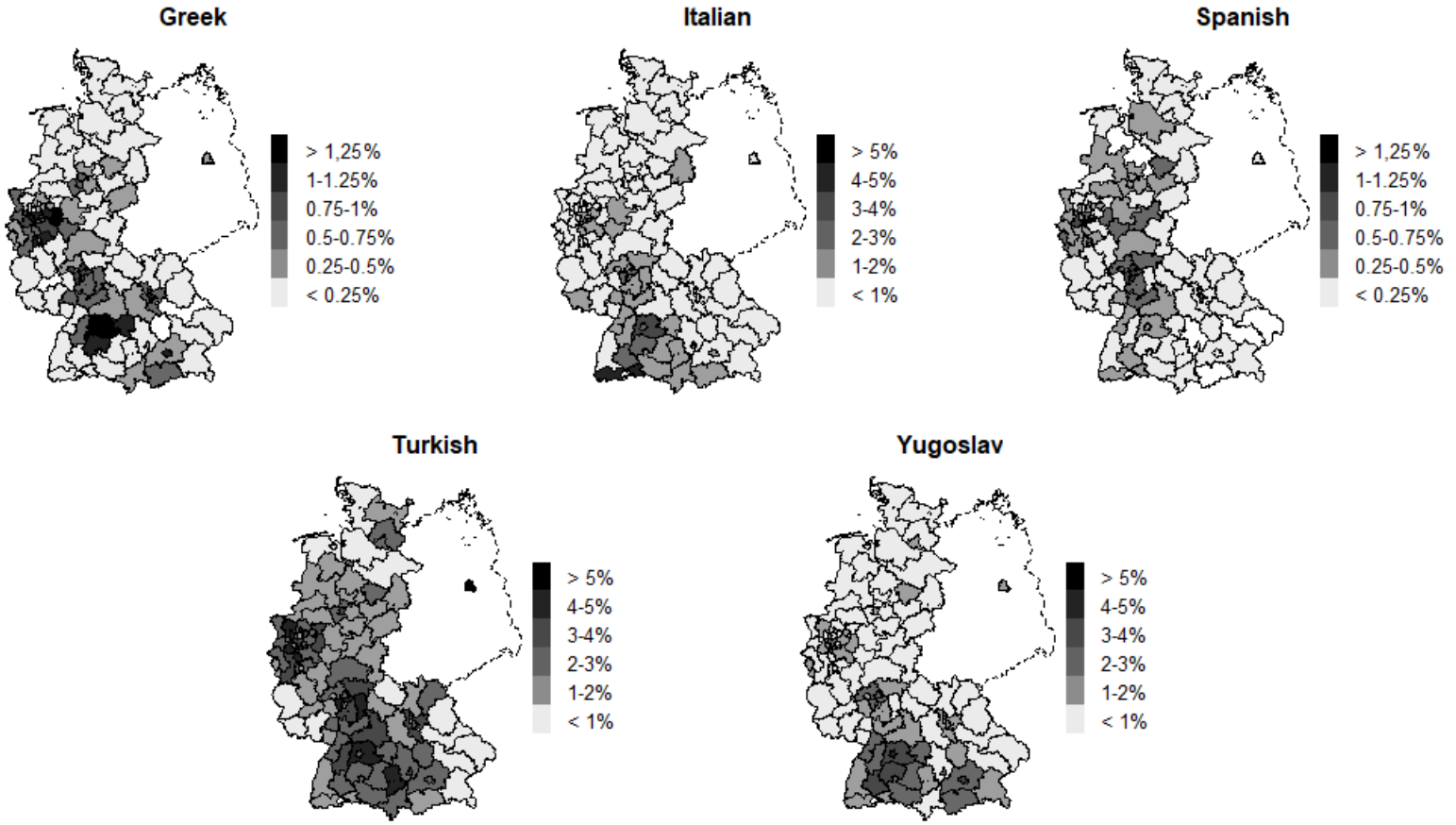
Penninx, Rinus, and Herman Van Renselaar. 1976. "Evolution of Turkish migration before and during the current European recession." In *Migration and Development: A Study of the Effects of International Labor Migration on Bogazliyan District*, edited by Nermin Abadan-Unat, Rusan Keles, Rinus Penninx, Leo Van Velzen, and Leyla Yenisey. Ankara: Ajans-Turk Press.

Schönwalder, Karen, and Janina Söhn. 2009. "Immigrant settlement structures in Germany: General patterns and urban levels of concentration of major groups." *Urban Studies* 46 (7): 1439-1460.

Schüller, Simone. 2016. "Ethnic enclaves and immigrant economic integration." *IZA World of Labor* 2016: 287 doi: 10.15185/izawol.287.

Voelker, Gottfried E. 1976. "More foreign workers - Germany's labour problem No. 1." In *Turkish Workers in Europe 1960-1975,* edited by Nermin Abadan-Unat. Leiden: E. J. Brill.

Figure 1: Ethnic Concentrations across West Germany, 1985



Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) of residence, 1985. Source: Institut für Arbeitsmarkt- und Berufsforschung (IAB). Own calculations of ethnic concentrations for 103 Anpassungsschichten. Figures based on a historical GIS datafile 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 1: Descriptive Statistics by Degree of Ethnic Concentration

Variable	Low EC	High EC	Diff.	P-Value	Obs.
Outcomes (Children)					
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
Children					
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
Mothers					
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
Education in country of origin					
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
Not employed (1984-1986)	0.57	0.51	0.05	0.36	1065
Unemployed (1984-1986)	0.06	0.04	0.02	0.27	1065
Fathers					
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
Education in country of origin					
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
Never moved flat since arrival in Germany	0.04	0.04	0.00	0.87	1065
Not employed (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
For Comparison					
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

Notes: Variable means by degree of ethnic concentration. 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 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, not employed, and unemployed refer to three-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).

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**	-0.185**	-0.167**	-0.173**
	(0.083)	(0.081)	(0.075)	(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

Notes: Random Effects Model. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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***	-0.056***	-0.059	0.002
	(0.019)	(0.021)	(0.052)	(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

Notes: OLS regressions. Dependent variables: 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. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. 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 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**	-0.200***	-0.060**	-0.022	-0.174**	-0.154**	-0.064***	0.009
	(0.082)	(0.070)	(0.024)	(0.051)	(0.082)	(0.074)	(0.021)	(0.051)
Size of ethnic group * born in Germany	0.062**	0.063**	0.009	0.049**				
	(0.028)	(0.026)	(0.018)	(0.021)				
Size of ethnic group * female					-0.027	-0.049	0.019*	-0.016
					(0.031)	(0.038)	(0.010)	(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			0.271	0.295		
Adjusted R^2			0.050	0.215			0.052	0.210

Notes: Columns 1-2 and 5-6: Random Effects Model. Columns 3-4 and 7-8: OLS regressions. Dependent variables: Speaking/writing proficiency: 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. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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 5: Mediating Factors - Effect of Ethnic Concentration on Host-Country Speaking Proficiency

	Baseline	Mediating Factors						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1985	-0.178**	-0.123	-0.011	-0.173**	-0.136	-0.181**	-0.169**	-0.182**
	(0.085)	(0.084)	(0.100)	(0.08 6)	(0.130)	(0.083)	(0.084)	(0.085)
Speaking abilities, parents		0.165***						
		(0.019)						
Speaking abilities, parents, IV lead + lag			0.519***					
			(0.097)					
Writing abilities, parents				0.073***				
				(0.024)				
Writing abilities, parents, IV lead + lag					0.617**			
					(0.240)			
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

Notes: Random Effects Model. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Speaking proficiency: self-assessed speaking ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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 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*	-0.095	0.027	-0.127	-0.092	-0.138*	-0.126	-0.138*
	(0.079)	(0.078)	(0.097)	(0.080)	(0.115)	(0.077)	(0.079)	(0.078)
Speaking abilities, parents		0.121***						
		(0.021)						
Speaking abilities, parents, IV lead + lag			0.562***					
			(0.115)					
Writing abilities, parents				0.103***				
				(0.024)				
Writing abilities, parents, IV lead + lag					0.662**			
					(0.281)			
First friend German						0.252***		
						(0.054)		
Visits from Germans, parents							0.082*	
							(0.046)	
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

Notes: Random Effects Model. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Writing proficiency: self-assessed writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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 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***	-0.062***	-0.065***	-0.062***	-0.064***	-0.056**	-0.062***	-0.065***
	(0.022)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.022)	(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

Notes: OLS regressions. Columns 3+5: IV models using lead and lag of parents' speaking/writing proficiency as instruments. Dependent variable: Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. 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 8: Measuring Ethnic Concentration by Share of Own Ethnicity in Regional Population

	Speaking proficiency	Writing proficiency	Any school degree	Intermediate school degree
	(1)	(2)	(3)	(4)
Share of own ethnicity in 1985	-0.080**	-0.080***	-0.025**	-0.051**
	(0.034)	(0.029)	(0.011)	(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

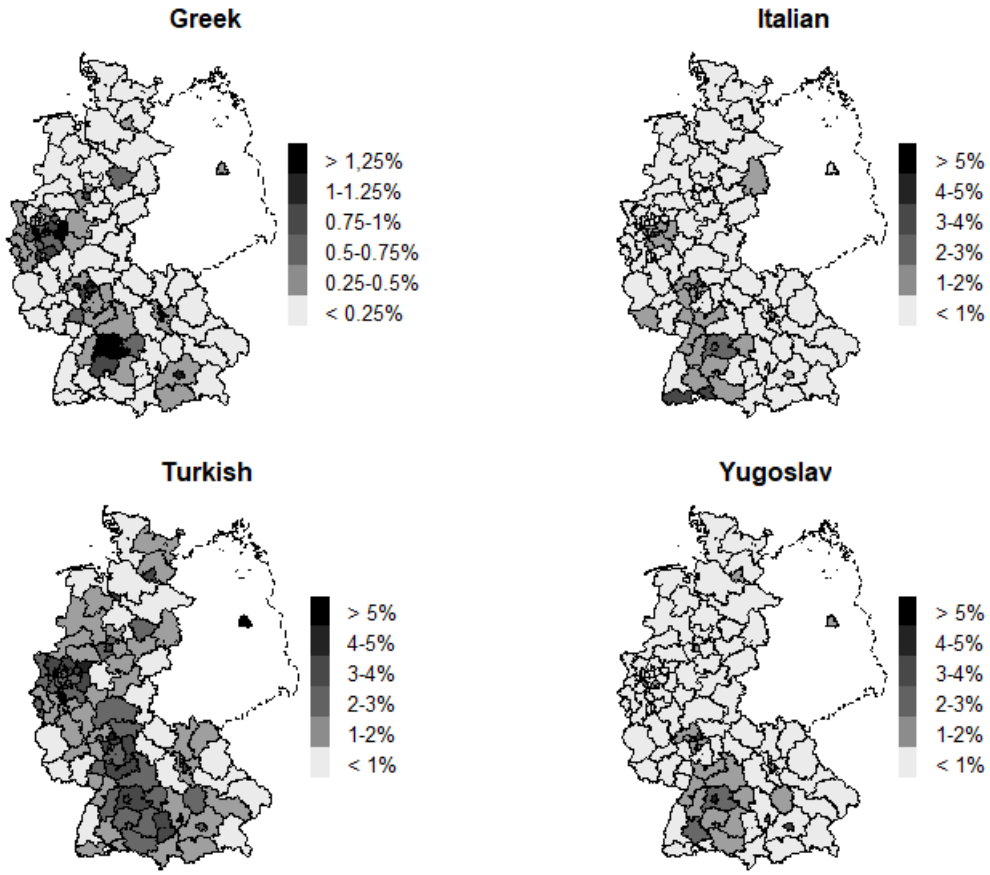
Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Share of own ethnicity in 1985: share of own ethnicity in the population of the region (*Anpassungsschicht*) 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 9: Instrumental-Variable Estimates using Ethnic Concentration in 1975

	Speaking proficiency	Writing proficiency	Any school degree	Intermediate school degree
	(1)	(2)	(3)	(4)
Size of ethnic group in 1985	-0.234**	-0.183**	-0.056***	0.032
	(0.103)	(0.075)	(0.019)	(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	248.87	248.87	321.309	321.309

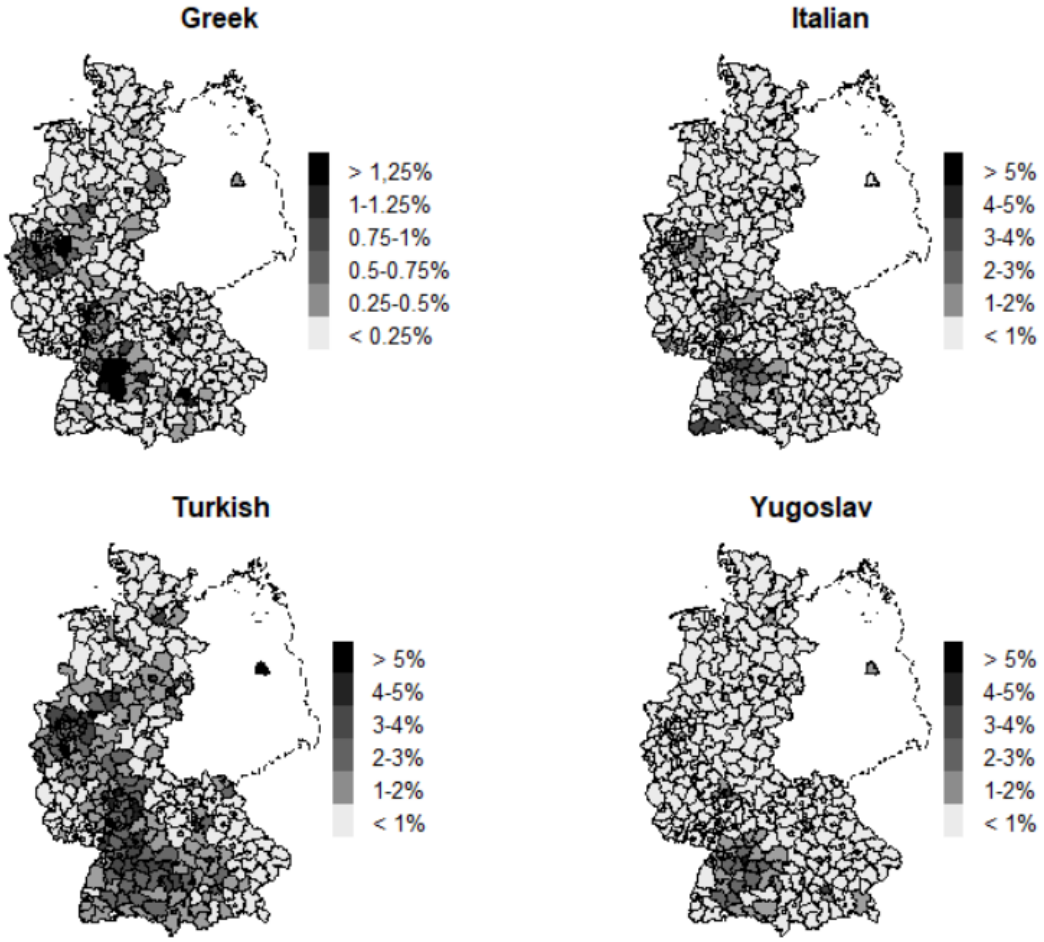
Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Size of ethnic group in 1985 is instrumented by size of ethnic group in 1975 (both variables in logs). Dependent variables: Speaking/writing proficiency: self-assessed speaking/writing ability in German (from 1="not at all" to 5="very 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. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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).

Figure A1: Ethnic Concentrations across West Germany: Census 1987



Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 103 Anpassungsschichten. Figures based on a historical GIS datafile 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).

Figure A2: County-Level Ethnic Concentrations across West Germany: Census 1987



Notes: Share of ethnicity in the total population of the county of residence, 1987. Source: German Census 1987. Own calculations of ethnic concentrations for 328 counties. Figures based on a historical GIS datafile 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 A1: Individual-Level Variables

Variable	Description
Outcomes (Children)	
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, *Random Effects Model*: 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, *Random Effects Model*: 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 *dropout*, *no school degree* or *no school degree yet*, 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 *dropout*, *no school degree* or *no school degree yet*, 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.
Demographics of Children	
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.
Demographics of Parents	
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 *No schooling*, *Incomplete compulsory schooling*, and *At least compulsory schooling*, based on survey question "Obtained School Degree Outside Germany" in survey year 1985.

Continued on next page

Table A1 *(Continued)*

Variable	Description
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 "sumkids" 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.
Not employed (1984-1986)	Mean of an indicator for being not employed during the survey years 1984-1986.
Unemployed (1984-1986)	Mean of an unemployment dummy during the survey years 1984-1986.
Mediating Factors	
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 %).

Continued on next page

Table A1 *(Continued)*

Variable	Description
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.
Robustness Checks	
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.
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".

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

Table A2: Regional Variables

Variable	Description
Size of ethnic group in 1985	Log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) 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 *Region of residence* 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: *Anpassungsschicht* (Table A5) or county (Table A6). Measure not available for immigrant children with Spanish ethnicity.
Official unemployment rate 1985	Unemployment rate in the year 1985, regional level: *Anpassungsschicht*, based on county-level data from Federal Employment Agency (2017).
Region of residence (used to construct ethnic concentration measures)	The region of residence (*Anpassungsschicht* 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 A3: 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

Notes: Share of ethnicity in the total population of the region (*Anpassungsschicht*) 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 A4: Subgroup Analysis by Ethnicity

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

Notes: Dependent variables: Panel A: Speaking proficiency: self-assessed speaking ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Panel B: Writing proficiency: self-assessed writing ability in German (from 1="not at all" to 5="very well"), normalized to mean 0 and standard deviation 1. Panel C: Any school degree: 1 if individual obtained any type of school degree, 0 otherwise. Panel D: At least intermediate school degree: 1 if individual obtained at least an intermediate school degree, 0 otherwise. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Panels A and B additionally include 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 A5: Ethnic Concentration Measured in 1987 Census

	Speaking proficiency		Writing proficiency		Any school degree		Intermediate school degree	
	Baseline	IV	Baseline	IV	Baseline	IV	Baseline	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size of ethnic group in 1987	-0.238**	-0.265**	-0.167**	-0.193**	-0.063**	-0.071***	0.000	0.004
	(0.094)	(0.111)	(0.083)	(0.084)	(0.024)	(0.022)	(0.049)	(0.049)
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	n.a.	n.a.	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	4523	4523	4514	4514	907	907	907	907
R^2 overall	0.272	0.271	0.301	0.300				
Adjusted R^2					0.038	0.038	0.229	0.229
First-stage F-statistic		278.58		278.58		384.913		384.913

Notes: Columns 1-4: Random Effects Model. Columns 5-8: OLS regressions. Columns 2, 4, 6, and 8: Size of ethnic group in 1987 is instrumented by size of ethnic group in 1975 (both variables in logs). Dependent variables: Speaking/writing proficiency: 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. Size of ethnic group in 1987: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, German Census 1987. 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), German Census 1987.

Table A6: Ethnic Concentration Measured at County Level (1987 Census)

	Speaking proficiency (1)	Writing proficiency (2)	Any school degree (3)	Intermediate school degree (4)
Size of ethnic group in 1987	-0.244**	-0.209**	-0.021	-0.003
	(0.097)	(0.087)	(0.035)	(0.054)
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	4523	4514	907	907
R^2 overall	0.305	0.333		
Adjusted R^2			0.033	0.234

Notes: Columns 1-2: Random Effects Model. Columns 3-4: OLS regressions. Dependent variables: Speaking/writing proficiency: 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. Size of ethnic group in 1987: log size of ethnic community (individuals of same ethnicity) in county of residence, German Census 1987. 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. 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 county-ethnicity level, in parentheses. Significance levels: * p<0.10, ** p<0.05,*** p<0.01. Data sources: German Socio-Economic Panel (SOEP), German Census 1987.

Table A7: Controlling for Interview Mode and Return Intention

	Speaking proficiency	Writing proficiency	Speaking proficiency	Writing proficiency	Any school degree	At least intermediate degree
	(1)	(2)	(3)	(4)	(5)	(6)
Size of ethnic group in 1985	-0.171**	-0.161**	-0.173**	-0.162**	-0.054**	0.006
	(0.083)	(0.068)	(0.080)	(0.069)	(0.021)	(0.049)
Stay in Germany, parents			0.105	0.085	0.036*	0.016
			(0.080)	(0.078)	(0.020)	(0.056)
Interview mode	Yes	Yes	No	No	No	No
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year of assessment	Yes	Yes	Yes	Yes	n.a.	n.a.
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4932	4922	4932	4922	1005	1005
R^2 overall	0.305	0.319	0.272	0.295		
Adjusted R^2					0.052	0.212

Notes: Columns 1-4: Random Effects Model. Columns 5-6: OLS regressions. Dependent variables: Speaking/writing proficiency: 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. Size of ethnic group in 1985: log size of ethnic community (individuals of same ethnicity) in region (*Anpassungsschicht*) of residence, 1985. Stay in Germany, parents: a binary dummy indicating the intent to stay in Germany (average of the variable of mother and father). Interview mode: dummies for different types of interview method such as "Oral Interview" and "Written, By Mail". 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).