Growing Up in Ethnic Enclaves: The Effects on Education and Language Proficiency

Alexander M. Danzer
KU Eichstätt Ingolstadt, IZA Bonn and CESifo

Carsten Feuerbaum
KU Eichstätt Ingolstadt

Marc Piopiunik
Ifo Institute and CESifo

Ludger Woessmann
University of Munich, Ifo Institute, IZA Bonn and CESifo

April 2017
WORK IN PROGRESS
PLEASE DO NOT CITE WITHOUT PERMISSION.

ABSTRACT

Does a high regional concentration of migrants of the same ethnicity affect the educational achievement and acquisition of host-country language skills of immigrant children? Exploiting the exogenous placement of guest workers from five ethnicities across German regions during the 1960s and 70s, we find that exposure to a higher own-ethnic concentration increases school dropout rates. Consistent with this finding, more co-ethnics in the region also impair immigrant children's proficiency in the host-country language. A key mediating factor for this effect is parents' lower speaking proficiency in the host-country language.

KEYWORDS: Immigrant children, ethnic concentration, education, language, guest worker program

JEL Classification: R23, I20, J15

We thank Seth Gershenson, Laura Viluma, Giovanni Mastrobuoni, Steven Stillman, Martin Halla, Jeffrey Grogger, Andreas Steinmayr, Elie Murard, Francesc Ortega, Jesús Fernández-Huertas Moraga, Uwe Sunde and Joachim Winter for helpful comments and suggestions. We also thank seminar audiences at the RES Symposium of Junior Researchers 2017 (Bristol), the Applied Microeconomics Workshop (Bolzano), the ZEW Workshop on Assimilation and Integration of Immigrants (Mannheim), the 4th MGSE Colloquium (Munich), the EBE Research Strategy Seminar (Munich), the ifo CEMIR Seminar (Munich), the Ifo Junior Economist Workshop on Migration Research (Munich) and the 22th BGPE Research Workshop (Munich).

1 Introduction

The economic and social integration of immigrant children is a main concern of policy makers. Within the context of the recent inflow of refugees to Europe, the EU-28 countries received more than two million asylum applications in 2015 and 2016 alone.[1] Given that 29 % of non-EU asylum applicants were younger than 18 years old in 2015[2], the integration of refugee children has also become an important issue on the political agenda in European countries. While failed integration bears substantial monetary and non-monetary costs, the educational attainment of immigrant children is considered a key factor for successful integration. Therefore, it is essential to understand the relevant factors influencing the educational achievement of immigrant children.

Recent evidence highlights the role of childhood environment for educational success. Chetty et al. (2016) study the effects of the Moving to Opportunity (MTO) experiment on childrens' long term outcomes[3], finding that moving to a better neighborhood during childhood significantly improves college attendance rates. Exploiting a quasi-experiment of more than five million families moving across counties in the U.S, Chetty and Hendren (2016) find that children who spent more time growing up in better neighborhoods are more likely to attend college later. Regarding immigrant children, Grönqvist (2006) shows that growing up in an ethnic enclave is associated with a decreased probability of graduating from higher education in Sweden. Contrasting to the latter, Åslund et al. (2011) find that an increase in the share of highly educated individuals in the ethnic communities has positive effects on immigrant childrens' school GPA.

We investigate the effect of the size of the ethnic community in the region of residence on immigrant childrens' educational achievement and host-country language proficiency within the context of the German Guest Worker Program. Besides education, proficiency in the host-country language is a relevant outcome of interest because it has been shown to be an important determinant of labour market integration (e.g., Dustmann and Soest (2001), Dustmann and Fabbri (2003), Aldashev et al. (2009)). While the precise mechanism through which a larger ethnic community might affect immigrant childrens' education and language skills are not well empirically studied, there are many competing hypothesis: e.g., ethnic concentration may reduce the host-country language proficiency of immigrant parents and their children due to fewer incentives to learn the language and a lack of op-

[1] Source: Eurostat, "Asylum and first time asylum applicants - annual aggregated data (rounded) (tps00191)", updated 11 November 2016.

[2] Source: Eurostat, "Asylum and first time asylum applicants by citizenship, age and sex Annual aggregated data (rounded) (migr_asyappctza)", updated 17 February 2017.

[3] The Moving to Opportunity (MTO) was a randomized social experiment sponsored by the U.S. Department of Housing and Urban Development (HUD). Starting in 1994, randomly selected families living in high poverty areas were offered housing vouchers to move to better neighborhoods.

portunities to do so. Given that Dustmann et al. (2010) highlight the role of language in explaining educational gaps between immigrant and native children in the UK, parents language proficiency might be a particularly important mediating factor: if parents' host-country language proficiency is lower in ethnic enclaves, this might in turn influence their childrens' language proficiency and their educational attainment in the host country. On the other hand, ethnic concentration has been shown to improve labour market outcomes of immigrants (see e.g., Edin et al. (2003)). Such an improvement in parents' labour market outcomes may lead to positive effects of ethnic concentration on immigrant childrens' educational attainment. However, in our setting immigrant childrens' parents - the guest workers - tended to be well integrated in the labour market due to the demand driven recruitment. Furthermore, the contact with co-ethnics might also provide immigrant parents with valuable information on institutional features of the host-country such as the educational system. Furthermore, adults of the same ethnicity might also act as role models for immigrant children (see Åslund et al. (2011)). While previous studies (e.g., Borjas (1995), Åslund and Fredriksson (2009)) indicate that enclave characteristics may affect the effects of the exposure to co-ethnics, guest workers' educational background was, on average, low compared to that of natives. Given that the sign of the effect of ethnic concentration is ambiguous from a theoretical perspective, we perform an empirical analysis on the effects of ethnic concentration on immigrant children using survey data from the German Socio-Economic Panel (SOEP) and measures of ethnic concentrations based on a large employee sample from the Institut für Arbeitsmarkt- und Berufsforschung (IAB).

To obtain quasi-random variation in ethnic concentrations, we exploit the exogenous placement of guest workers from five ethnicities across German regions during the 1960s and early 1970s. Since the initial job location was exogenous from the perspective of an individual guest worker, the placement procedure rules out differences in immigrant parents' unobserved characteristics such as their willingness to integrate. While we control for region fixed effects and country-of-origin fixed effects, the presence of five different immigrant groups with different exposures within the same region allows us to identify the effect of ethnic concentration.

With the inclusion of region and country-of-origin fixed effects and the focus on immigrant children of guest workers, the causal interpretation of our estimates hinges on the assumption that guest workers didn't endogenously sort into ethnic enclaves between the initial exogenous placement and the year for which their children report their educational attainment and language proficiency. In particular, the sorting of immigrants with unobserved unfavorable characteristics into regions with high ethnic concentrations is a primary threat to validity (see e.g., Cutler and Glaeser (1997)). Several facts support our key identifying assumption: first, prior research has documented that demographic characteristics

are balanced between guest workers living in low concentration regions and those living in high concentration regions (Danzer and Yaman (2013), Constant et al. (2013), Danzer and Yaman (2016)). Second, comparing demographic characteristics of immigrant children and their parents by degree of ethnic concentration, we find no evidence of sorting on observables such as parents' educational background. Given that parents' educational background and their moving patterns might be good measures of potential unobserved factors, this finding is particularly reassuring. Third, the evidence against endogenous sorting is consistent with specific features of the German guest worker program: guest workers were restricted in their residential choice because the work permit required guest workers to stay with their first employer for at least three years and within the same occupation and region for eight years (Dahnen and Kozlowicz (1963)). Fourth, the evidence against sorting is consistent with the type of migration in our setting: guest worker tended to be well integrated in the labour market compared to other immigrant groups resulting in lower incentives to move across regions. Moreover, Schonwalder and Sohn (2009) find that immigrants' settlement structures in Germany still reflect the lack of available manpower in the 1960s and 1970s. Finally, ethnic segregation has been quite stable across workplaces and residential locations over the period from 1975 to 2008 (Glitz (2014)). Nevertheless, we use a high regional level of aggregation in the empirical analysis because it produces conservative estimates and allows for sorting within large regions.

Results indicate that a higher co-ethnic concentration increases school dropout rates of immigrant children. Moreover, growing up in regions with high ethnic concentrations impairs immigrant childrens' proficiency in the host-country language. Our findings indicate that parents' speaking proficiency in the host-country language is a key channel through which the effect of ethnic concentration on immigrant childrens' language proficiency operates. We also analyze potential effect heterogeneity of growing up in regions with high ethnic concentrations by country of birth and gender: we find that the negative effects of ethnic concentration on host-country language proficiency tend to be larger in magnitude for foreign born children.

We contribute to the literature by studying the long-run impact of ethnic concentration on the education attainment and language proficiency of immigrant children. Existing studies on the German Guest Worker Program, in contrast, have investigated the impact of ethnic concentration only on first-generation immigrants (see Danzer and Yaman (2013), Constant et al. (2013), Danzer and Yaman (2016)). Considering that integration is a long-term process, this paper evaluates the effects of co-ethnic concentration in a broader perspective across generations.

The current study is also related to the literature on educational achievement gaps between immigrant children and natives. While the share of immigrant students has increased

to 12.5% in OECD countries in 2015, immigrant students typically have poorer educational outcomes compared to native ones (see e.g., OECD (2016)). Algan et al. (2010) find that the educational achievement gaps tend to be lower for second-generation immigrants compared to first-generation immigrants in France, Germany and the UK. The findings of Dustmann et al. (2012) indicate that the achievement gap between immigrant and native children is significantly associated with the language spoken at home. We contribute to this literature by investigating in how far growing up in a region with a high ethnic concentration might contribute to the achievement gap between immigrant and native children.

This paper contributes to the large literature on the relationship between ethnic enclaves and economic outcomes of migrants. For example, while some studies within this literature strand have focused on the effects of the exposure to co-ethnics on immigrants' labour market outcomes (see e.g., Damm (2009), Beaman (2012), Battisti et al. (2016)), other studies have focused on the effects of social networks within ethnic enclaves on welfare participation (see e.g., Bertrand et al. (2000)). In general, these studies suggest that different economic effects arise within the context of ethnic enclaves.

More specifically, the present paper adds to the small economic literature on the effects of ethnic concentration on immigrant childrens' educational attainment. The early study of Cortes (2006), for example, finds that attending an enclave school is at most weakly associated with reading and math test scores. Using PISA data from Denmark, Jensen and Rasmussen (2011) find that attending a school with a high immigrant concentration is associated with lower reading and math skills for immigrant children. Greater emphasis on the identification of causal effects can be found in Åslund et al. (2011): they exploit exogenous variation from a refugee placement policy in Sweden to account for residential self-selection of immigrants. Their study is based on administrative data and focuses on education. In our study we combine administrative data with survey data which allows us to investigate language proficiency as an outcome variable and to shed more light on the underlying mechanisms of the effects of ethnic concentration. To the best of our knowledge, this paper is the first to investigate whether parents' proficiency in the host-country language mediates the effects of ethnic concentration on immigrant children.

Finally, we discuss our findings on the effects of ethnic concentration on immigrant childrens' educational attainment and language proficiency in the host-country in the light of the recent inflow of refugees to European countries.

The paper is organized as follows: Section 2 summarizes the key features of the German Guest Worker Program and describes the identification strategy. In Section 3, the data is introduced. We present our main results on the effects of ethnic concentration in Section 4. We further provide econometric evidence on potential mediating factors, the robustness of the results, and the heterogeneity of effects in this section as well. Section 5 discusses

the results with regards to their policy implications. Section 6 concludes.

2 The German Guest Worker Program as a Quasi-Experiment

The German Guest Worker program was one of the largest guest worker programs ever undertaken. West Germany started to recruit foreign guest workers to reduce the lack of available manpower at a time of rapid economic growth during the 1950s and 1960s. Guest worker treaties were initiated with Italy in 1955, Greece and Spain in 1960, Turkey in 1961, and Yugoslavia in 1968. Given that the recruitment was designed to attract unskilled workers within certain age brackets, the group of guest workers constituted a rather homogeneous immigrant population: the incoming guest workers were, on average, less educated than German workers. West Germany stopped the active recruitment of guest workers in 1973 due to an economic recession within the context of the oil crisis. After 1973, many guest workers stayed in Germany and brought their family members into the country.

The placement procedure of the German Guest Worker Program provides an opportunity to isolate the causal effect of ethnic concentration on the educational attainment and language skills of immigrant children: given that the initial job location was exogenous from the perspective of a guest worker[4], the placement procedure rules out differences in unobserved characteristics such as unobserved ability and other relevant characteristics of immigrant parents that might influence their childrens' educational achievement and language proficiency in the host-country language. The demand-driven placement during this large-scale immigration generated regional differences in ethnic concentrations: as can be seen in Figure 1, the distribution of ethnic concentrations in West Germany varied across ethnicities. For example, while the Spanish population tended to be concentrated in central West Germany, Yugoslavs were concentrated mostly in the southern regions. By including region fixed effects, we can control for systematic regional differences such as the strength of the regional economy, the regional share of migrants in the population and unobserved school quality. By including country-of-origin fixed effects, we can control for systematic differences between ethnicities such as linguistic distance or school quality in the country of origin. The presence of five different immigrant groups with different exposures to own ethnic concentration within the same region allows us to identify the effect of ethnic concentration.

While country of origin fixed effects and region fixed effects address many threats to validity, a potential threat to identification remains: parents may have sorted by unobserved characteristics into ethnic enclaves between the initial placement and the year for which

[4]For more details of the placement procedure, see Danzer and Yaman (2016).

their children report their educational attainment and language proficiency.[5] For example, if parents with relevant unfavorable characteristics moved to regions with high ethnic concentrations during this period and if these characteristics are unobservable and determine their childrens' educational attainment and host-country language proficiency, then the estimated coefficient of ethnic concentration would be upward biased.

Several facts support our identification strategy: previous studies within the context of the German guest worker programm (see Constant et al. (2013), Danzer and Yaman (2013), Danzer and Yaman (2016)) found no evidence of significant demographic differences between guest workers living in high concentration regions and those living in low concentration regions. This evidence against sorting may be surprising, given that other studies in different settings have found evidence of endogenous sorting of migrants or refugees into ethnic enclaves (see e.g. Åslund et al. (2011)).

However, the evidence against sorting is consistent with specific features of the German Guest Program: guest workers in our setting were restricted in their residential choice because the work permit required guest workers to stay with their first employer for at least three years and within the same occupation and region for eight years (Dahnen and Kozlowicz (1963)). In addition, the evidence against sorting is also consistent with the specific type of migration in our setting: compared to immigrant groups such as refugees, the guest workers in our setting tended to be well integrated in the labor market due to the demand driven recruitment. The guest worker gained also work experience in the host country since their arrival in Germany. Therefore, incentives to move into ethnic enclaves are lower compared to other immigrant groups because their potential to improve their labor market integration through ethnic networks is lower.

Against this background, it is not surprising that Schonwalder and Sohn (2009) find that immigrants' settlement structures in Germany still reflect the demand for labor in the 1960s and 1970s. Furthermore, the recent results of Glitz (2014) indicate that ethnic segregation has been quite stable across workplaces and residential locations over the period 1975 to 2008. In Section 3.3, we will provide further evidence in support of the identifying assumption by testing whether immigrant childrens' parents have sorted on observables such as their educational background into regions with high ethnic concentrations.

3 Data

We draw information from two datasets: individual-level data from the German-Socio-Economic Panel (GSOEP) and measures of ethnic concentrations based on two large em-

[5] Immigrant children of guest workers report the corresponding outcome variables from 1984 onwards.

ployee samples from the Institut für Arbeitsmarkt- und Berufsforschung (IAB).

3.1 Individual-level Data

We obtain information on guest workers and their children from the German-Socio-Economic Panel (GSOEP). The data provide detailed information on household and individual characteristics including language ability and educational attainment. While the GSOEP started in 1984, the level of geographic disaggregation of the first wave in 1984 is insufficient for merging with other data sources. Hence, we identify former guest workers and their residence region using the 1985 wave. In a second step, we link these immigrants of the 1985 wave to their children. We identify the children of mothers using the data set on the birth biography of female respondents.[6] Since there is no similar dataset on the children of fathers available for the guest worker sample, we identify the children of a male immigrant from their relationship to the mother.[7]

While it is not possible to directly identify guest worker from our data, we identify former guest workers by country of origin, year of immigration and age at migration. We exclude children whose parents migrated before the guest worker agreement.[8] In the main sample, we keep only children if at least one parent was older than 18 at immigration and arrived in Germany at the time of the guest worker program. We keep only children who were 13 or younger at migration because the focus of our paper is on the effects of growing up in ethnic enclaves. Finally, we base our analysis on children with non-missing values for both the schooling and the language variables.[9] Our final sample provides rich information on 942 immigrant children with either a Greek, Italian, Spanish, Turkish, or Yugoslavian background.[10]

We construct two binary outcomes of educational attainment: the outcome *Any School Degree* corresponds to one if an immigrant child obtained any type of school degree and zero if a child dropped out of school without a certificate. Our second outcome *At least Intermediate School Degree* corresponds to one if a child obtained an intermediate school degree or a higher secondary school degree and zero if otherwise.[11] The educational outcomes are

[6]The title of the dataset is BIOBIRTH. It is based on a comprehensive set of questions on the number, birth year, sex of the children and their relationship to the head of household.

[7]If a male immigrant lives in the mother's household in 1985 and is the husband or the spouse of the mother, we assume that he is the father of her children.

[8]The individual cut-off year depends on the country of origin.

[9]Given that we examine the impact of ethnic concentration on outcomes observed at age 16 or older, the main reason for the missing values is that some children didn't stay in the SOEP until reaching adulthood. The share of children with non-missing values of the language and schooling outcomes is similar in regions with low and regions with high ethnic concentration (see Table A1 in the Appendix).

[10]Given that the focus of our paper is on children with immigrant parents, we exclude 49 children with only one immigrant parent and non-missing values of the educational outcomes.

[11]While we could assign a value to most children in our sample, a small minority of children stated that

based on the latest available educational level in the SOEP.[12]

In addition, we examine two outcomes of immigrant childrens' host-country language proficiency: speaking and writing proficiency in German. Both language outcomes are self-reported and based on a five-point scale: very good, good, fairly, poorly to not at all. The corresponding questions are usually asked for the first time when the children are 17 or 18 years old. Our language outcomes correspond to the mean of the first two self-reported assessments in order to reduce the bias due to measurement errors being independent over time (see e.g. Solon (1992), Dustmann and Soest (2001)). We normalize childrens' language outcomes to have a mean of zero and standard deviation one.

Our data set contains a rich set of explanatory variables including demographic characteristics, variables of the individual immigration history, and the educational attainment and labor market outcomes of parents.[13] Furthermore, we construct variables that may be potential mediating factors of the effects of ethnic concentration: parents' speaking and writing abilities in German, parents' employment status and household income, parents' visits from Germans and a binary indicator if a child's' first friend was German. Parents' mediating factors are based on the average of the corresponding variables of mother and father.

3.2 Ethnic Concentration

We identify immigrant childrens' ethnicity based on their father's country of origin or nationality.[14] Following Danzer and Yaman (2016), we estimate measures for co-ethnic concentrations for five nationalities (Greek, Italian, Spanish, Turkish, and Yugoslav) at the regional level of the so-called *Anpassungsschichten* using the IAB employee sample of the year 1985.[15] Then we match the data on ethnic concentrations with the individual-level data on guest workers and their children. The main explanatory variable used in this paper is the log of the size of the ethnic community in the 1985 region of residence.

---

they had obtained another leaving certificate. We assumed that this school leaving certificate is equivalent to an intermediate school degree.

[12] If the last available educational level corresponds to "dropout no school degree" or "no school degree yet", we checked for school leaving certificates reported in previous years. In 5 cases, we corrected our educational outcome variables based on previous reported school leaving certificates.

[13] As usual for surveys, our data on guest workers and their children has missing value for some questions. Since our set of control variables is large, dropping all children with any missing value would substantially reduce sample size. We therefore impute missing values by using the mean of each control variable. Also we construct an indicator for each variable with missing values that corresponds to one for imputed values and zero otherwise. In all regressions, we include these indicators to ensure that the imputed data are not driving the results.

[14] Father's ethnicity is identical to mother's ethnicity for more than 99% of the immigrant children in our final sample.

[15] The IAB employee sample is a 2% random sample of the German employee population including also the recipients of social transfers.

We use the so-called *Anpassungsschichten* as a level of regional aggregation because it is a wide concept of ethnic enclaves: it allows not only for sorting of guest workers across city quarters within cities, but also for sorting across cities within large regions.[16] In West Germany, including West Berlin, there were 110 Anpassungsschichten in 1985 with an average region size of about half a million people. Hence, using this level of aggregation produces conservative results and strengthens the identification strategy.

3.3 Descriptives Statistics by Degree of Ethnic Concentration

In Table 1, we provide descriptive statistics on the main variables of guest worker children in our sample and their parents. Table 1 reports the mean levels of variables by degree of ethnic concentration at the immigrant children level. The treatment intensity corresponds to low (high) ethnic concentration, if the individual ethnic concentration is below (above) the ethnicity-specific median of the share of ethnic concentration in the year 1985. The upper panel of Table 1 reveals that immigrant children in high concentration areas are less likely, on average, to obtain any school degree. While immigrant children living in high concentration regions have a lower speaking and writing proficiency in the host-country language, the corresponding differences in unconditional means are not significant at conventional levels.

Sorting of parents with relevant unobserved characteristics into ethnic enclaves might be a primary threat to our identification strategy. Table 1 aims to test whether immigrant childrens' parents have sorted on observables such as their educational attainment into ethnic enclaves. Testing parents' sorting based on unobservables is possible, if there are variables of parents which are highly correlated with unobserved characteristics that might be related to endogenous sorting. We consider variables on parents' educational background and moving patterns as particularly good measures of potential unobserved factors. If parents' educational background is different in regions with high ethnic concentration compared to parents' educational background in regions with low ethnic concentrations, our key identifying assumption that immigrant childrens' parents didn't endogenously sort might be violated.

We find no evidence of sorting: most importantly, variables of parents' educational background in low concentration regions are similar to the ones of parents in high concentration regions. In addition, there are no significant differences in parents' demographic characteristics and their moving patterns. We also find no evidence for significant demographic differences between immigrant children residing in high vs. low ethnic concentration regions.

[16] Figure 1 shows the regional level of aggregation used in this paper as well as the distribution of ethnic concentrations across West Germany and West Berlin.

Given that regions with high ethnic concentration had an excess labour demand, it makes sense that unemployment rates of the year 1985 are significantly lower in these regions (see the lower panel of Tabel 1). Within this context, it is not surprising that the household income of parents living in high concentration areas is significantly higher and that fathers in these regions are significantly less unemployed. On the other hand, these differences in parents' labour market outcomes might also reflect that ethnic networks within enclaves might increase immigrant's opportunities in the labor market (see e.g., Edin et al. (2003)).

All remaining t-tests between parents residing in high vs. low concentration areas are insignificant. Overall, we interpret this as evidence in favor of our identifying assumption of no endogenous sorting of parents after their initial exogenous placement in Germany.

4 Empirical Analysis

We estimate the effects of ethnic concentration on immigrant childrens' educational attainment and host-country language proficiency using OLS regression specifications of the form:

$$y_i = \beta_0 + \beta_1 EC_i + X_i'\gamma + P_i'\delta + M_i'\lambda + \tau_j + \sigma_c + \varepsilon_i \tag{1}$$

where the dependent variable y_i is one of the four outcomes of children i, EC_i is the key explanatory variable size of ethnic group in the 1985 region of residence (entered in logs), X is a vector of child characteristics (birth-cohort fixed effects (2-year intervals), gender, age at migration dummies), P is a vector of parent characteristics (year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, dummies for the number of mother's children), the τ_j's denote region fixed effects and the σ_c's denote country-of-origin fixed effects. ε_i is an individual error term. By including country-of-origin fixed effects we can control for systematic differences between immigrant groups such as linguistic or cultural distance. By including region fixed effects we exploit only variation in ethnic concentrations that is not systematic across regions. We identify the effect of ethnic concentration β_1 from the presence of different immigrant groups with different exposures to own ethnic concentration within the same region. M is an optional vector of potential mediating factors such as parents' language proficiency in the host-country language. We cluster standard errors at the *Region-Nationality* level. All regressions with childrens' outcomes of language proficiency contain dummies for age at assessment.

We begin by analyzing the direct effects of the size of the ethnic community on immigrant childrens' educational achievement and host-country language proficiency in Section

4.1. We then turn to potential channels of the effects in Section 4.2. To investigate the robustness of the results, we also estimate the effects of ethnic concentration on immigrant childrens' language proficiency using random effects models in Section 4.3. In Section 4.4, we shed light on the extent to which the effects of ethnic concentration are heterogeneous across country of birth and gender.

4.1 The Impact of Ethnic Concentration on Educational Achievement and Host-Country Language Proficiency

We begin this section by investigating the effects of the regional size of the ethnic community on immigrant childrens' educational achievement: Table 2 reports the regression results with the binary indicators *Any School Degree* (columns 1-2) and *At least Intermediate School Degree* (columns 3-4) being the outcome variables. The size of the ethnic community is negatively associated with the probability of obtaining a school degree. The significant and negative effect of ethnic concentration on the probability of obtaining a school degree persists after controlling for region fixed effects, ethnicity fixed effects, child controls and parent controls: living in a region with a 170 percent larger own ethnic community reduces the probability of obtaining a school degree by 6.5%-points (column 2). We do not find significant effects of ethnic concentration on the probability of obtaining at least an intermediate school degree (columns 3-4). Overall, these findings are consistent with the descriptive evidence in Table 1.

In Table 3, we examine the impact of the size of the ethnic group on immigrant childrens' host-country language proficiency: columns 1-2 of Table 3 indicate that growing up in regions with high ethnic concentrations reduces childrens' speaking proficiency in German. The size of the ethnic community is also negatively associated with the writing proficiency in German (columns 3-4). Again, including individual control variables has little impact on the qualitative conclusions. An increase in the size of the ethnic group by around 170 percent decreases speaking proficiency in German by 0.309 standard deviations (column 2).

4.2 Mediating Factors of the Effects of Ethnic Concentration

In this section, we investigate to what extent potential mediating factors can explain the negative effects of ethnic concentration on immigrant childrens' educational attainment and host-country language proficiency. While previous studies on the German Guest Worker program found negative effects on language proficiency (Danzer and Yaman (2016)) and social integration of guest workers (Danzer and Yaman (2013)), these effects may be mediating

factors of the impact of ethnic concentration on their children. In particular, intergenerational aspects of language proficiency may partly explain the effects of ethnic concentration on immigrant children within the context of this paper. On the other hand, ethnic networks within ethnic enclaves may improve parents' labor market outcomes (see e.g., Edin et al. (2003)). This in turn may have beneficial effects on childrens' educational and language outcomes.

To better understand the effects of ethnic concentration, we investigate several competing hypothesis on the potential channels of the effects of ethnic concentration on immigrant childrens' educational attainment and language proficiency in the host-country language: parents' speaking and writing abilities in German, a binary dummy indicating if the first friend was German, a binary indicator which equals one if the parents received visits from Germans, and parents' employment status. We do so by examining whether the inclusion of a potential mediating factor as an additional explanatory variable affects the coefficient of ethnic concentration.

In Table 4, we investigate potential channels of the effects of ethnic concentration on the probability of obtaining any school degree by stepwise controlling for them. In order to reduce time-independent measurement errors in parents' language variables, we instrument parents' self-reported language ability with the corresponding leads and lags (see Dustmann and Soest (2001)). More precisely, we instrument the second available observation of parents' language skills with the corresponding leads and lags (column 2 and 3).[17] None of the potential mediating factors can significantly reduce the effect of ethnic concentration on this educational outcome (columns 2-5, Table 4). Controlling for the potential channel of parents' labour market outcomes does not significantly increase the effect of ethnic concentration.

Regarding the effects of ethnic concentration on immigrant childrens' host-county language proficiency, we test specific channels of causation in Table 5. While immigrant childrens' speaking proficiency in German is the outcome in Panel A, their writing proficiency in German is the outcome in Panel B. Since we use parents' language proficiency at the time of the childrens' first reported speaking proficiency in German, we focus on children for which parents' language proficiency is available at that time.[18] Regarding both language outcomes, the negative effect of ethnic concentration becomes smaller and less significant after including parents' instrumented speaking proficiency (column 2). Interestingly, including parents' writing proficiency does not substantially reduce the negative effect of ethnic concentration (column 3). Including the variable *First Friend German* and the vari-

[17]The second available observation of parents' speaking ability refers to the year 1985 in most cases.

[18]To increase the sample size, we instrument parents' current language proficiency with itself, if a lead or lag is not available in the data. While this increases the representativeness of the sample, it does not affect key results.

able on visits from Germans only marginally changes the effects of ethnic concentration (columns 4 and 5). Overall, these results in Table 5 suggest that parents' speaking proficiency is a significant mediating factor of the negative effect of ethnic concentration on childrens' speaking and writing proficiency in the host-country language.

4.3 Robustness: Random Effects Model

To investigate the robustness of the effects of ethnic concentration on immigrant childrens' language skills, we also estimate random effects models (see Table 6 and Table 7). These are based on the full sample of children (as in Table 3). While our main results in Section 4.1 and 4.2 are based on a cross-section of the immigrant children in our sample, the advantage of the random effects models is that we can use on average five observations of language proficiency for each child resulting in a substantially larger sample size. Each observation is an immigrant children-year pair based on self-reported language proficiency in the years 1984 - 1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005.[19] The main sample includes around 4600 children-year observations. The key explanatory variable is again the regional *Size of Ethnic Group in 1985*.

Our panel data findings in Table 6 suggest significant negative effects of ethnic concentration on immigrant childrens' speaking proficiency (columns 1-2) and writing proficiency (columns 3-4) across all specifications.

In a second step, we investigate potential mediating factors of this negative effect of ethnic concentration (Table 7). In Panel A, immigrant childrens' speaking proficiency in German is the dependent variable. In Panel B, immigrant childrens' writing proficiency is the outcome variable. To deal with the problem of random measurement errors in self-reported speaking abilities, we again instrument parents' language proficiency with leads and lags of their self-reported language proficiency (see Dustmann and Soest (2001)). For this robustness check, we focus on children-year pairs for which parents' language proficiency is available at the time their children report their own language proficiency.[20] As can be seen in Table 7, the inclusion of parents' speaking abilities substantially reduces the negative effect of ethnic concentration (column 2 in Panel A and Panel B). The effect of ethnic concentration on immigrant childrens' speaking proficiency becomes insignificant. Other potential mediating factors than parents' language proficiency (columns 4-6) do not

[19] Our panel data set is unbalanced for mainly two reasons: first, while questions on language proficiency are usually asked for the first time when the children are 17 or 18 years old, not all the children of our sample are 17 or older in 1984. The year 1984 corresponds to the first wave of the German Socio- Economic Panel. Second, immigrant children of our sample may stay in the SOEP until different waves: e.g., while some individuals left the SOEP in 2001, others stayed until the 2005 wave.

[20] The corresponding leads and lags of parents' language proficiency are for some children-year observations not available. Regarding these observation, we simply instrument with parents' current language proficiency. While this increases sample size and makes the sample more representative, it does not affect key results.

play such a significant role here. These findings confirm the results in Section 4.2: parents' speaking proficiency is a key mediating factor of the negative effect of ethnic concentration on immigrant childrens' proficiency in the host-country language.

4.4 Heterogeneity of Effects

In Table 8 we shed light on the extent to which the effects of ethnic concentration are heterogeneous across immigrant childrens' country of birth (Panel A) and gender (Panel B). We include interaction terms to investigate differential effects with respect to these characteristics. Interestingly, the negative effects of ethnic concentration on the language proficiency in German are significantly stronger for foreign born children and less pronounced for migrant children already born in Germany (columns 3-4). Regarding the outcome at least intermediate school degree, the effects of ethnic concentration significantly differ by country of birth. While we find that the negative effect of ethnic concentration on the probability to obtain any school degree is less pronounced for girls, we do not find other significant gender differences in the effects of ethnic concentration.

5 Policy Implications

In this section, we discuss the experience of the guest worker program in light of the recent inflow of refugees to European countries. Germany alone received 1,221,665 asylum seekers between 2015 and 2016.[21] Germany's Integration Act, enacted in August 2016, requires asylum seekers to stay at an assigned place of residence for 3 years. [22] The aim of this act is to foster sustainable integration of immigrants and prevent social segregation. While the current refugee crisis differs from the German Guest Worker Program (in particular, current asylum seekers have migrated for different reasons and are not assigned as laborers to specific firms), refugees today and guest workers of the 1960s and 1970s share similar characteristics that are crucial for integration success: low education levels and low German-language skills. Therefore, our results on the ethnic concentration impacts on the children of guest workers also addresses the current debate as to whether asylum seekers should be assigned fixed places of residence which aim at distributing refugees equally across the country, thereby preventing the development of ethnic enclaves.

[21] Source: Eurostat, "Asylum and first time asylum applicants - annual aggregated data (rounded) (tps00191)", updated 11 November 2016.

[22] Exceptions are refugees who are in vocational training or employment. For more details: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGB&jumpTo=bgbl116s1939.pdf .

6 Conclusion

This paper has exploited the exogenous placement of guest workers during the German Guest Worker Program to estimate the effect of ethnic concentration on immigrant childrens' educational attainment and proficiency in the host-country language. Consistent with previous studies within the context of the German Guest Worker Program, we found no evidence that immigrant childrens' parents have endogenously sorted on observables such as their educational attainment into regions with high ethnic concentrations.

Results indicate that growing up in regions with high ethnic concentrations has negative effects on immigrant childrens' educational attainment in the host-country. More specifically, ethnic concentration reduces the probability that immigrant children obtain any school degree. Consistent with this finding, our results suggest negative effects of ethnic concentration on immigrant childrens' proficiency in the host-country language. We show that parents' lower speaking proficiency explains much of the negative effect of ethnic concentration on immigrant children' proficiency in the host-country language. This finding may help to better understand intergenerational aspects of language proficiency within ethnic enclaves. Our findings also suggest that the effects of ethnic concentration on host-country language proficiency are particularly salient for foreign born immigrant children.

Overall, these findings may be particularly important within the context of education and host-country language proficiency being considered key factors for immigrant childrens' social and labor market integration, as well as upward mobility.

References

ALDASHEV, ALISHER, JOHANNES GERNANDT, and STEPHAN L. THOMSEN (2009), "Language usage, participation, employment and earnings. Evidence for foreigners in West Germany with multiple sources of selection". *Labour Economics* 16(3), pp. 330–341.

ALGAN, YANN, CHRISTIAN DUSTMANN, ALBRECHT GLITZ, and ALAN MANNING (2010), "The economic situation of first and Second-Generation immigrants in France, Germany and the United Kingdom". *Economic Journal* 120(542), F4–F30.

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), pp. 67–95.

ÅSLUND, OLOF and PETER FREDRIKSSON (2009), "Peer Effects in Welfare Dependence". *Journal of Human Ressources* 44(3), pp. 798–825.

BATTISTI, MICHELE, GIOVANNI PERI, and AGNESE ROMITI (2016), *Dynamic Effects of Co-Ethnic Networks on Immigrants' Economic Success*. NBER Working Paper No. 22389.

BEAMAN, LORI A. (2012), "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S." *Review of Economic Studies* 79(1), pp. 128–161.

BERTRAND, MARIANNE, ERZO LUTTMER, and SENDHIL MULLAINATHAN (2000), "NETWORK EFFECTS AND WELFARE CULTURES". *The Quarterly Journal of Economics* (August), pp. 1019–1055.

BORJAS, GEORGE J. (1995), "Ethnicity, Neighborhoods, and Human Capital Externalities". *The American Economic Review* 85(3), pp. 365–390.

CHETTY, RAJ and NATHANIEL HENDREN (2016), *The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects*. Tech. rep. No. 23001.

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), pp. 855–902.

CONSTANT, AMELIE F., SIMONE SCHÜLLER, and KLAUS F. ZIMMERMANN (2013), "Ethnic Spatial Dispersion and Immigrant Identity". *IZA DP No. 7868*.

CORTES, KALENA E. (2006), "The effects of age at arrival and enclave schools on the academic performance of immigrant children". *Economics of Education Review* 25(2), pp. 121–132.

CUTLER, DAVID M. and EDWARD L. GLAESER (1997), "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics* 112(3), pp. 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), pp. 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(2), pp. 311–325.

— (2016), "Ethnic concentration and language fluency of immigrants: Evidence from the guest-worker placement in Germany". *Journal of Economic Behavior and Organization* 131, pp. 151–165.

DUSTMANN, CHRISTIAN and FRANCESCA FABBRI (2003), "Language proficiency and labour market performance of immigrants in the UK". *Economic Journal* 113(489), pp. 695–717.

DUSTMANN, CHRISTIAN, TOMMASO FRATTINI, and GIANANDREA LANZARA (2012), "Educational achievement of second-generation immigrants: An international comparison". *Economic Policy* 27(69), pp. 143–185.

DUSTMANN, CHRISTIAN, STEPHEN MACHIN, and UTA SCHNBERG (2010), "Ethnicity and educational achievement in compulsory schooling". *Economic Journal* 120(546).

DUSTMANN, CHRISTIAN and ARTHUR *van* SOEST (2001), "Language fluency and earnings: Estimation with misclassified language indicators". *Review of Economics and Statistics* 83(November), pp. 663–674.

EDIN, PER-ANDERS, PETER FREDRIKSSON, and OLOF Ã...SLUND (2003), "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from A Natural Experiment". *Quarterly Journal of Economics* 118(1), pp. 329–357.

GLITZ, ALBRECHT (2014), "Ethnic segregation in Germany". *Labour Economics* 29, pp. 28–40.

GRÖNQVIST, HANS (2006), *Ethnic enclaves and the attainments of immigrant children*.

JENSEN, PETER and A. W. RASMUSSEN (2011), "The effect of immigrant concentration in schools on native and immigrant children's reading and math skills". *Economics of Education Review* 30(6), pp. 1503–1515.

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.

SCHONWALDER, K. and J. SOHN (2009), "Immigrant Settlement Structures in Germany: General Patterns and Urban Levels of Concentration of Major Groups". *Urban Studies* 46(7), pp. 1439–1460.

SOLON, GARY (1992), "Intergenerational Income Mobility in the United-states". *American Economic Review* 82(3), pp. 393–408.

Table 1: Descriptive Statistics by Degree of Ethnic Concentration

Variable	Low EC	High EC	Diff.	P-Value	Obs.
Outcomes (Children)					
Any School Degree	0.95	0.90	0.04	0.04	886
At least Intermediate School Degree	0.40	0.39	0.01	0.88	886
Speaking Proficiency	0.06	-0.06	0.12	0.25	920
Writing Proficiency	0.07	-0.06	0.13	0.18	920
Children					
Male	0.54	0.57	-0.03	0.41	942
Year of Birth	1971.5	1970.7	0.76	0.22	942
Age at Migration	2.81	3.19	-0.37	0.35	942
Born in Germany	0.55	0.52	0.03	0.54	942
Turk	0.44	0.40	0.04	0.70	942
Yugoslaw	0.17	0.18	-0.00	0.99	942
Italian	0.14	0.17	-0.03	0.67	942
Spaniard	0.10	0.11	-0.01	0.84	942
Greek	0.15	0.15	0.00	0.99	942
Mothers					
Year of Birth	1944.3	1943.5	0.80	0.35	942
Year of Immigration	1970.9	1970.1	0.82	0.24	942
Age at Migration	26.57	26.60	-0.02	0.98	942
Education in Country of Origin					
No Schooling	0.20	0.20	-0.00	0.95	942
Incomplete Compulsory Schooling	0.43	0.41	0.02	0.74	942
At least Compulsory Schooling	0.37	0.39	-0.02	0.78	942
Years of Education	8.17	8.21	-0.04	0.82	942
Never Moved	0.16	0.13	0.03	0.39	942
Moved into Flat before 1976	0.39	0.34	0.04	0.39	942
Nr Kids	3.98	3.72	0.26	0.33	942
Household Income (1984-1986)	1694.16	1835.52	-141.36	0.07	942
Unemployed (1984-1986)	0.07	0.05	0.02	0.19	942
Not Employed (1984-1986)	0.60	0.52	0.08	0.19	942
Fathers					
Year of Birth	1940.0	1939.9	0.08	0.91	942
Year of Immigration	1968.2	1967.7	0.41	0.41	942
Age at Migration	28.16	27.80	0.36	0.54	942
Education in Country of Origin					
No Schooling	0.09	0.11	-0.03	0.47	942
Incomplete Compulsory Schooling	0.28	0.29	-0.02	0.77	942
At least Compulsory Schooling	0.37	0.39	-0.02	0.78	942
Years of Education	9.06	9.06	-0.00	1.00	942
Never Moved	0.05	0.04	0.01	0.54	942
Moved into Flat before 1976	0.39	0.34	0.04	0.39	942
Household Income (1984-1986)	1694.16	1835.52	-141.36	0.07	942
Unemployed (1984-1986)	0.11	0.04	0.06	0.02	942
Not Employed (1984-1986)	0.10	0.07	0.02	0.36	942
For Comparison					
Regional Unemployment Rate 1985	0.11	0.08	0.03	0.00	942
Number of all Observations	447	495			942

Notes: Table reports the means of variables by degree of ethnic concentration. Low ("Low EC") and high ethnic concentrations ("High EC") are split at the ethnicity-specific median of the share of ethnic concentration in the year 1985. P-values for two-sided tests reported in column "P-Value". During the calculation of P-values, standard errors were clustered by region-nationality level. Outcomes: *Any School Degree*: binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise; *At least Intermediate School Degree*: binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 otherwise; *Speaking Proficiency* (*Writing Proficiency*), generated from self-assessed speaking (writing) ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; The variables Household Income (1984-1986), Unemployed (1984-1986), and Not Employed (1984-1986) are means over 3 years (1984-1986). The F-statistic for joint significance in regressing the high-concentration dummy on the individual characteristics is: 0.44, p-value 0.8928 (*children*); 1.13, p-value 0.3354 (*mothers*); 3.25, p-value 0.0010 (*fathers*). Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Bundesagentur für Arbeit.

Table 2: Effect of Ethnic Concentration on Educational Attainment

Dep. Var.:	Any School Degree		At least Intermediate School Degree	
	(1)	(2)	(3)	(4)
Size of Ethnic Group in 1985	-0.080***	-0.065***	-0.074	-0.007
	(0.019)	(0.022)	(0.052)	(0.048)
Region Fixed Effects	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes
Child Controls	No	Yes	No	Yes
Parent Controls	No	Yes	No	Yes
Observations	886	886	886	886
Adjusted R^2	0.03	0.06	0.10	0.21

Notes: OLS regressions. Dependent variables: *Any School Degree* (Columns 1-2): binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise; *At least Intermediate School Degree* (Columns 3-4): binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 otherwise. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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 Host-Country Language Proficiency

Dep. Var.:	Speaking Proficiency		Writing Proficiency	
	(1)	(2)	(3)	(4)
Size of Ethnic Group in 1985	-0.304***	-0.309***	-0.241***	-0.241**
	(0.102)	(0.110)	(0.089)	(0.097)
Year of Assessment, Dummies	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes
Child Controls	No	Yes	No	Yes
Parent Controls	No	Yes	No	Yes
Observations	920	920	920	920
Adjusted R^2	0.17	0.23	0.18	0.23

Notes: OLS regressions. Dependent variables: *Speaking Proficiency* (Columns 1-2), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; *Writing Proficiency* (Columns 3-4), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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: Mediating Factors - Effect of Ethnic Concentration on Educational Attainment

Dep. Var.:	Any School Degree					
	(1)	(2)	(3)	(4)	(5)	(6)
Size of Ethnic Group in 1985	-0.066***	-0.065***	-0.066***	-0.060***	-0.069***	-0.066***
	(0.022)	(0.021)	(0.020)	(0.022)	(0.022)	(0.023)
Speaking Abilities, Parents, IV leads + lags		0.041**				
		(0.019)				
Writing Abilities, Parents, IV leads + lags			0.002			
			(0.020)			
First Friend German				0.026		
				(0.021)		
Visits From Germans, Parents					-0.004	
					(0.027)	
Unemployed (1984-1986), Parents						-0.096
						(0.098)
Observations	885	885	885	885	885	885
Adjusted R^2	0.06	0.05	0.06	0.07	0.06	0.06
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Child Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Income + Household Income Squared	No	No	No	No	No	Yes

Notes: OLS regressions (Columns 1, 4-6); IV estimates (Columns 2-3), using leads and lags of parents' speaking (writing) proficiency as instruments. Dependent variable: *Any School Degree*: binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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 Language Proficiency

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Speaking Proficiency in German						
Size of Ethnic Group in 1985	-0.304***	-0.192*	-0.289***	-0.304***	-0.284**	-0.304***
	(0.114)	(0.106)	(0.108)	(0.112)	(0.110)	(0.114)
Speaking Abilities, Parents, IV Leads + Lags		0.495***				
		(0.093)				
Writing Abilities, Parents, IV Leads + Lags			0.193**			
			(0.077)			
First Friend German				0.332***		
				(0.080)		
Visits From Germans, Parents					0.266**	
					(0.110)	
Unemployed (1984-1986), Parents						-0.786**
						(0.386)
Observations	882	882	882	882	882	882
Adjusted R^2	0.22	0.20	0.21	0.24	0.22	0.22
Panel B: Writing Proficiency in German						
Size of Ethnic Group in 1985	-0.186*	-0.068	-0.164*	-0.182*	-0.173*	-0.188*
	(0.098)	(0.095)	(0.093)	(0.095)	(0.095)	(0.098)
Speaking Abilities, Parents, IV Leads + Lags		0.516***				
		(0.092)				
Writing Abilities, Parents, IV Leads + Lags			0.284***			
			(0.078)			
First Friend German				0.313***		
				(0.076)		
Visits From Germans, Parents					0.167	
					(0.120)	
Unemployed (1984-1986), Parents						-0.580
						(0.386)
Observations	882	882	882	882	882	882
Adjusted R^2	0.22	0.18	0.21	0.24	0.22	0.22
Control variables in Panels A + B						
Year of Assessment, Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Child Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Income + Household Income Squared	No	No	No	No	No	Yes

Notes: OLS regressions (Columns 1, 4-6); IV estimates (Columns 2-3), using leads and lags of parents' speaking (writing) proficiency as instruments. Dependent variables: *Speaking Proficiency* (Panel A), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; *Writing Proficiency* (Panel B), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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: Random Effects Model: Effect of Ethnic Concentration on Host-Country Language Proficiency

Dep. Var.:	Speaking Proficiency		Writing Proficiency	
	(1)	(2)	(3)	(4)
Size of Ethnic Group in 1985	-0.152*	-0.135*	-0.130*	-0.125*
	(0.083)	(0.078)	(0.073)	(0.069)
Year of Assessment, Dummies	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes
Child Controls	No	Yes	No	Yes
Parent Controls	No	Yes	No	Yes
Observations	4613	4613	4603	4603
R^2 overall	0.17	0.26	0.18	0.29

Notes: Random Effects Model. Dependent variables: *Speaking Proficiency* (Columns 1-2), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; *Writing Proficiency* (Columns 3-4), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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: Random Effects Model: Effect of Ethnic Concentration on Host-Country Language Proficiency

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Speaking Proficiency in German						
Size of Ethnic Group in 1985	-0.172**	-0.038	-0.140	-0.166**	-0.163*	-0.178**
	(0.085)	(0.095)	(0.090)	(0.083)	(0.084)	(0.084)
Speaking Abilities, Parents, IV leads + lags		0.408***				
		(0.059)				
Writing Abilities, Parents, IV leads + lags			0.307***			
			(0.065)			
First Friend German				0.233***		
				(0.062)		
Visits from Germans, Parents					0.084*	
					(0.046)	
Unemployed (1984-1986), Parents						-0.719**
						(0.294)
Observations	3959	3959	3959	3959	3959	3959
R^2 overall	0.26	0.27	0.26	0.27	0.26	0.27
Panel B: Writing Proficiency in German						
Size of Ethnic Group in 1985	-0.106	0.004	-0.080	-0.101	-0.098	-0.108
	(0.075)	(0.093)	(0.091)	(0.073)	(0.075)	(0.074)
Speaking Abilities, Parents, IV leads + lags		0.336***				
		(0.058)				
Writing Abilities, Parents, IV leads + lags			0.263***			
			(0.066)			
First Friend German				0.259***		
				(0.057)		
Visits from Germans, Parents					0.075	
					(0.048)	
Unemployed (1984-1986), Parents						-0.642**
						(0.300)
Observations	3954	3954	3954	3954	3954	3954
R^2 overall	0.29	0.30	0.30	0.31	0.30	0.30
Control variables in Panels A + B						
Year of Assessment, Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Child Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household Income + Household Income Squared	No	No	No	No	No	Yes

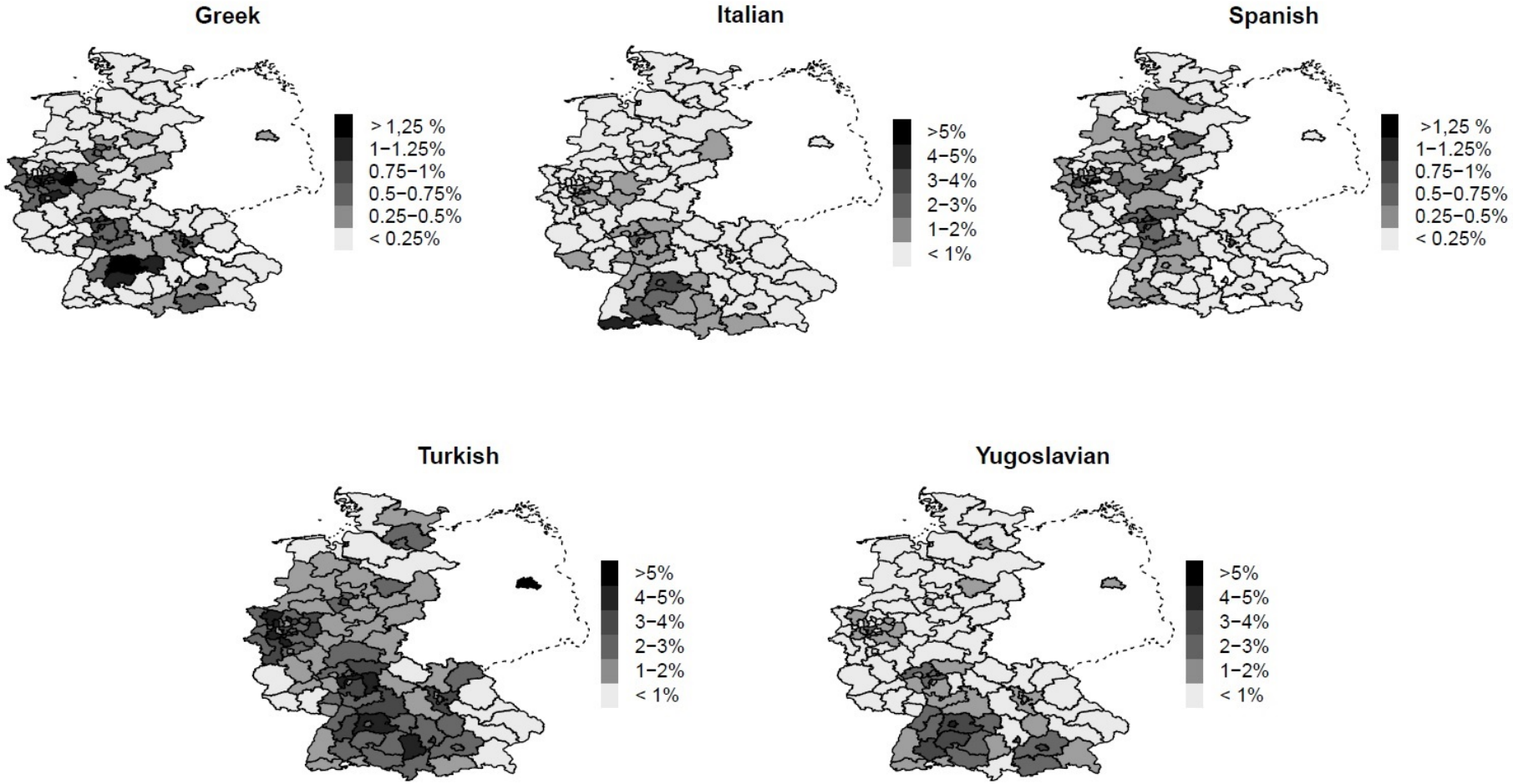
Notes: Random Effects Model. Dependent variables: *Speaking Proficiency* (Panel A), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; *Writing Proficiency* (Panel B), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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: Heterogeneity of Effects

Dep. Var.:	Any School Degree (1)	At Least Intermediate Degree (2)	Speaking Proficiency (3)	Writing Proficiency (4)
Panel A: by Country of Birth				
Size of Ethnic Group in 1985	-0.069***	-0.033	-0.354***	-0.287***
	(0.025)	(0.050)	(0.114)	(0.098)
Size of Ethnic Group * Born in Germany	0.008	0.054***	0.095**	0.100***
	(0.020)	(0.018)	(0.040)	(0.036)
Observations	886	886	920	920
Adjusted R^2 0.06	0.21	0.23	0.23	0.23
Panel B: by Gender				
Size of Ethnic Group in 1985	-0.074***	0.009	-0.313***	-0.207*
	(0.023)	(0.050)	(0.114)	(0.109)
Size of Ethnic Group * Female	0.022*	-0.037	-0.006	-0.041
	(0.012)	(0.031)	(0.045)	(0.049)
Observations	886	886	920	920
Adjusted R^2	0.06	0.21	0.25	0.24
Control variables in Panels A + B				
Year of Assessment, Dummies	No	No	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes
Child Controls	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variables: *Any School Degree* (Column 1): binary indicator that equals 1 if individual obtained any type of school degree and 0 otherwise; *At least Intermediate School Degree* (Column 2): binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 otherwise; *Speaking Proficiency* (Column 3), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one; *Writing Proficiency* (Column 4), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one. *Size of Ethnic Group in 1985* is the log of individuals of the same ethnicity in the region of residence in the year 1985. *Child Controls* contain birth-cohort fixed effects (2-year intervals), gender, and age at migration dummies. *Parent Controls* contain the following variables for father and mother: year of birth, fixed effects for arrival cohort (2-year intervals), education dummies, binary indicators for incomplete compulsory schooling and at least compulsory schooling in country of origin, and dummies for the number of mother's children. Standard errors, clustered at the region-nationality 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 1: Ethnic Concentrations across West Germany and Berlin



Source: IAB 1985. Own calculations.

Appendices

Table A1: The Share of Children with Non-Missing Variables by Degree of Ethnic Concentration

	Low EC	High EC	Diff.	P-Value	Obs.
Info School Degree	0.70	0.69	0.01	0.86	1274
Info Language Proficiency	0.72	0.73	-0.01	0.80	1274

Notes: This Table reports the means of variables by degree of ethnic concentration. Low ("Low EC") and high ethnic concentrations ("High EC") are split at ethnic-specific median concentrations of the year 1985. Binary indicators *Info School Degree* and *Info Language Proficiency* equal to 1 if information on the corresponding outcome is available in the SOEP data. The main reason for the missing values is that some children didn't stay in the SOEP until reaching adulthood. P-values for two-sided tests reported in column "P-Value". During the calculation of P-values, standard errors were clustered by region-nationality level. Data sources: German Socio-Economic Panel (SOEP), Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Table A2: Variable Definitions (SOEP-Data)

Variable	Description
Outcomes (Children)	
Any School Degree	Binary indicator that equals 1 if individual obtained any type of school degree and 0 if otherwise, based on the latest available educational level
At least Intermediate School Degree	Binary indicator that equals 1 if individual obtained at least an intermediate school degree and 0 if otherwise, based on the latest available educational level
Speaking Proficiency	Generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one, *Cross-Section Analysis* (Tables 3, 5): outcome based on the mean of the first two self-reported assessments, *Panel Data Analysis* (Tables 6, 7): each observation is an immigrant children-year pair 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 (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), normalized to have a mean zero and standard deviation one, *Cross-Section Analysis* (Tables 3, 5): outcome based on the mean of the first two self-reported assessments, *Panel Data Analysis* (Tables 6, 7): each observation is an immigrant children-year pair based on self-reported language proficiency in the years 1984 - 1987, 1989, 1991, 1993, 1997, 1999, 2001, 2003, and 2005

Continued on next page

Table A2 – *Continued from previous page*

Variable	Description
Variables of Children	
Country-of-Origin Dummies (Turk, Yugoslav, Italian, Spaniard, Greek)	Binary indicators based on father's country of origin or nationality (father's nationality is identical to the mother's nationality for more than 99% of the immigrant children in the analytic sample)
Variables of Parents	
Variables on Education in Country of Origin (No Schooling, Incomplete Compulsory Schooling, At least Compulsory Schooling)	Binary indicators based on survey question "Obtained School Degree Outside Germany", based on survey year 1985
Years of Education	Amount of education or training (in years), generated variable by SOEP, based on survey year 1985
Never Moved	Binary indicator equal to 1 if the individual 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, based on survey year 1985
Moved into Flat before 1976	Binary indicator equal to 1 if the household moved into the dwelling before 1976, based on survey year 1985
Household Income (1984-1986)	The mean of parents' household income over three years, based on survey years 1984-1986
Unemployed (1984-1986)	The mean of an unemployment dummy over three years, based on survey years 1984-1986
Not Employed (1984-1986)	The mean of an not employed dummy over three years, based on survey years 1984-1986
Mediating Factors	
Speaking Abilities, Parents, IV Leads + Lags	Parents' self-reported speaking ability instrumented with the corresponding leads and lags to reduce measurement error (see Dustmann and Soest (2001)), generated from self-assessed speaking ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), based on the average of self-reported speaking ability of the mother and the father, normalized to have a mean zero and standard deviation one, Table 4: the second available observation of parents' language skills is instrumented with the corresponding leads and lags, Table 5: parents' language proficiency at the time of the childrens' first reported language proficiency in German is instrumented with the corresponding leads and lags, Table 7: parents' language proficiency at the time of the childrens' reported language proficiency in German is instrumented with the corresponding leads and lags

Continued on next page

Table A2 – *Continued from previous page*

Variable	Description
Writing Abilities, Parents, IV Leads + Lags	Parents' self-reported writing ability instrumented with the corresponding leads and lags to reduce measurement error (see Dustmann and Soest (2001)), generated from self-assessed writing ability in German (very well = 5, good = 4, fairly = 3, poorly = 2, not at all = 1), based on the average of self-reported writing ability of the mother and the father, normalized to have a mean zero and standard deviation one, Table 4: the second available observation of parents' language skills is instrumented with the corresponding leads and lags, Table 5: parents' language proficiency at the time of the childrens' first reported language proficiency in German is instrumented with the corresponding leads and lags, Table 7: parents' language proficiency at the time of the childrens' reported language proficiency in German is instrumented with the corresponding leads and lags
First Friend German	Binary indicator equal to 1 if a childs' first friend was German
Visits from Germans, Parents	Variable based on the average of the following variable of the mother and the father: a binary indicator equal to one if she or he received visits from Germans, Table 4: variable refers to the survey year 1985, Table 5: variable refers to the year of the childrens' first reported language proficiency, Table 7: variable refers to the year of the childrens' reported language proficiency
Unemployed (1984-1986), Parents	Variable based on the average of the following variable of the mother and the father: The mean of an unemployment dummy over three years, based on survey years 1984-1986
Household Income (1984-1986), Parents	The mean of parents' household income over three years, based on survey years 1984-1986

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

Table A3: Regional Variables

Variable	Description
Size of Ethnic Group in 1985	Log of individuals of the same ethnicity in the region of residence in the year 1985, Level of aggregation: *Anpassungsschicht*, estimated ethnic concentrations based on a two percent sample of the German employee population (incl. recipients of social transfers) from the Institut für Arbeitsmarkt- und Berufsforschung (IAB)
Regional Unemployment Rate 1985	Unemployment rate in the year 1985, Regional level: *Anpassungsschicht*, based on county-level data from the Bundesagentur für Arbeit

Notes: Source: Institut für Arbeitsmarkt- und Berufsforschung (IAB), Bundesagentur für Arbeit