Freie Universität Berlin

Language Proficiency and Homeownership:
Evidence from U.S. Immigrants

Marc-André Luik
Max Friedrich Steinhardt
Simon Voss

School of Business & Economics
Discussion Paper

Economics

2023/3

Language Proficiency and Homeownership: Evidence from U.S. Immigrants

MARC-ANDRÉ LUIK
Helmut Schmidt University and GLO

MAX FRIEDRICH STEINHARDT
Freie Universität Berlin, IZA and LdA

SIMON VOSS
Freie Universität Berlin

March 17, 2023

Abstract

In this paper we deliver first causal evidence on the relationship between immigrant host country language proficiency and homeownership. Using an instrumental variable strategy, we find a substantial positive impact of language skills on the propensity to own a home and the quality of housing. While this effect is mediated by education and household income, our estimates also speak in favor of a direct effect. Our results highlight the importance of host-country-specific human capital and, in particular, language proficiency for socio-economic assimilation.

JEL codes: J11, J13, J61, R21, Z13

Keywords: language, immigrants, assimilation, homeownership

Acknowledgments: We would like to thank Ingrid Gould Ellen, Emma von Essen, Tobias Heidland, Sarah Miller, Christopher Parsons, Carsten Schroeder, Juergen Schupp, and Luca Stella as well as participants and seminar audiences at the BeNa winter workshop, the DEPRISK workshop, Freie Universität Berlin, Helmut Schmidt University, Leuphana Universität Lüneburg, Luxembourg Institute for Socio-Economic Research, and University of Trier for useful comments.

1 Introduction

In recent decades, the US has experienced a large influx of immigrants, feeding into a growing proportion of foreign-born people in the population. As a consequence, it is imperative to identify drivers of social and economic integration. In this work we focus on integration into the U.S. real estate market, i.e. homeownership for which the literature documents a substantial gap between natives and immigrants. For instance, using census data Borjas (2002) reports an ownership differential of 15 and 20 percentage points in the years 1990 and 2000. For the year 2010, Camarota (2012) also finds a gap of 15 percentage points between both groups, drawing on data from the U.S. Census Bureau's American Community Survey.[1]

Homeownership is an important facet of migrant integration for many reasons. Owning a home symbolizes prosperity, stability, and economic success (Amuedo-Dorantes and Mundra, 2013). Moreover, owning a home is related to income and financial integration as it unlocks tax reliefs and allows the accumulation of equity wealth. The latter can serve as collateral for loans of educational or entrepreneurial purpose. Homeownership also creates incentives to invest in the local environment and social capital (DiPasquale and Glaeser, 1999), so that it might also serve as an accelerator of social integration. It is, hence, not surprising, that homeownership is heavily promoted by the U.S. government.[2]

One of the first studies that highlighted a positive association between homeownership and language skills of immigrants was the one by Alba and Logan (1992). As a consequence, it became common to include a language proxy in the homeownership equation in order to control for differential access to the real estate, credit and labor market (e.g. Constant et al., 2009, Painter and Yu, 2010, Amuedo-Dorantes and Mundra, 2013, Gobillon and Solignac, 2020). However, to date, the literature still lacks empirical confirmation of a causal effect. We fill this gap, by delivering first causal evidence on the impact of language skills on

[1] From 2010 onward, respondents in decennial census were no longer asked whether they had immigrated.

[2] For example, the federal government promotes homeownership through FHA loans (HUD, 2023b) and the HUD's Housing Counseling Program (HUD, 2023a). Additionally, nonprofit organizations such as UnidosUS provide support specifically targeted to immigrants (UnidosUS, 2023).

homeownership among immigrants. For this purpose, we use an instrumental variable (IV) approach which exploits a critical childhood period to acquire language skills introduced by Bleakley and Chin (2004) in the context of immigrant earnings. Comparing childhood immigrants in the U.S. from English- (control) and non-English-speaking (treatment) countries, we instrument English language proficiency with age at migration.

Using data from the American Community Survey for the years 2010-2019, we find that language skills have a substantial positive impact on immigrant homeownership. Particularly, a one-unit increase in language proficiency raises the likelihood of owning a home by 9.4 percentage points, which corresponds to 16% of the mean homeownership rate for immigrants from Non-English speaking countries. Our results are robust to different instruments and sample specifications such as focusing on refugee groups with high intentions to stay. While we cannot rule out that there are other indirect channels between language and homeownership, our results suggest that, in contrast to the language-earning association (Bleakley and Chin, 2004), the indirect mediation through factors such as education and household income do not seem to account for the entire observed relationship. Hence, direct effects of language through awareness, discriminatory practice of real estate agents and lending behavior are likely to play a role, too.

Our paper contributes to the literature on the determinants of immigrant homeownership (see Gobillon and Solignac (2020) for an overview). Previous studies have found factors such as education, income, and wealth, as well as location and ethnicity, to be most central for the likelihood of immigrant homeownership (e.g. Borjas, 2002 and Tesfai, 2016). Besides these, further determinants identified by previous research include ethnic enclave residency, time spent in the host country, immigration status, and discrimination in the labor market (e.g. Constant et al., 2009, Flippen, 2010, and Amuedo-Dorantes and Mundra, 2013).

Our work is also implicitly related to the literature on the effect of immigrant language skills on labor market outcomes. In contrast to homeownership, since the early work of Chiswick (1991), this field accumulated rich correlational and causal evidence documenting

a positive association. In order to tackle inherent endogeneity and measurement error, most works utilized an instrumental variable approach. Early arguably exogenous instruments were minority-language concentration in place of residence, whether married overseas, number of children (Chiswick and Miller, 1995), parental education (Dustmann and Van Soest, 2001), or veteran status (Chiswick and Miller, 2002). Later, the early childhood instrument, which we use in this paper, was successfully applied to identify a positive effect of language on U.S. earnings, offspring language acquisition (Bleakley and Chin, 2008) and social assimilation (Bleakley and Chin, 2010). Moreover, using the same setup, there is empirical evidence in favor of a positive effect of language on labor market and social integration in Australia (Guven and Islam, 2015) and the Netherlands (Yao and van Ours, 2015) as well as education, fertility, and health in the UK (Aoki and Santiago, 2018b).

The paper is organized as follows: Section 2 describes the data used and provides first descriptive statistics, while Section 3 presents our empirical design. Section 4 presents our empirical results and discusses potential channels. Section 5 concludes the paper.

2 Data

Our analysis uses data from the American Community Survey for the years 2010-2019, combining two IPUMS 5-year files (Ruggles et al., 2022). We limit our sample to households who have at least one immigrant spouse born outside the U.S. mainland.[3] In the case of two immigrants, we focus on the household head. Moreover, we limit our sample to childhood immigrants who arrived in the United States before adulthood (<18 years) and were between 25 and 75 years old during the observation period. This has two main reasons: First, the critical period of language skill acquisition is during early childhood. Second, children are less likely to be self-selected into the initial migration decision, although they are part of their parents' decision-making process. Following the classification of Bleakley

[3]We conduct our analysis on alternative sample definitions. Corresponding results are presented in the robustness section.

and Chin (2004), we distinguish immigrant households into being from English- or non-English-speaking countries of origin such as the United Kingdom and Mexico, respectively. As this distinction is critical for our analysis, we also exclude immigrants from countries for which it cannot be uniquely identified if English is the predominant language. Examples for the latter are countries with multiple official languages such as Hong Kong and India. Our final sample includes 533,764 immigrant households from English- (12%) and non-English (88%) speaking countries-of-origin.

Our outcome variable of interest is homeownership as being indicated by a reported dwelling ownership. We measure the effect of language through a self-reported English ability scale. Respondents can report to speak no English at all (0), to speak English, but not well (1), to speak English well (2), to speak English very well (3) or to speak only English (4). In line with Bleakley and Chin (2004), we aggregate categories (3) and (4) so that our proficiency measure increases from zero to three.

Figure 1 shows that average homeownership increases almost monotonically in language proficiency. The maximum homeownership gap spans a substantial 26 percentage points between immigrants with no and very well English proficiency. In order to learn more about unobserved heterogeneity over different levels of proficiency, we are interested in average group characteristics. For simplicity, in Table 1, we compare immigrants who speak English less than very well (column 1) with immigrants who speak English very well (column 2).[4] In line with Figure 1, homeownership is far more prevalent for the latter group. Moreover, English-proficient immigrants also tend to arrive at a younger age (7.6 vs. 13) and are less likely to originate from a Non-English-speaking country (84% vs. 99%). Moreover, they are more likely to be Black and less likely to have a Hispanic background. Finally, the majority of household heads among immigrant households with low English proficiency are male, which could also reflect differences in gender roles between the two groups. Arguably the observed heterogeneity also suggests unobserved heterogeneity with respect to factors such as ability

[4]Descriptive statistics of home ownership, language proficiency and demographic characteristics are presented in Table A.1 in the Appendix.

and intention-to-stay.

3 Empirical Design

Measuring the causal effect of language proficiency on homeownership is not straightforward as endogeneity can arise for three reasons: first, there are likely to be omitted factors that are correlated with homeownership and language skills at the same time, such as ability and family endowment. A second source of endogeneity is measurement error in the self-reported language measure. As outlined by Dustmann and Van Soest (2001), measurement error could be either unsystematic and independent over time or time-persistent. Finally, endogeneity could arise through simultaneity if foreign language proficiency is jointly determined with homeownership. In other words, homeownership and language acquisition could affect each other simultaneously.

In a perfect world, we would tackle the endogeneity issue by randomly allocating English proficiency or language training among immigrants and observing the outcome of interest. However, in the absence of this opportunity, we need to rely on natural experiments or apply quasi-experimental techniques. In this study we use an elegant instrumental-variable framework introduced by Bleakley and Chin (2004) which exploits the well-documented association between age and language acquisition from psychology and neuroscience (Lenneberg, 1967).[5] This setting has been successfully applied to study the impact of language skills on U.S. and Australian immigrants‘ social and labor market outcomes (Bleakley and Chin, 2004, Bleakley and Chin, 2010, Guven and Islam, 2015), their health insurance coverage (Dillender, 2017) as well as their participation in stock markets (Gan et al., 2022). However, to the best of our knowledge there has been no attempt to use this or any other quasi-experimental research design to identify the causal effect of language proficiency on homeownership.

[5]Alternative identification approaches in the economic literature exploit school reforms as natural experiments or use instruments such as minority-language concentrations in place of residence, veteran status, whether married overseas, number of children or parental education (Chiswick and Miller, 1995, Angrist and Lavy, 1997, Dustmann and Van Soest, 2001, Dustmann and Soest, 2002)

In particular, the instrumental variable approach exploits that there is a critical age range to develop foreign language skills. After this period it is far less likely to acquire native-level proficiency.[6] We would hence expect immigrants who arrived at a younger age to have on average better host country language skills than those who arrived later in life. In an instrumental variable framework this would be detected in a successful first stage. But what about the exclusion restriction? As Bleakley and Chin (2004) point out, age at migration alone might not fulfill this criterion due to non-language differences that also affect homeownership. For instance, younger and older migrants could differ with respect to host-country education, cultural assimilation and parental characteristics. In order to partial out non-language effects, we follow Bleakley and Chin (2004) and contrast the effect of age at arrival between immigrants from English- and non-English-speaking countries of origin. Put differently, our language effect is identified through an interaction between age at arrival and a dummy for English-speaking country of origin.

Figure 2 shows that the described relationship between English language skills and age at arrival can also be observed in our data. The red, circle marker line displays the average English skills of immigrants from non-English-speaking countries for different ages of arrival, while the blue triangle marker line illustrates the respective pattern for immigrants from English-speaking countries.[7] For immigrants originating from non-English speaking countries the picture is very clear: the earlier they immigrated to the US, the greater the language proficiency later in life. In fact, there seems to be an age threshold around the age of ten after which acquisition is impeded. In the same vein, childhood immigrants of English-speaking origin have very good English language skills irrespective of their age at arrival. The pattern delivers first graphical evidence for a successful first stage of our instrument and is in line with the empirical evidence (Bleakley and Chin, 2004).

Figure 3 displays the average rate of homeownership for our two groups of immigrants

[6]There is a large economic literature on the critical periods for human development (e.g. Cunha and Heckman, 2007, Van den Berg et al., 2014)

[7]Figure is regression-adjusted for cohort fixed effects, age, race, Hispanic background and sex.

by age at arrival.[8] The associations between age at arrival, country of origin and homeownership mirror our first-stage patterns. In particular, we find homeownership to decrease in age at migration for immigrants of non-English-speaking origin around the age of ten.[9] The similarity of Figure 2 and 3 suggests a potential positive causal effect of English language proficiency on homeownership. To investigate this further we next turn to regression analysis.[10]

4 Empirical Analysis

4.1 Main Results

As a starting point we estimate a naïve linear probability model of the following form:

$$HomeOwn_{ija} = \alpha + \beta EngSkills_{ija} + \beta X_{ija} + \mu_j + \nu_a + \xi_i + \epsilon_{ija}, \tag{1}$$

whereas $HomeOwn_{ija}$ is a dummy taking the value of one if individual i born in country j who immigrated to the United States at age a owns a house. Our explanatory variable of interest is $EngSkills_{ija}$ which measures the respondent's English language proficiency on a 3-point scale. X_{ija} is a vector of exogenous controls such as race, age and sex, while μ_j, ν_a, and ξ_i capture country-of-origin, age-at-arrival, and immigration-cohort fixed effects.

The corresponding estimates are presented in column (1) of Table 2.[11] We find the likelihood to own a home to increase by almost 8 percentage points if language proficiency increases by one unit. The estimate is significantly different from zero and economically meaningful. Our results are in line with the graphical evidence presented in section 3 and

[8] Once again, the figure is regression-adjusted for cohort fixed effects, age, race, Hispanic background and sex.

[9] We will address potential non-language effects of age at arrival later in the robustness section.

[10] Detailed information on average group characteristics of child immigrants from (non-) English-speaking origin can be found in Table A.1.

[11] As suggested by Schoellman (2016), we use person-weights. Using alternative weights delivers nearly identical estimates. See Section 4.2.

conditional on age at migration, country of origin and demographics. However, so far we have treated the endogenous regressor language proficiency as exogenous. Hence, our results might well be biased. In a next step, we therefore turn to our two-stage least squares (2SLS) estimates.

Our instrument of language ability is an interaction of two biographical migrant characteristics (Bleakley and Chin, 2004). The first component is a dummy taking the value of one if the migrant originated from a non-English-speaking country. The second element is a function of age at arrival of the following form: $max(0, a_i - 9)$, with a_i being the age at arrival in the United States of individual i. This means that our instrument switches to zero for immigrants that were younger than 10 when they entered the United States or originated from an English-speaking country. For the remaining immigrants the instrument increases monotonically in age at migration.

The estimate in column (2) of Table 2 indicates that our first stage works well. Our instrument is a significant and negative predictor of English language proficiency, which means that language skills decrease in age at arrival starting at the age ten for immigrants from non-English-speaking origin. Moreover, the F-Statistics of the first stage is over 17 – reasonably high, and clearly larger than 10, which is the commonly used threshold for a weak identification test. Our second stage estimate in column (3) indicates a causal positive and significant effect on homeownership. In particular, we find an increase in the language proficiency by one unit to raise the likelihood of owning a home by 9.4 percentage points which corresponds to approximately 16 percent of the mean homeownership rate among migrants from non-English-speaking countries.[12]

The OLS estimate in column (1) therefore does not suffer from an upward bias (through an ability bias), but instead seems to be slightly downward biased although it is statistically not different from the IV estimate. This observation has also been made by other studies dealing with the same measure of English language proficiency in the U.S. (Bleakley and

[12] A one-unit increase in the ordinal English skill measure approximately corresponds to an increase by 1.3 standard deviations among immigrants from non-English-speaking countries, see Table A.1 in the Appendix.

Chin, 2004) and Australia (Guven and Islam, 2015). An often-cited explanation for this finding is severe measurement error of the language proficiency variable leading to an attenuation bias which offsets the upward ability bias (Dustmann and Van Soest, 2001; Bleakley and Chin, 2004).

4.2 Sensitivity Checks

Before we discuss potential channels behind our result, we test its robustness in several ways. As mentioned earlier, immigrants of English-speaking origin could be seen as a poor control group for age-at-arrival effects. For instance, Bleakley and Chin (2010) argue that what we interpret as a language effect could in fact be closer cultural proximity of English-speaking countries such as Canada, UK, New Zealand and Australia. This argument is even more worrying as Canadian migrants compose the largest fraction of our control group. However, when excluding these groups from our sample the estimate remains positive and highly significant and is very similar to our benchmark estimate (column 1), 2SLS-DD, Table 3). The same holds true if we exclude Puerto Ricans or Mexicans, the latter of whom are by far the largest group of Spanish-speaking immigrants in the U.S., or if we limit our sample to immigrants from the Caribbean.[13] As our results are not driven by a particular group of immigrants, it is very unlikely that our estimates simply reflect cultural proximity or historical ties between particular sending countries and the United States.

Another way to avoid pitfalls of country of origin comparability is to follow Bleakley and Chin (2008) and exploit age-at-arrival variation within immigrants from non-English-speaking origin only. In other words, we use age-at-arrival dummies as the identifying instruments and focus only on immigrants from non-English-speaking countries. This, however, comes at the price of a strong zero non-language effects assumption. The corresponding estimates are presented in column (2) of Table 3 and are labeled 2SLS-D. Both in the full and reduced samples we find significant and positive language effects, which are very similar

[13]It is noteworthy that the effect is more pronounced when focusing on immigrants from the Caribbean Islands. Our results are robust to excluding Caribbean immigrants from the sample.

to the estimates in column (1) when using our preferred instrument (2SLS-DD).

Our 2SLS-DD estimates are also robust to modifying the age cut-off point of our instrument. As the results in columns (1) and (2) of Table 4 show, our benchmark estimate is almost unaffected whether we assume the critical childhood age to acquire language skills to be 8 or 10. As a final robustness check, we modify our preferred instrument by interacting the non-English-speaking country dummy with age-at-migration dummies. This delivers an estimate which is almost identical to the one of our benchmark specification.[14] The same holds true when using household weights or no weights (see table A.2 in the Appendix) or when using state and survey year fixed effects instead of cohort fixed effects (see table A.3) in the Appendix.

To further test the robustness of our findings, we control for spousal characteristics as migrants with no or poor English skills could rely on the help of other household members with better English skills for purchasing a house. The corresponding results are shown in Table 5 in which we divide the sample into migrants that are not married (1), married with a native (2), or married with an immigrant (3). Doing so, we find a significant effect on homeownership of similar magnitude in all subgroups.

To summarize, our results are robust to changes in the sample composition, modifications of the instrument, the use of alternative instruments, as well as the use of alternative weights, and controlling for state and survey year fixed effects or spousal background.

4.3 Potential Channels

As we know from previous studies, host country language skills have positive effects on human capital accumulation (e.g. Bleakley and Chin, 2004, 2008), labor market outcomes (e.g. Dustmann and Van Soest, 2001), and social outcomes (Bleakley and Chin, 2010) of immigrants. This is also documented in Table 6, in which we report results from 2SLS-DD regressions using the aforementioned outcomes as dependent variables. In particular, we find

[14] We also tested whether there is any impact heterogeneity with respect to years since migration or arrival cohorts. We do not find any evidence for it. The results are available on request.

that language skills positively affect earnings and the probability of having a college degree, while it has negative effects on fertility.

An obvious explanation for our findings would therefore be that the estimated effects of language on homeownership work through the aforementioned mediators. For suggestive evidence on this point, we estimate our benchmark 2SLS specification and sequentially add the potential mediators as explanatory variables. The corresponding estimates, which should be interpreted as suggestive since the potential mechanisms are endogenous, are presented in Table 7. A comparison of the estimates in column (1) and (2) suggests that fertility and marital status are not relevant factors in explaining our result as the estimate is very similar to the one from our benchmark model. Educational attainment through better language skills seems instead to be one important mechanism, as the estimate of our language coefficient falls by almost 40 percent after controlling for education. Adding household income as a further control reduces the significance of the coefficient, while its size changes relatively little. Overall, the estimates in Table 7 indicate, in line with the existing evidence on social assimilation-language association (Bleakley and Chin, 2004, 2010), that large parts of the language effects can be explained by indirect effects through mediators such as education and income. However, they seem to not account for the entire effect that we find.

To test this further we run our benchmark specification for a set of subsamples in which we distinguish between immigrants with different kind of socio-economic characteristics. In column (1) of Table A.4 we split the sample into immigrants who were never married and those who were and find positive effects of similar magnitude for both groups. The same holds true for immigrants with and without children (column 2). These results are in line with estimates in Table 7 which suggest that neither marital status nor fertility are major drivers behind our findings. Comparing immigrants with and without a college degree in column 3 shows that the effect is more pronounced for high-skilled immigrants, although the point estimate is statistically not different from the one for immigrants without a college degree. When splitting the sample by the level of household income in column 4, we do not

find any statistically significant difference in impact, with the point estimate being larger for immigrants in the upper half of the income distribution.

Another possibility could be that our language estimates reflect differences in return intentions if childhood immigrants from non-anglophone countries have systematically different return plans than those from English-speaking countries. To rule this out we follow Schoellman (2016) and construct a subsample of Indochinese immigrants who entered the US as refugees between the mid-1970s and the mid-1990s. These have in common that they generally do not have any return intentions and instead plan to stay in the United States permanently. The corresponding results, displayed in Column (1) of Table 8, are very similar to those when using the full sample (Table 2).

Moreover, our results could be driven by cultural differences with respect to the attitude towards homeownership among immigrants. To test this, we distinguish between immigrants from sending countries with low and high homeownership rates.[15] Doing so does not deliver any substantial differences in estimates suggesting that our results are not driven by institutional and historical differences in homeownership rates across countries of origin (see columns 2 and 3 in Table 8).

Finally, we test in Table A.5 whether our results could be due to the sorting of immigrants into specific regions of the United States. First, we limit our sample to migrants living in linguistic enclaves, which we define as local jurisdictions with a disproportionately high share of people speaking a specific language. This means that a Spanish-speaking immigrant is considered to live in a linguistic enclave if they live in a local jurisdiction where the proportion of Spanish-speakers is higher than the proportion of Spanish-speakers in the U.S. overall. With this definition, we follow Aoki and Santiago (2018a), while using the public-use microdata area (PUMA) as our geographical unit. While it could be assumed that English skills are less important in linguistic enclaves, we find that the results of our main analysis

[15]For origin country homeownership rates, we follow Marcén and Morales (2020), who use data for 48 countries from the IPUMS International.

hold also for this subgroup (column 1).[16]

Second, we distinguish between regions with different levels of homeownership and income. In particular, we distinguish between PUMAs with homeownership rates below (column 2 of Table A.5) and above the median rate (column 3), before looking at PUMAs with an average household income below (column 4) or above the national median household income (column 5). Once again, we find the effect of interest to be significant and of similar size in all subsamples. Therefore, we show that our results are not driven by the region in which the different groups of migrants settle.

To sum up, the evidence presented in this section suggests that human capital acquisition and household income are important channels through which language affects homeownership, while mediators such as marital status and residential sorting are of minor relevance. Return intentions also do not seem to be a driving mechanism. The fact that mediators such as education are unlikely to account for the entire observed relationship between language skills and homeownership suggests that direct effects of language skills through information or discrimination are likely to play a role, too. We will analyze this further in the following section in which we study the effect of language on the quality and financing of housing.

4.4 Financing and Quality of Housing

We start by focusing only on homeowners and distinguishing between those having a mortgage[17] and those without. The corresponding results are provided in Table 9. We find that language skills positively affect the likelihood to live in a house with a mortgage, even after controlling for endowment with human capital, household income including revenues from investments, as well as house value. The corresponding estimate in column 2 suggests that a one-standard-deviation change in language skills increases the likelihood to own a house financed with a mortgage by 10 percentage points, which corresponds to approximately 13

[16]It is noteworthy, that the difference between the OLS and IV estimate is more pronounced than in the full sample, suggesting that measurement bias among immigrants in linguistic enclaves is larger than outside of these.

[17]In the definition applied by the ACS, this also includes loans and other kinds of debt.

percent of the mean outcome among immigrants from non-English-speaking countries.

Next, we look at the effect of language skills on the likelihood of living in an overcrowded household, this time including households who live in rented apartments.[18] Previous studies have shown that US immigrants historically have often lived in particularly crowded conditions, which are, among other things, associated with high infection rates for contagious diseases (Ager et al., 2023). Overcrowding can be measured in different ways. Table 10 shows the results when two different measures for overcrowding are applied, both of which are used by the US Department of Housing and Urban Development (Blake et al., 2007).[19] In column (1), we simply look at the number of rooms and the number of persons living in the household. If there is on average less than one room for each household member, the household is considered to be overcrowded. The measure used in column (2) follows the same approach, except that now only the number of bedrooms per person is considered. In both cases, we find a significant negative effect of language skills on overcrowding, which holds even after controlling for endogenous controls such as income and education. The conditional estimate in column (2) implies that an increase in language proficiency by one unit lowers the risk to live in an overcrowded home by 10 percentage points which corresponds to 20 percent of the mean overcrowding rate among immigrants from Non-English-speaking counties.

Finally, we look at the US Census Multidimensional Deprivation Index (Glassman, 2019), according to which a housing unit is considered being of low quality if at least two of the following conditions are fulfilled: 1) Incomplete kitchen; 2) incomplete plumbing; 3) overcrowded ($>$ 2 persons per bedroom); 4) high cost burden. Column (3) of Table 10 shows the second stage results of the IV regression. Consistent with our previous results, we find that language skills reduce the risk of living in poor-quality housing.

To summarize, we find that language skills significantly affect not only the likelihood of

[18] If we focus on homeowners, results are qualitatively similar. However, we decided to report results including renter households as those are particularly affected by overcrowding (Myers et al., 1996; Blake et al., 2007) and other kinds of low-quality housing such as high cost burden (DeLuca and Rosen, 2022).

[19] Descriptive statistics of the three outcomes are provided in table A.1.

immigrants to become homeowners but also the quality of housing they find themselves in and the way they finance their housing - even after controlling for education and income. This is suggestive evidence that language also has a direct impact on homeownership by improving the access to housing and credit markets.

5 Conclusion

Immigrant homeownership is an important indicator of socioeconomic assimilation. Therefore it is important to improve our understanding of the determinants of homeownership in immigrant households. This paper contributes to the literature by analyzing the role of English language skills for homeownership in the United States. To address the endogenous nature of language skills, we instrument self-reported language proficiency by an interaction of a function of age at arrival before the age of 10 and a dummy for being born in a non-English-speaking country. Doing so, we find language proficiency to increase the chance to own a home substantially. Naïve OLS tends to underestimate the true effect, suggesting that measurement error outweighs ability bias. Moreover, we find that languages skills also positively affect the quality of housing and the likelihood of living in a home financed via a mortgage.

Our analysis further suggests that language affects homeownership both indirectly and directly. In other words, indirect effects through better education and higher household income seem not to account for the entire observed relationship. Instead, better English language skills seem also to directly increase the chances of homeownership. We provide first suggestive evidence that this might be among other factors due to better access to capital markets. Other potential explanations are an improved access to relevant information and a reduction of discriminatory behavior in the real estate market. Future research using field or lab experiments is needed to investigate this further.

Our results have important implications for integration policy and real estate markets.

First, our estimates highlight the importance of language programs for immigrant assimilation in the real estate market. As stressed by the literature, this inclusion is likely to have positive external effects on social integration. Second, immigrants in the U.S. are a key driver of current and future demand in real estate markets (Mussa et al., 2017). It should therefore be of interest to policymakers and real estate agencies to explicitly target foreign-speaking households.

References

Ager, P., J. J. Feigenbaum, C. W. Hansen, and H. R. Tan (2023). How the other half died: Immigration and mortality in US cities. *Review of Economic Studies, forthcoming*.

Alba, R. D. and J. R. Logan (1992). Assimilation and stratification in the homeownership patterns of racial and ethnic groups. *International Migration Review 26*, 1314–1341.

Amuedo-Dorantes, C. and K. Mundra (2013). Immigrant homeownership and immigration status: Evidence from Spain. *Review of International Economics 21*(2), 204–218.

Angrist, J. D. and V. Lavy (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labor Economics 15*(1, Part 2), 48–76.

Aoki, Y. and L. Santiago (2018a). Deprivation, segregation, and socioeconomic class of UK immigrants: Does English proficiency matter? *IZA Discussion Paper Series Number No. 11368*.

Aoki, Y. and L. Santiago (2018b). Speak better, do better? Education and health of migrants in the UK. *Labour Economics 52*, 1–17.

Blake, K. S., R. L. Kellerson, and A. Simic (2007). Measuring overcrowding in housing. *U.S. Department of Housing and Urban Development Office of Policy Development and Research Project No. 017-002*.

Bleakley, H. and A. Chin (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics 86*(2), 481–496.

Bleakley, H. and A. Chin (2008). What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *Journal of Human Resources 43*(2), 267–298.

Bleakley, H. and A. Chin (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal. Applied Economics 2*(1), 165–192.

Borjas, G. J. (2002). Homeownership in the immigrant population. *Journal of Urban Economics 52*(3), 448–476.

Camarota, S. (2012). Immigrants in the United States: A profile of America's foreign-born population. *Center for Immigration Studies*.

Chiswick, B. R. (1991). Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics 9*(2), 149–170.

Chiswick, B. R. and P. W. Miller (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics 13*(2), 246–288.

Chiswick, B. R. and P. W. Miller (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics 15*(1), 31–57.

Constant, A. F., R. Roberts, and K. F. Zimmermann (2009). Ethnic identity and immigrant homeownership. *Urban Studies 46*(9), 1879–1898.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *The American Economic Review 97*(2), 31–47.

DeLuca, S. and E. Rosen (2022). Housing insecurity among the poor today. *Annual Review of Sociology 48*, 343–371.

Dillender, M. (2017). English skills and the health insurance coverage of immigrants. *American Journal of Health Economics 3*(3), 312–345.

DiPasquale, D. and E. L. Glaeser (1999). Incentives and social capital: Are homeowners better citizens? *Journal of Urban Economics 45*(2), 354–384.

Dustmann, C. and A. V. Soest (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review 55*(3), 473–492.

Dustmann, C. and A. Van Soest (2001). Language fluency and earnings: Estimation with misclassified language indicators. *Review of Economics and Statistics 83*(4), 663–674.

Flippen, C. A. (2010). The spatial dynamics of stratification: Metropolitan context, population redistribution, and black and Hispanic homeownership. *Demography 47*(4), 845–868.

Gan, X., F. M. Song, and Y. Zhou (2022). Language skills and stock market participation: Evidence from immigrants. *Journal of Financial and Quantitative Analysis 57*(2), 1 – 32.

Gobillon, L. and M. Solignac (2020). Homeownership of immigrants in France: Selection effects related to international migration flows. *Journal of Economic Geography 20*(2), 355–396.

Guven, C. and A. Islam (2015). Age at migration, language proficiency, and socioeconomic outcomes: Evidence from Australia. *Demography 52*(2), 513–542.

HUD (2023a). Housing counseling program overview. https://www.hudexchange.info/programs/housing-counseling/program-description/. Accessed: 2023-03-06.

HUD (2023b). Let FHA loans help you. https://www.hud.gov/buying/loans. Accessed: 2023-03-06.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley and Sons.

Marcén, M. and M. Morales (2020). Ethnic identity and immigrant homeownership. *Journal of Regional Science 60*(1), 1–32.

Mussa, A., U. G. Nwaogu, and S. Pozo (2017). Immigration and housing: A spatial econometric analysis. *Journal of Housing Economics 35*, 13–25.

Myers, D., W. C. Baer, and S.-Y. Choi (1996). The changing problem of overcrowded housing. *Journal of the American Planning Association 62*(1), 66–84.

Painter, G. and Z. Yu (2010). Immigrants and housing markets in mid-size metropolitan areas. *International Migration Review 44*(2), 442–476.

Ruggles, S., S. Flood, R. Goeken, M. Schouweiler, and M. Sobek (2022). Integrated public use microdata series: Version 12.0 IPUMS USA. *Minneapolis: University of Minnesota*.

Schoellman, T. (2016). Early childhood human capital and development. *American Economic Journal: Macroeconomics 8*(3), 145–174.

Tesfai, R. (2016). The interaction between race and nativity on the housing market: Homeownership and house value of black immigrants in the United States. *International Migration Review 50*(4), 1005–1045.

UnidosUS (2023). Wealth and housing alliance. https://www.uushfe.org/index.cfm?fuseaction=Page.ViewPagepageId=1192. Accessed: 2023-03-06.

Van den Berg, G. J., P. Lundborg, P. Nystedt, and D.-O. Rooth (2014). Critical periods during childhood and adolescence. *Journal of the European Economic Association 12*(6), 1521–1557.

Yao, Y. and J. C. van Ours (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics 34*, 76–85.

Figure 1: Average Homeownership by English Proficiency

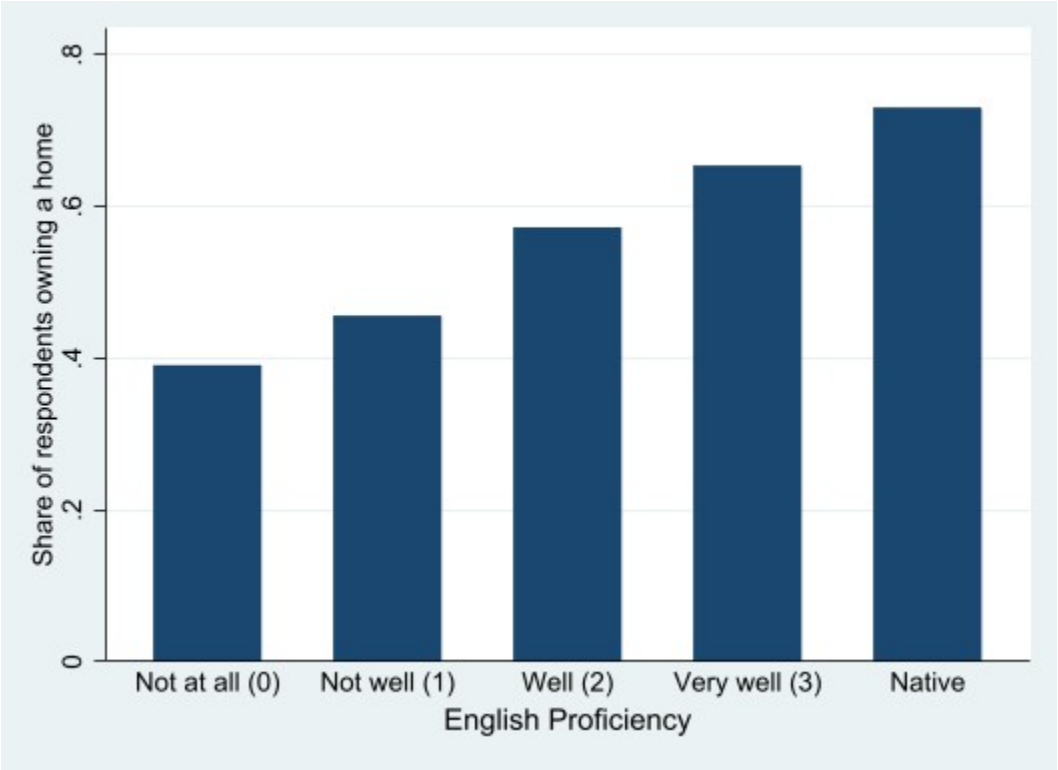
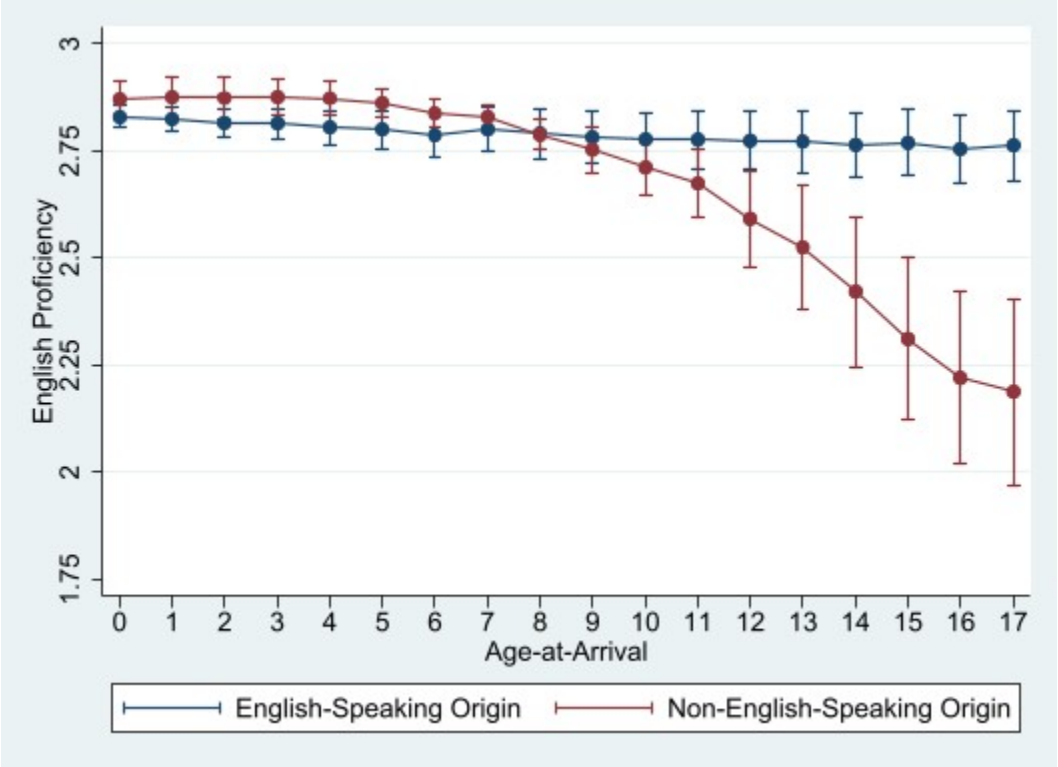
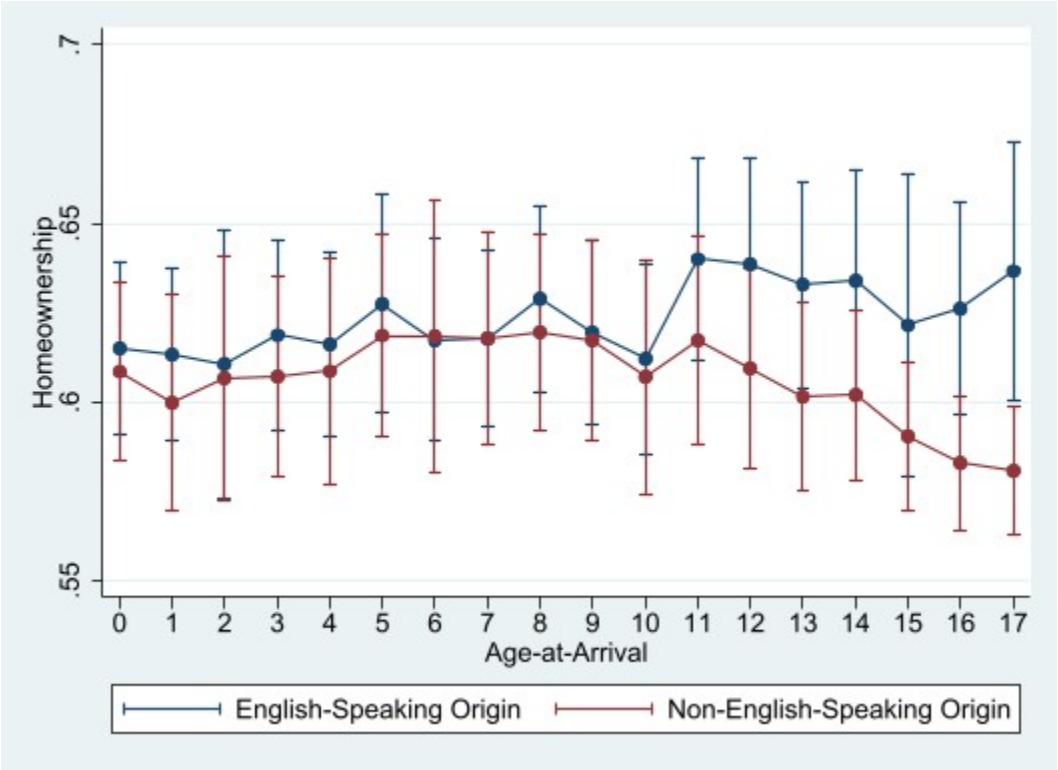


Figure 2: Regression-Adjusted English Proficiency by Age at Arrival and Origin



The figure shows the relationship between English proficiency and age at arrival, adjusted for cohort fixed effects age, sex, race, and Hispanic background.

Figure 3: Regression-Adjusted Homeownership by Age at Arrival and Origin



The figure shows the relationship between Homeownership and age at arrival, adjusted for cohort fixed effects, age, sex, race, and Hispanic background.

Table 1: Summary Statistics

	English Proficiency		
	(1)	(2)	(3)
	Less than very well	Very well	Difference
Home Ownership	0.49	0.64	0.16***
Home Ownership with Mortgage	0.68	0.76	0.08***
Rooms per person < 1	0.20	0.06	-0.15***
Bedrooms per person < 1	0.42	0.17	-0.25***
Low-Quality Housing (US Census Def., 2019)	0.03	0.01	-0.02***
Non-English Origin	0.99	0.84	-0.15**
Age At Arrival	13.00	7.64	-5.36***
Age	42.89	45.25	2.36***
White	0.55	0.64	0.08
Black	0.03	0.09	0.06***
Asian/Pacific	0.11	0.12	0.01
Other Single Race	0.29	0.12	-0.17***
Multiracial	0.02	0.03	0.02***
Hispanic	0.81	0.41	-0.40***
Female	0.44	0.50	0.06***
N	117,090	416,674	533,764

The table reports mean values. The sample consists of immigrants aged 25 to 75 during the time of observation (2010-2019), who migrated to the U.S. as children. Columns (1) and (2) report mean values. Column (3) reports the difference between column (1) and column (2). * $p < .10$, ** $p < .05$, *** $p < .01$

Table 2: OLS and 2SLS Results

	(1) OLS	(2) First Stage IV	(3) Second Stage IV
English	0.079***		0.094***
	(0.01)		(0.02)
Instrument		-0.074***	
		(0.02)	
N	533,764	533,764	533,764
adj. R^2	0.170	0.277	0.115
K-P-F statistic		17.681	

The table reports OLS and IV (2SLS-DD) estimates. The instrument used for the latter is an interaction between a function that takes the value of 0 for individuals who entered the U.S. when they were younger than 10 and increases monotonically for the remaining immigrants, and a dummy variable for being born in a non–English–speaking country. All regressions control for cohort- and age-at-arrival fixed effects, age, sex, race, and Hispanic background. Standard errors, clustered at country-of-origin level, are reported in parentheses. The K-P F statistic refers to the first-stage Kleibergen-Paap Wald rk F statistic. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3: Robustness: Sample Composition and Alternative IV

	(1) 2SLS-DD	(2) 2SLS-D
Benchmark Sample	0.0938***	0.0691***
	(0.02)	(0.01)
Control w/o Canada, U.K. Australia and New Zealand	0.0958***	0.0691***
	(0.02)	(0.01)
Treatment w/o Mexico	0.102***	0.0656***
	(0.03)	(0.02)
Treatment w/o Puerto Rico	0.0884***	0.0670***
	(0.02)	(0.01)
Carribean Sample	0.169***	0.108***
	(0.04)	(0.02)

The table reports coefficients from instrumental variable regression, with each column using a different specification of the identifying instrument. In column (1), we use the same instrument as in column (2) of Table 2 (2SLS-DD), applying the benchmark controls described in Table 2. In column (2), we use an alternative instrument and exploit the difference between younger and older child migrants from non-English-speaking country of origin, controlling for cohort fixed effects, age, sex, race, and Hispanic background (2SLS-D). Each row refers to a different sample. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 4: Robustness: Instrument Functional Form

	(1) Age Cutoff 8	(2) Age Cutoff 10	(3) Age at Migration FE
English	0.0916***	0.0911***	0.0945***
	(0.01)	(0.01)	(0.02)
N	533,764	533,764	533,764

Each column employs the benchmark set of controls, while referring to an alternative instrument specification. Columns (1) and (2) vary the cut-off age for the critical period for language acquisition. Column (3) interacts language in country of origin with age-at-migration dummies. Standard errors, clustered at country-of-origin level, are reported in parentheses.* $p < .10$, ** $p < .05$, *** $p < .01$

Table 5: Robustness: Background of Spouse

	(1) Not Married	(2) Married with Native	(3) Married with Immigrant
English	0.076***	0.080**	0.098***
	(0.02)	(0.04)	(0.02)
N	190,141	96,925	245,814
adj. R^2	0.140	0.156	0.166

The table reports OLS and IV estimates (2SLS-DD). Each column employs the benchmark set of controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 6: Potential Channels

	(1) Married	(2) Any Children	(3) Number of Children	(4) Some College Degree	(5) HH Income
English	0.00438	-0.0759***	-0.219***	0.205***	0.464***
	(0.02)	(0.02)	(0.05)	(0.04)	(0.09)
N	533,764	533,764	533,764	533,764	533,764
adj. R^2	0.037	0.135	0.153	0.039	0.033

Each column refers to a different dependent variable and reports 2SLS-DD estimates using the benchmark controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 7: Potential Channels: Sequential Inclusion

	(1) Benchmark	(2) +Married/Children	(3) +Education	(4) +HH Income
English	0.0938***	0.0943***	0.0546**	0.0420*
	(0.02)	(0.02)	(0.02)	(0.02)
N	533,764	533,764	533,764	533,764
adj. R^2	0.115	0.177	0.190	0.203

Each column refers to a different dependent variable and reports 2SLS-DD estimates using the benchmark controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 8: Indochinese Refugees and Homeownership in Country of Origin

	(1) Indochinese Refugees	(2) Above median	(3) Below/equal to median
English	0.050**	0.051***	0.075***
	(0.03)	(0.02)	(0.01)
N	15,514	285,641	178,360
adj. R^2	0.053	0.175	0.145

Column (1) reports IV estimates (2SLS-D) for the subsample of migrants that arrived in the US as refugees from Indochina between 1976 and 1996, controlling for age, sex, race, and Hispanic background. Columns (2) and (3) report IV estimates (2SLS-D) for migrants born in countries with an average homeownership rate above or below the median, respectively, controlling for immigrant cohort, age, sex, race, and Hispanic background. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 9: Effect of Language Skills on the Likelihood of Having a Mortgage

	(1) Including exogenous controls	(2) Including exogenous + endogenous controls
English	0.122***	0.102***
	(0.03)	(0.04)
N	354,928	354,928
adj. R^2	0.084	0.104

Both columns report 2SLS-DD estimates focusing on homeowners. Column (1) includes the benchmark set of controls, while Column (2) adds the marriage status, the number and age of children, the level of education, the household income, and the house value as endogenous controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table 10: Effect of Language Skills on the Quality of Housing

	(1) Rooms per person < 1	(2) Bedrooms per person < 1	(3) US Census Definition (2019)
English	-0.070***	-0.109***	-0.014***
	(0.01)	(0.01)	(0.00)
adj. R^2	0.102	0.174	0.014
Including endogenous controls	-0.034***	-0.052***	-0.005*
	(0.01)	(0.01)	(0.00)
adj. R^2	0.210	0.404	0.028
N	533,764	533,764	533,764

All columns report 2SLS-DD estimates. The table reports results for the benchmark set of controls as well as an alternative setup, where the endogenous controls marriage status, number of children, level of education, and household income are added. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table A.1: Descriptive Statistics

	(1) Full sample	(2) Nonenglish	(3) English
Home Ownership	0.61	0.60	0.68
	(0.49)	(0.49)	(0.47)
Home Ownership with Mortgage	0.75	0.74	0.77
	(0.43)	(0.44)	(0.42)
Rooms per person < 1	0.09	0.10	0.03
	(0.29)	(0.30)	(0.16)
Bedrooms per person < 1	0.23	0.25	0.10
	(0.42)	(0.43)	(0.30)
Low-Quality Housing (US Census Def., 2019)	0.01	0.01	0.01
	(0.11)	(0.12)	(0.08)
English Proficiency	2.64	2.59	2.99
	(0.72)	(0.75)	(0.14)
Speaks English not at all	0.02	0.02	0.00
	(0.14)	(0.15)	(0.01)
Speaks English not well	0.08	0.09	0.00
	(0.27)	(0.28)	(0.05)
English Proficiency well	0.14	0.16	0.01
	(0.35)	(0.36)	(0.09)
English Proficiency very well	0.76	0.73	0.99
	(0.43)	(0.44)	(0.10)
Age At Arrival	8.92	9.09	7.66
	(5.80)	(5.80)	(5.61)
Age	44.68	44.21	48.15
	(12.58)	(12.46)	(12.89)
White	0.62	0.62	0.60
	(0.49)	(0.49)	(0.49)
Black	0.07	0.04	0.30
	(0.26)	(0.20)	(0.46)
Asian/Pacific	0.12	0.13	0.04
	(0.32)	(0.33)	(0.20)
Other Single Race	0.16	0.18	0.02
	(0.37)	(0.39)	(0.14)
Multiracial	0.03	0.03	0.03
	(0.18)	(0.18)	(0.18)
Hispanic	0.51	0.57	0.02
	(0.50)	(0.49)	(0.13)
Female	0.49	0.48	0.54
	(0.50)	(0.50)	(0.50)
N	533,764	464,001	69,763

The table reports mean values. The sample consists of immigrants aged 25 to 75 during the time of observation (2010-2019), who migrated to the U.S. as children. Column (1) refers to the full sample, while Column (2) and Column (3) correspond to immigrants from Non-anglophone and anglophone countries, respectively. Standard deviations are reported in parentheses.

Table A.2: 2SLS Results - Different Weights

	(1) Second Stage IV (Person-Level Weights)	(2) Second Stage IV (Household-Level Weights)	(3) Second Stage IV (Unweighted)
English	0.094***	0.094***	0.092***
	(0.02)	(0.02)	(0.01)
N	533,764	533,764	533,764
adj. R^2	0.115	0.115	0.114

The table reports OLS and IV (2SLS-DD) estimates. Each setup employs the benchmark set of controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table A.3: OLS and 2SLS Results - Controlling for survey year and state

	(1) OLS	(2) First Stage IV	(3) Second Stage IV
English	0.076***		0.102***
	(0.01)		(0.01)
Instrument		-0.076***	
		(0.02)	
N	533,764	533,764	533,764
adj. R^2	0.189	0.279	0.135
K-P-F statistic		16.553	

The table reports OLS and IV (2SLS-DD) estimates. Each setup employs the benchmark set of controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table A.4: Potential Channels: Subsamples

	(1) Ever Married	(2) Children	(3) College Degree	(4) Upper Median HH Income
Yes	0.0856***	0.108***	0.158**	0.0740***
	(0.02)	(0.02)	(0.08)	(0.03)
adj. R^2	0.139	0.170	0.152	0.183
No	0.103***	0.0986***	0.0604***	0.0355**
	(0.02)	(0.02)	(0.02)	(0.02)
Difference (P-Value)	0.608	0.875	0.212	0.130
adj. R^2	0.116	0.202	0.163	0.161

Each column refers to a different subsample, reporting IV (2SLS-DD) results with the benchmark set of controls. Median HH Income is defined as the median household income within the same survey year, race, and age. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Table A.5: Regional Heterogeneity

	(1) Linguistic Enclave	(2) Low Homeown.	(3) High Homeown.	(4) Low Income	(5) High Income
English	0.124***	0.077***	0.101***	0.103***	0.081***
	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
N	268,718	263,930	269,834	265,266	268,498
adj. R^2	0.157	0.162	0.134	0.157	0.168

The Table reports IV estimates (2SLS-DD) for the subsamples of migrants that live in linguistic enclaves (1), in a PUMA with a homeownership rate below (2) or above (3) the median, as well as those living in a PUMA with an average household income below (3) or above (4) the median. Each setup employs the benchmark set of controls. Standard errors, clustered at country-of-origin level, are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

Diskussionsbeiträge - Fachbereich Wirtschaftswissenschaft - Freie Universität Berlin
Discussion Paper - School of Business & Economics - Freie Universität Berlin

2023 erschienen:

2023/1 RENDTEL, Ulrich; Yeonjoo LEE und Hartmut GERKS: Eine Analyse des Studienerfolgs im Masterstudium auf der Basis von Umfrage- und administrativen Prüfungsdaten: ein Vergleich von fünf Masterstudiengängen am Fachbereich Wirtschaftswissenschaft der Freien Universität Berlin
Economics

2023/2 Alho, Juha; Ulrich Rendtel und Mursala Khan: The fade away effect of initial nonresponse bias in regression analysis
Economics