The impact of host language proficiency on employability and workplace language use

Joan Comet Donoso

Advisor: Antonio Di Paolo

June 2024

A master thesis submitted in fulfilment of the requirements for the Master of Economics at University of Barcelona – School of Economics.

Abstract:

The paper investigates the impact of Spanish language proficiency among immigrants on employability and the use of Spanish in the workplace in Spain, using data from the *"Encuesta de Características Esenciales de la Población y las Viviendas"* which complements the 2021 census. Employing an Instrumental Variable (IV) strategy based on the critical period hypothesis of language acquisition, we refine the methodological approach by considering the mother tongue effects of immigrants and enhancing the exogenous assignment of Spanish proficiency through the instrument. Contrary to previous findings in Spain, the IV results suggest that Spanish proficiency does not significantly affect employment probability. However, proficient individuals show a substantial increase in the use of Spanish in the workplace, with slightly heterogeneous effects by sex. Correcting for endogeneity and classification error, Spanish-proficient immigrants present a significantly higher workplace host language use. The findings suggest a considerable downward bias in the OLS estimates, indicating that the effect of Spanish proficiency results in much greater integration benefits in the workplace, measured by host language use on the job. The study underscores the role of language in facilitating labor market integration by examining a new labor market integration outcome in a recent context. Our findings provide nuanced insights into immigrant labor market dynamics, offering valuable implications for policymakers aiming to foster inclusive workplaces.

JEL Classification: C26, J15, J24, J61, J82

Keywords: Spanish language proficiency, immigrant labor market outcomes, instrumental variable strategy, workplace integration.

Acknowledgments

I would like to thank my supervisor, Antonio Di Paolo, for his helpful comments. His guidance has been instrumental in shaping this work. Likewise, I am grateful for the dedication and professionalism of many professors in the economics master's program. Undoubtedly, their teaching has been crucial in enabling me to complete this work. Of course, all remaining errors are my own.

I would also like to thank my university colleagues, specially, Natalia, Marc, and Alejandro, with whom I have shared doubts and explanations, enduring my boring talks on the subject.

Finally, a heartfelt thank you for the support during the master's program to my mother and my sister, whom I can never love as much as they love me.

Correspondence author: Joan Comet Donoso, phone: +34 640371208, e-mail: joancometdonoso@gmail.com

1. Introduction

Migration is one of the most important demographic phenomena in today's globalized world. Each year, hundreds of millions of people move to different regions. Often, this move means that immigrants find themselves in an environment where the predominant language is different from their mother tongue. This has deep implications for their well-being and for their host societies, affecting both immigrants' labor market outcomes and their social integration (Aparicio-Fenoll & Di Paolo 2023).

A cohesive and stable society requires the economic and social assimilation of newcomers. Immigrants now constitute a significant part of the population in many countries and, consequently, an important part of the labor force. According to Eurostat, in 2023, 9.3% of active labor market participants in Europe were individuals residing in a country other than their own. In Spain, this rate is slightly higher, with the foreign population representing 14% of the labor force participation. Modern labor markets feature diverse cultural backgrounds, fostering a workplace environment where both native inhabitants and immigrants coexist in the workplace. This research aligns with the language economic literature that highlights the importance of documenting immigrant's labor market outcomes for integration and analyzing their underlying causes.

Labor market success and economic status are essential aspects of social integration (Chiswick 1991). To achieve this, individuals rely on the stock of knowledge, skills, and other attributes they have acquired over their lifetime, which are used to produce goods, services, or ideas. This collection of attributes is known as human capital (Becker 1976). Communication skills in the host country's language are undoubtedly a crucial component of human capital and have significant implications for the social and cultural integration of immigrants. Therefore, analyzing the impact of destination language proficiency on labor market outcomes is important for understanding the overall socioeconomic well-being of immigrants and their economic, social, and political involvement in the host country's society.

Returns on host language skills has been a popular topic for economic and sociological analysis over the past three decades. During this time, evidence from a variety of countries indicates that proficiency in the host country language has a significant impact on crucial aspects of integration, such as employability (e.g., Chiswick 1991, 1998; Dustmann & Fabbri 2003), earnings (e.g., Bleakley & Chin 2004; Chiswick & Miller 2007) overqualification risks and educational mismatches (Budría et al. 2021; Greene et al. 2006; Imai et al. 2019), job satisfaction and perceived professional status (Bloemen 2013), and other indicators of social integration (e.g., Bleakley & Chin 2010; Guven & Islam 2015). These results have been widely attributed to the role of language as a communication device.

An important aspect in this literature is that the study of the returns on language skills can suffer from endogeneity if language acquisition depends on unobservable features that are correlated with labor market

outcomes. The problem of endogeneity in this setting can be explained by three factors. First, unobserved heterogeneity, for instance, more able and motivated individuals may be better at language learning, and this characteristic may explain their labor market outcomes rather than their language skills. Second, reverse causality: labor market performance may contribute to language skills leading to a bias in the parameter estimate. Third, potential measurement errors in self-reported language skills could lead to bias since people tend to overestimate their proficiency levels (Bleakley & Chin 2004; Dustmann & van Soest 2002).

To address the issue of potential bias in the Ordinary Least Squares (OLS) estimates, this research adopts the methodological proposal of Bleakley & Chin (2004), which is probably the most convincing IV strategy for estimating the returns to language proficiency to date. This method is based on the "critical period hypothesis" of language acquisition (Lenneberg 1967) to construct an instrumental variable (IV) for language proficiency. The authors exploit the phenomenon that younger individuals, within the critical period of language acquisition, learn languages more quickly than older ones. Under the IV strategy, the authors prove a substantial downward bias in the OLS estimate due to measurement error and a more negligible upward bias due to endogeneity.

The use of this type of instrument is widespread in the language economics literature. For instance, studies by Ghio et al. (2023) in Italy, Yao & van Ours (2015) in the Netherlands, and Budría & Swedberg (2012) in Spain employed this approach to find causal evidence on host language acquisition and its impact on earnings and employment outcomes. Yet, the IV strategy has been extended to estimate how host language proficiency impacts immigrants' social outcomes related to integration or social assimilation, including educational attainment, health, fertility, spousal characteristics, and residential choice (see, among others; Aoki & Santiago 2018, 2024; Aparicio-Fenoll 2018; Bleakley & Chin 2010; Clarke & Isphording 2017; Guven & Islam 2015). In addition, the methodology has been used to explore the causal effect of acquiring an additional language on narrowing the native-immigrant wage gap, as evidenced by Miranda & Zhu (2013) concerning English proficiency in the UK and Di Paolo & Raymond's (2012) research on Catalan proficiency in Spain.

A crucial aspect in the IV methodology is how to determine the language skills by the immigrant upon arrival in the host country. Most of the studies use the country of origin and the predominant language as indicators of the individual's language, following the initial methodology proposed by Bleakley & Chin (2004). However, an individual's language may differ from the predominant in their country, or they may speak multiple languages. This introduces a problem in the IV identification if an immigrant is classified as a non-speaker of the host country's language by the instrument but still speaks it. This misidentification would bias the results of the analysis. To address this, the present research aims to refine this methodology by better distinguishing individuals who do or do not speak the language under analysis at their age at arrival. The analysis considers the individual's mother tongue and that of their parents to enhance the accuracy and reliability of the findings, ensuring that language proficiency is correctly identified, and the instrument

accurately assigns each individual into the control or treatment group. Additionally, the mother tongue is included as a control to capture information regarding language attribution based on the country of origin. This is in line with the evidence of Isphording & Otten (2013) that linguistic distance respect to the host country language is negatively correlated with reported language skills of immigrants.

The present research focuses on Spain, where a variety of articles highlight how knowledge of Spanish significantly contributed to the labor market outcomes of immigrants (see, among others; Budría & Swedberg 2010, 2015; Budría et al. 2017; Budría et al. 2019; Swedberg 2010). However, the evidence on these issues in Spain is largely based on the National Immigration Survey (NIS-2007), a cross-sectional survey conducted just before the outbreak of the economic crisis in 2008. The NIS-2007 data reflect a very specific situation in Spain characterized by two main events: First, a boom in the net immigration flows since the early 2000s. In 2007, Spain ranked second among OECD countries in the aggregate number of annual immigrant inflows, just behind the United States (OECD 2008). Second, a robust economic growth period and a considerable decline in the unemployment rate. Additionally, this period coincided with a massive demand in the market for activities requiring unskilled labor, such as construction, tourism, and personal services. It is crucial to recognize that this situation captured by the database changed drastically with the economic downturn that started in the third quarter of 2008. Migration inflows significantly slowed, migration outflows increased, and a large number of immigrants, most of them in the initial stages of labor market assimilation, faced substantial employment losses, particularly in low-qualified jobs (Miyar-Busto et al. 2019).

The investigation uses an updated survey, the "*Encuesta de Características Esenciales de la Población y las Viviendas*" (ECEPOV-2021), which complements the 2021 population and housing census. The data collection for this database primarily took place during the second quarter of 2021 and early 2022, capturing a period when Spain was experiencing an intense economic cycle, with significant impacts on employment outcomes due to the COVID-19 pandemic. In 2020, Spain faced important challenges, with a 7.6% decrease in employment and a 11.3% decline in GDP, the highest within the EU. An impact that was largely attributed to the heavy reliance of the Spanish economy on sectors like tourism, heavily impacted by mobility restrictions. Despite the 5.1% increase in GDP and the 6.6% rise in employment in 2021, the recovery was insufficient to offset the previous year's impacts (*Ministerio de Asuntos Económicos y Transformación Digital* 2021; *Ministerio de Trabajo y Economía Social*. 2022). These recent features make this investigation a particularly interesting setting for examining whether employment among Spanish proficiency immigrants follows the dominant patterns identified in the literature.

Overall, this research has two main objectives: first, to examine the direct effect of Spanish language skills on employment outcomes in the recent context; and second, to study the relationship between language acquisition and its use in the workplace. The primary contribution of this study lies in the analysis of Spanish

usage among immigrants in the workplace, which serves as an indicator of integration and offers new evidence on a previously unstudied aspect of labor market integration.

Inadequate access to cultural knowledge, insufficient local language skills, and lack of opportunities to interact with native speakers are significant barriers to immigrant integration. In that sense, using the host language at work not only improves job and social integration overall but also helps prevent workplace exclusion and discrimination (Bergman et al. 2008; Nelson 2014; Schaeffer & Bukenya 2010). Furthermore, from the worker's perspective it may be a crucial component for career advancement within an organization (Lønsmann 2014). Therefore, this document addresses the issue of inclusion and knowledge exchange in the workplace.

The research differentiates the hypothetical relationship between language and employment for men and women. We acknowledge that the Spanish labor market may impose additional constraints on the use of human capital based on gender. For instance, immigrant women are more likely to work for native employers, typically as housekeepers or caregivers. Consequently, this constraint could lead to different workplace behaviours and impact differently labor market and social integration outcomes (Barone & Mocetti 2011; Miyar-Busto et al. 2019).

Under the IV strategy, the results indicate that Spanish proficiency does not significantly affect the probability of employment, which contrasts with findings from previous studies. However, there is a substantial and significant effect, approximately 82,8 percentage points (pp), on the likelihood of using Spanish in the workplace for proficient immigrants, with a slightly higher impact for men. Yet, this effect is reduced by about 10pp when refining the assignment of language skills through the instrument, suggesting that considering the country of origin of the immigrant as an indicator of their language skills may not accurately reflect their host language proficiency upon arrival and introduces a considerable upward bias in the estimates. All in all, we argue that the effect on language use in the workplace is not due to selection effects into occupation related to host language proficiency but rather that Spanish proficiency facilitates labor market integration.

In short, the study aims to uncover causal evidence to inform effective policy designs and offer recommendations to policymakers and individuals alike. By identifying causal relationships between language proficiency and socioeconomic performance, we gain insight into the underlying mechanism of acquiring additional language skills beyond merely reflecting other unobserved traits at play.

The document is organized as follows: Section 2 provides the theoretical framework of the research via a literature review. Section 3 describes the database and key variables. Section 4 outlines the identification strategy and methodology employed. Section 5 presents and discusses the results. Section 6 conducts robustness checks and sensitivity analysis. Finally, Section 7 concludes the study.

2. Literature review

The present work adds to the body of language economics literature, which aims to provide a better understanding of the relationship between language skills and economic returns. A key concept in this literature is that language proficiency is a crucial component of human capital that contributes to an individual's productivity and the development of their social capital with an eventual impact on their labor market outcomes and socioeconomic characteristics. This literature provides a reasonable explanation for why cultural differences have an impact on economic outcomes by highlighting one of the most distinguishable cultural characteristics: language (Epstein & Gang 2010).

Over the last few decades, the amount of research on this general issue has substantially increased, emphasizing the importance of language skills, similar to other forms of human capital. It is worth noting the existence of main strands on the language economics literature, however, in this study, we focus on the chapter that considers the effect of host country language proficiency among immigrants. The general findings in this chapter seem to have a clear message: better proficiency in the host country language enhances migrants' socioeconomic outcomes and positively affects their integration. Moreover, this relationship appears to be causal, as the positive association between language skills and labor market and social outcomes persists even after controlling for migrants' unobservable characteristics using econometric methods that allow for estimating causal relationships (Aparicio-Fenoll & Di Paolo 2023).

2.1 Impact of host country language proficiency on immigrants

The strand of the literature on the implications of immigrants' proficiency in the language of the host country is vast. There is consensus that immigrants' host language proficiency significantly improves their labor market outcomes in the host country. However, empirically disentangling language barriers from other effects associated with migrant status and labor market abilities presents a significant challenge. This review particularly focuses on studies that address endogeneity in the returns to the destination country's language for immigrants, aligning with the focus of the current investigation.

Early works by Carliner (1981) and McManus et al. (1983) laid the foundation for this area of study and key papers employing instrumental variables have advanced the field significantly. Bleakley & Chin (2004) is probably the most influential paper from the methodological point of view. The authors use the 1990 U.S. Census to instrument host country language proficiency by the interaction of having arrived young to the host country and being born in a non-English-speaking country. Using the proposed IV strategy, the authors proof a substantial downward bias in the OLS estimate due to measurement error and a more negligible upward bias due to endogeneity. They find a significant positive effect of English-language skills on wages among individuals who immigrated to the United States as children. Dustmann & Fabbri (2003) employ an IV strategy with a matching estimator to analyze English-language fluency determinants among ethnic minority immigrants in the UK and the effect of language on labor market outcomes. Using UK survey data, they find that accounting for selection and measurement errors leads to higher employment

probabilities than those estimated by OLS. Chiswick & Miller (2010) extend previous research by examining the required English-language proficiency of occupations in the US labor market. Their study highlights the importance of occupational selection and language skills required in explaining the relationship between language proficiency and wages.

This research employs the phenomenon that language learning efficiency declines with age at the time of migration (Bleakley & Chin 2004; Budría et al. 2017; Miranda & Zhu 2013). This is related to cognitive scientists' critical period hypothesis, according to which there is an age range in which children are particularly efficient at language learning. Language proficiency increases as well with immigrants' educational attainment because of the latter's potential correlation with pre-migration exposure to the host language (Chiswick & Miller 1995; Isphording & Otten 2014). Additionally, parents' language skills have a positive and significant impact on their children's language abilities (Bleakley & Chin 2008). Moreover, the linguistic distance between the native and host languages significantly impacts the acquisition of the host language (Isphording & Otten 2011; 2013). Based on this evidence, we consider the individual's educational attainment, parents' language skills, and mother tongue into the identification strategy and sensitivity analysis.

Despite most research being conducted in English-speaking countries, there is also significant work focused on non-English-speaking countries. International evidence consistently shows that immigrants who arrive at a younger age are more fluent in the destination language. Studies in Spain (Budría et al. 2017; Budría & Swedberg 2012), the Netherlands (Zorlu & Hartog 2018), Germany (Dustmann & van Soest 2002), and Israel (Berman et al. 2003; Chiswick & Repetto 2001) similarity report positive impacts of host language proficiency on immigrant labor market outcomes, with estimates being considerably higher under IV strategies.

Nevertheless, the benefits of knowing the host country language for immigrants do not restrict themselves to the labor market. Proficient immigrants are more successful in integrating into the host society. Notably, there is substantial evidence on the effects of host language proficiency on various social outcomes for immigrants, such as health and demographic outcomes, children's education, and residential choice (Aoki & Santiago 2018, 2024; Aparicio-Fenoll 2018; Bleakley & Chin 2010; Chen 2013; Clarke & Isphording 2017; Guven & Islam 2015).

Finally, we find that the use of the local language by immigrants in the workplace significantly promotes labor integration. Insufficient host country language proficiency can act as a natural barrier to intercultural communication and information flow, negatively affecting foreign workers' work-related adjustment. A low level of proficiency in the host country's language can restrict a foreign worker's social interactions to other foreigners or to a small number of host nationals who are proficient enough in English or another shared language (Schaeffer & Bukenya 2010). Immigrants with limited host language proficiency are more likely to be categorized as out-group members (Peltokorpi 2007; Toh & Denisi 2007). Additionally, they can be excluded from communication networks due to the natural tendency of people to interact in their native

languages (Froese & Peltokorpi 2011). Therefore, a lack of host country language skills isolates immigrant workers in the workplace, leading to lower-quality relationships and perceived poorer development within their companies (Rodríguez & Yepes 2018; Lønsmann 2014; Froese 2010). However, the studies we have come across regarding host country language proficiency and workplace language use primarily originate from the field of sociology, lacking substantial empirical support from an economic standpoint. Significantly, no studies have been found that specifically examine the causal effect of language acquisition and its application in the workplace.

2.2 Language proficiency and immigrant outcomes in Spain

Research conducted in Spain consistently reveals the dominant pattern regarding the relationship between immigrants' language proficiency and their labor market outcomes. Studies indicate that immigrants with proficient language skills generally experience better job prospects (Budría et al. 2019; Miyar-Busto et al. 2019), an increased likelihood of securing permanent contracts (Budría et al. 2019), and higher income levels, particularly among non-Hispanic immigrants (Budría et al. 2017; Budría & Swedberg 2012, 2015; Swedberg 2010). In addition, research has examined the role of regional languages, such as Catalan, in Catalonia. Rendon (2007) investigated the impact of Catalan proficiency on job opportunities in Catalonia, finding a significant and positive effect. Similarly, Di Paolo & Raymond (2012) find a positive return to Catalan language proficiency, with an 18% increase in earnings for individuals fluent in Catalan.

For instance, Budría et al. (2019), using data from 2006-2007, find that proficient language skills increase the probability of employment by 15 to 22pp in Spain. However, the impact on income remains somewhat limited (Budría et al. 2017; Isphording 2015; Swedberg 2010). This limitation can be partly attributed to the inclusion of Hispanic immigrants in the analysis. Hispanic immigrants often face segmentation and occupational segregation in the Spanish labor market. Many native Spanish-speaking immigrants from Latin American countries end up in low- and middle-skill occupations, some of which require advanced Spanish proficiency. This finding is consistent with the research by Budría et al. (2019) suggesting that Spanish proficiency does not guarantee access to white-collar jobs and may even reduce the likelihood of obtaining one.

Consistent with previous findings, research by Davia et al. (2022) using the 2014 Spanish Labour Force Survey finds that the impact on occupational prestige is often less clear because of occupational segregation in the Spanish labour market amongst workers from different regions of the world but it turns positive and significant among non-Hispanic immigrants when endogeneity in language skills is considered through a bivariate analysis. The authors argue that white-collar positions demand technical expertise beyond host language skills. Interestingly, immigrants from OECD, despite not always having a strong command of Spanish, they occupy highly skilled and well-paid roles, highlighting the importance of other qualifications, languages and experiences in the labor market.

3. Data

The analysis looks at the first available series of the *"Encuesta de Características Esenciales de la Población y las Viviendas"* (ECEPOV-2021), which complements the 2021 population and housing census. The ECEPOV-2021 is an official statistical survey conducted by the National Institute of Statistics of Spain (INE, by its Spanish acronym), designed to provide detailed information about people, housing, and buildings that cannot be obtained through administrative records. This survey is planned to be conducted every five years, with the next edition scheduled for 2026.

The ECEPOV fieldwork was carried out between March 2021 and February 2022, targeting over 170,000 households across Spain, collecting data on more than 420,000 individuals. Data collection was conducted through self-completed interviews (INE 2021). The ECEPOV-21 data is divided into three files: adults, households, and housing. For this research, the household and adult files were merged to obtain all necessary variables for the analysis.

The study focuses on a subsample of immigrants, specifically individuals born outside of Spain. For model estimation, only active labor market participants are included, as the analyzed outcomes are strictly labor-related. Consequently, the sample is restricted to individuals aged 18 to 61. The variable 'age at arrival' is calculated by subtracting the year the individual started residing in Spain from their birth year. Only individuals who arrived in Spain before the age of 41 are considered. This restriction is based on the assumption that those arriving later may have non-labor-related reasons for immigrating, which could confound labor-related indicators.

The mother tongue variable is identified through the initial languages in the ECEPOV-21 survey, which include English, French, Italian, German, Romanian, Arabic, and 33 linguistic families that are identified in the "Other" mother tongue category, where respondents provided their own responses in the survey. For the classification of languages in the "other" mother tongue category into aggregated linguistic families, we based on information from ethnologue.com, a reference source providing information and statistics on the world's living languages. To see this classification, refer to Table A in the appendix. Finally, the sample excludes individuals whose mother tongue is a co-official language (Galician, Basque, Catalan, or Valencian) but not Spanish. This exclusion assumes that these individuals likely know Spanish. Since they do not declare Spanish as an additional mother tongue, we are not able to capture their Spanish proficiency. After these adjustments, the final sample comprises 25,073 individuals: 11,088 men (44.2%) and 13,985 women (55.8%).

The key variable, 'Spanish proficiency,' is a binary indicator derived from the question assessing the individual's level of Spanish. Respondents rated their Spanish skills in understanding, reading, speaking, and writing as (i) Well, (ii) With difficulty, or (iii) Not at all. An individual is considered proficient if they selected "Well" for all four skills. According to this criterion, 83,7% of the sample reports being proficient in Spanish.

The variable 'non-Spanish speaking country' is a binary indicator that identifies the country of origin of the immigrant, with 0 for Spanish-speaking countries and 1 for non-Spanish-speaking countries. This variable

is used in the construction of the IV-strategy described in the next section. For the classification of countries by official language, refer to Table B in the appendices. According to this classification, 43,3% of the sample comes from Spanish-speaking countries.

The study investigates an outcome of employability and the use of Spanish at work. Firstly, the " Working" variable measures the probability of being employed. It is a binary indicator set to 1 for individuals working (either part-time or full-time) during the week prior to the interview, excluding those engaged in household tasks (Note: In the regression tables, this variable is labeled as "probability of employment."). Secondly, the "Spanish workplace use" variable is identified from the question: "How frequently do you use Spanish at work or school?" with possible responses being: (i) Never, (ii) Sometimes, (iii) Frequently, and (iv) Always. The subsample excludes students, and the variable identifies only working individuals who select "Always."

Table 1 presents the descriptive statistics of the final sample under study. The table reports p-values from t-tests for statistical significance by gender at conventional levels. For descriptive analysis, the countries of origin are grouped into: Eastern Europe, Western Europe, Northern Africa, Sub-Saharan Africa, Latin America, Asia & Oceania, and Australia & North America. Latin America is the predominant region of origin, accounting for 45% of the sample. The proportion of working immigrants in the overall sample is 74.2%, with approximately 8pp gap between men and women. Individuals who use Spanish in the workplace constitute 77.4% of the sample, with the magnitude being slightly higher among women. Educational attainment categories include dummies for primary, secondary, post-secondary, and higher education levels. More women have higher education, but more men have completed secondary education. The 'mother tongue' category includes dummies for the primary languages identified in the questionnaire, with the "Other" category where the remaining native languages are included. It is important to note that an individual may possess multiple mother tongues. Despite 43.3% of individuals coming from Spanish-speaking countries, 55.4% declare Spanish as their mother tongue.

Tables C in the appendices show the conditional descriptive statistics with respect to Spanish proficiency. We observe that 75.7% of immigrants with Spanish proficiency are employed, compared to 65.3% of non-proficient individuals. This difference is mainly due to the gap between non-proficient men and women, with a difference of over 17pp; 73.1% of non-proficient men are employed compared to 55.7% of non-proficient women. Interestingly, the gender gap among proficient individuals is relatively smaller, around 7pp between men and women. The use of Spanish in the workplace among employed individuals is 82.7% for those with proficiency and 41.9% for those without, with slightly higher percentages for women in both cases. Proficient immigrants tend to arrive in Spain earlier, at around 23 years of age compared to 27 for non-proficient immigrants and they have higher education levels, with 31.6% compared to 9.3% of non-proficient individuals. The most common mother tongue among proficient individuals is Spanish (64.1%), and they primarily come from Latin American countries (52.9%). In contrast, non-proficient individuals mainly have Arabic as their mother tongue (39.7%) and specially come from sub-Saharan Africa (49.2%).

Table 1. Descriptive Statistics

	Overall		Man		Woman		
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Statistical sign. of difference (by gender)
Sex							
• Female	0.558	0.497					
Age	41.32	9.966	41.293	10.226	41.341	9.755	0.705
Working	0.742	0.438	0.785	0.411	0.702	0.457	0.000
Spanish workplace use	0.774	0.418	0.750	0.433	0.800	0.400	0.000
Spanish proficiency	0.837	0.369	0.83	0.376	0.843	0.364	0.005
Non-Spanish-speaking country	0.567	0.495	0.595	0.491	0.546	0.498	0.000
Country of origin							
• Eastern Europe	0.062	0.241	0.053	0.225	0.069	0.253	0.000
• Western Europe	0.255	0.436	0.266	0.442	0.247	0.431	0.001
• Northern Africa	0.007	0.086	0.01	0.097	0.006	0.075	0.000
• Sub-Saharan Africa	0.172	0.377	0.195	0.396	0.153	0.36	0.000
• Latin America	0.45	0.498	0.417	0.493	0.477	0.499	0.000
• Asia & Oceania	0.046	0.21	0.052	0.221	0.042	0.2	0.000
• Australia & North America	0.007	0.083	0.007	0.082	0.007	0.084	0.765
Age at arrival	23.331	10.111	23.131	10.328	23.49	9.934	0.005
Educational attainment							
• Higher education	0.28	0.449	0.251	0.433	0.302	0.459	0.000
• Post-secondary	0.259	0.438	0.261	0.439	0.257	0.437	0.572
• Secondary	0.325	0.469	0.351	0.477	0.305	0.461	0.000
• Primary	0.136	0.343	0.138	0.345	0.135	0.341	0.466
Mother tongue							
• Spanish	0.554	0.497	0.533	0.499	0.57	0.495	0.000
• Co-official	0.012	0.108	0.013	0.112	0.011	0.104	0.191
• English	0.042	0.201	0.044	0.204	0.041	0.199	0.336
• French	0.042	0.201	0.046	0.209	0.039	0.194	0.009
• Italian	0.019	0.137	0.025	0.157	0.014	0.119	0.000
• German	0.019	0.135	0.017	0.131	0.02	0.139	0.190
• Romanian	0.096	0.294	0.096	0.295	0.096	0.294	0.848
• Arabic	0.13	0.337	0.146	0.353	0.118	0.322	0.000
• Other	0.187	0.39	0.182	0.386	0.191	0.393	0.081

Note: The overall sample consists of 25,073 individuals, including 11,088 men (44.2%) and 13,985 women (55.8%). Excluding those engaged in household tasks, the working sample is reduced to 23,063 individuals, with 11,014 men (47.7%) and 12,049 women (52.3%). Focusing on the Spanish workplace use, the sample of working individuals includes 17,105 people, comprising 8,644 men (50.5%) and 8,461 women (49.5%). It is important to note that individuals can have more than one mother tongue. The table presents the percentage of speakers for each mother tongue. The table presents the p-values from the t-test, which assess statistical significance of differences in means between genders.

4. Empirical Strategy

The research analyses the relationship between Spanish language skills and labor market indicators. This relation can be represented through the following regression model:

$$y_{ijal} = \alpha + \beta SpanishProficiency_{ijal} + \delta_a + \gamma_j + \rho_l + \theta X_{ijal} + \varepsilon_{ijal} \quad (1)$$

Where y refers to labor market outcomes for individual i, born in country j, arriving in Spain at age a, with mother tongue l.

The variable "Spanish Proficiency" indicates the Spanish language proficiency of respondents. Dummy variables are also included for age at arrival (delta), country of birth (gamma), and mother tongue (rho). The inclusion of mother tongue fixed effects aligns with Isphording and Otten (2013), who, in their IV strategy, employ a linguistic measure in the identification. They specifically use the interaction effects between a linguistic distance measure and the years of residence in the host country. X represents the set of controls included in the model, which consist of age, and its squared, a dummy variable for gender, educational level dummies, dummies for the autonomous community where the individual resides, and the size of the municipality in four categories: (1) 50,000 inhabitants or fewer, (2) 50,001 to 100,000 inhabitants, (3) 100,001 to 500,000 inhabitants, (4) Over 500,000 inhabitants. Finally, ε is the error term.

The analysis focuses on the parameter β to identify the causal effects of Spanish language proficiency among immigrant's labor market performance. However, identifying the impact of language on socio-economic outcomes can be challenging due to the endogeneity of language skills. The endogeneity issue in the context of language skills can be attributed to three factors. Firstly, unobserved heterogeneity: for example, individuals who are more capable and motivated may be better at language learning, and this characteristic might drive their labor market outcomes rather than their language skills alone. Indeed, previous literature suggests that both language proficiency and labor market outcomes could be correlated with unobserved heterogeneity. Secondly, reverse causality: labor market performance might influence language skills, introducing bias into the parameter estimate. For instance, employed immigrants are more likely to engage with the local community, thereby reversing the causal effect of language proficiency on labor market performance. Thirdly, measures of language proficiency are self-reported and classified in a binary variable, leading to potential classification error (we use the term "classification error" over "measurement error" due to the categorical nature of the Spanish proficiency variable, which entails classification among options rather than a continuous measurement).

Considering the potential endogeneity and classification error associated with the Spanish proficiency variable in the regression, conducting OLS estimation would likely result in a biased and inconsistent estimate of the coefficient β. To address this problem, we employ a two-stage least squares estimation (2SLS) approach by using an IV that offers an exogenous assignment of the level of Spanish proficiency among individuals. The objective of the instrument is thus to predict exogenous levels in language skills and their

eventual impact. Following Bleakley & Chin (2004) we employed an IV defined as the interaction between immigrants' age at arrival and whether individuals come from a Spanish-speaking country. This choice of instrument was similarly adopted in Spain by Budría et al. (2019), Budría et al. (2017), and Budría & Swedberg (2015).

The IV strategy builds upon the strong association between age of arrival in Spain and Spanish proficiency in adulthood. Age at arrival is negatively correlated with language proficiency, as younger individuals tend to acquire languages easier than adults. Cognitive scientists refer to this phenomenon as the critical period hypothesis according to which there exists an age range during which individuals learn languages more efficiently (Chiswick & Miller 2008). Figure 1 captures much of the correlation between age at arrival and Spanish-language proficiency and illustrates the relationship between age at arrival and Spanish-language skills among immigrants from non-Spanish-speaking and Spanish-speaking countries.

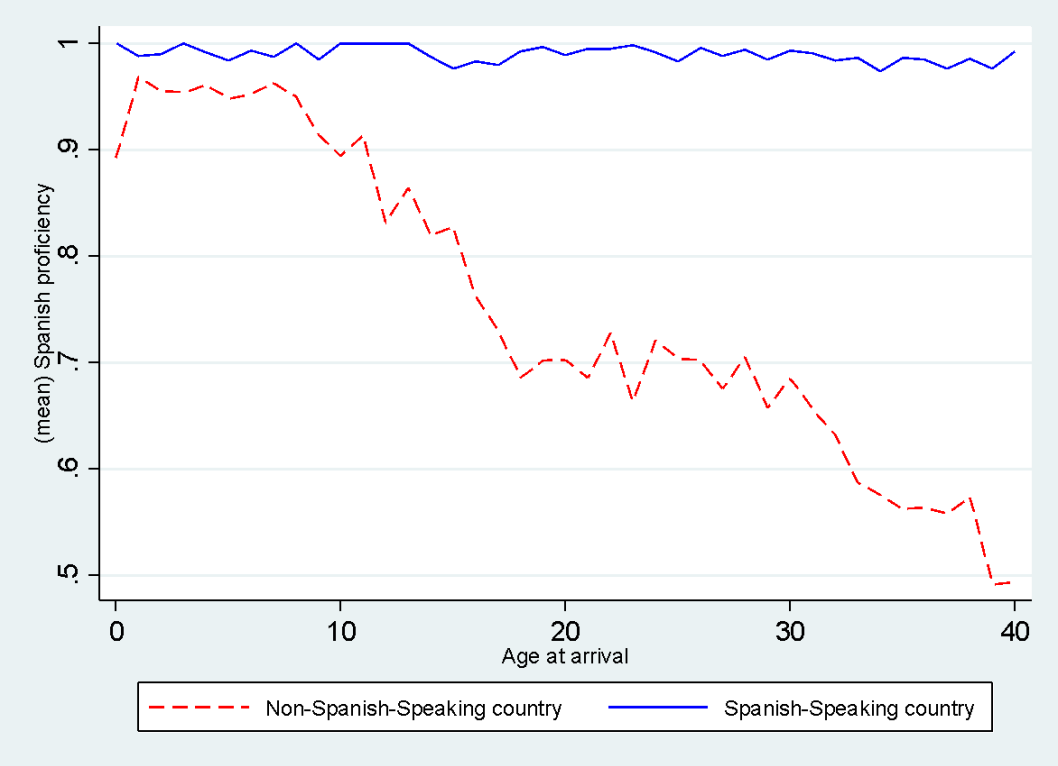

Figure 1. Spanish proficiency among immigrants from non-Spanish-speaking countries and immigrants from Spanish-speaking countries, by age at arrival.

Figure 1 shows that immigrants from Spanish-speaking countries clearly exhibit fluency in Spanish, regardless of their age at arrival. In line with the critical period hypothesis, immigrants from non-Spanish-speaking countries who were first exposed to Spanish at a young age achieve language skills comparable to those of immigrants from Spanish-speaking countries. However, immigrants whose initial exposure to Spanish occurred later in life present lower proficiency levels. In the study, we assume that the disparity between the two groups increases linearly with age at arrival. We note that up to a certain early age at arrival there is no discrepancy in language skills between immigrants from non-Spanish-speaking and Spanish-

speaking countries. Our selection of this age at arrival threshold differs from Bleakley & Chin (2004), who used 9 years old as the cutoff. They focused on individuals who migrated below the age of 18 to the USA, which is not directly applicable in our case, where the majority of migrants arrived after this age to Spain (75% of the sample). After testing different age at arrival thresholds in a first-stage model, we determined that the optimal predictive results for Spanish proficiency were obtained at 15 years old at arrival. For further reference on the first-stage results using different age at arrival cutoffs in the instrument, please see Table D in the appendices. Nevertheless, as we shall see, the instrument is highly significant in all the first-stage equation models. Consequently, we use the following parameterization for the instrument:

$$Z_{ija} = max(0, ageatarrival - 15) \times I(j\ is\ a\ non - Spanish - speaking\ country) \quad (2)$$

where I () is the indicator function, j is the country of birth and a the age at arrival. The indicator function is equal to 1 if the country of birth is a non-Spanish-speaking country and 0 otherwise.

It is crucial to note the inclusion of immigrants from Spanish-speaking countries in the equation to isolate the effects of age at arrival that affect the immigrant social outcomes through channels different from language acquisition. Age at arrival alone is not a valid instrument due to potential direct effects on the immigrant social outcomes through channels different from language acquisition (immigrants who arrive early may benefit from factors such as cultural assimilation or a better understanding of institutions and social services than those who arrive later), we separately control for having arrived young and being born in a non-Spanish-speaking country. Thus, any difference in outcomes between early and late arrivers from non-Spanish-speaking countries beyond the equivalent difference observed for those immigrants from Spanish-speaking countries could reasonably be attributed to language effects.

To refine our identification strategy, we implement sample restrictions using differentiated estimations. Firstly, we exclude individuals with a non-Spanish-speaking country of origin but with Spanish as their mother tongue. Secondly, individuals with a non-Spanish-speaking country of origin and a non-Spanish mother tongue but with Hispanic-origin parents are also excluded from the analysis. Thirdly, we exclude the bilingual autonomous communities. On the other hand, we conducted an analysis examining the effects of linguistic skills by measuring both speaking proficiency and writing proficiency. We consider that for certain jobs, especially low-skill positions, strong writing skills are not necessary and there could be differences between the causal effects of communication skills and formal skills. All of these findings are detailed in Section 6, "Robustness Checks."

Finally, we present the first-stage regression for Spanish proficiency, where the identifying instrument targets the compliers. These are the only individuals for whom the assignment of language skills is assumed exogenous in the model. This emphasis is crucial because, by design and under standard assumptions, 2SLS produces the Local Average Treatment Effect (LATE). Our LATE focuses on individuals whose Spanish proficiency was influenced by the instrument. Specifically, we focus on individuals meeting the general

specification criteria: those with an age at arrival equal to or before 15, originating from a non-Spanish speaking country. The formal equation for the first-stage model is as follows:

$$SpanishProficiency_{ijal} = \alpha + \pi Z_{ija} + \delta_a + \gamma_j + \rho_l + \theta X_{ijal} + \varepsilon_{ijal} \quad (3)$$

We expect the parameter of the instrument, π, to be statistically significant and negative, assuming a negative linear trend between age at arrival and language skills. Additionally, since residuals of labor market outcomes are likely to be correlated within regions of origin, we clustered the standard errors by country of birth.

The main underlying assumption behind the validity of this identification strategy is that non-linguistic cohort effects are common for both linguistic immigrant communities. Specifically, non-language age-at-arrival effects are assumed to be the same for immigrants from Spanish-speaking and non-Spanish-speaking countries. However, we consider the potential for differences in age-at-arrival effects on labor market outcomes between Spanish-speaking immigrants and, specifically, English-speaking immigrants. English plays a significant role in global business, particularly for high-ranking positions in countries like Spain. Following the findings of Budría et al. (2019) and Davia et al. (2022), white-collar positions often demand technical expertise beyond host language skills. Despite not always having a strong command of Spanish, immigrants from developed countries occupy highly skilled and well-paid roles, highlighting the importance of other qualifications, languages and experiences in the labor market. To address this issue, we restrict the sample by excluding individuals who speak only English at work. The results and discussion of this analysis is detailed in Section 6, "Robustness Checks."

5. Results

In this section, we present the results of our analysis. All tables follow the same structure: estimates are provided for the entire sample and separately for men and women. Additionally, the models are presented with both basic and extended controls. The basic controls include age and its quadratic term and dummy variables for gender, age at arrival, and country of birth. The extended controls add dummy variables for educational level, mother tongue, autonomous community of residence, and municipality size.

Tables 2 and Table 3 below present the OLS regressions for the labor market indicators described in Section 4, "Data.". Table 1 shows a significant effect on the probability of working, with a 7.9pp increase for immigrants with Spanish proficiency. The coefficient is slightly higher for women at 9.8pp, compared to 6pp for men. When extended controls are included, this effect reduces to 5.5pp, with 4.2pp for men and 6.6pp for women. The coefficient for age—a proxy for professional experience—is associated with a higher probability of employment, though at a decreasing rate as indicated by its square. As expected, higher education level has a positive effect on the probability of employment, increasing it by 10pp respect to primary education, similarly for both sexes. Additionally, post-secondary education presents a positive effect, about 6.8pp for men and 3.6pp for women compared to primary education level. Results on secondary education are not statistically significant.

Table 2 presents the effect on the probability of using Spanish at work, revealing a significant positive result of 27.4pp increase for those proficiency in Spanish in the basic model. This effect is similar for both men and women and remains analogous in magnitude when extended controls are included. Interestingly, we find a negative association of highly educated men, with a 4pp decrease in the probability of using Spanish at work compared to those with low education levels.

Table 1. OLS regression

	Dependent variable: Probability of employment					
	Basic controls			Extended controls		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All sample	Men	Women	All sample	Men	Women
Spanish Proficiency	0.079***	0.060***	0.098***	0.055***	0.042**	0.066***
	(0.008)	(0.014)	(0.018)	(0.009)	(0.018)	(0.018)
Sex	-0.100***			-0.105***		
	(0.019)			(0.019)		
Age	0.029***	0.033***	0.024***	0.028***	0.033***	0.023***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Age2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education				0.010	0.014	0.005
				(0.012)	(0.020)	(0.020)
Post-secondary				0.050***	0.068***	0.036*
				(0.018)	(0.026)	(0.021)
Higher education				0.105***	0.106***	0.104***
				(0.021)	(0.027)	(0.024)
Constant	0.091	0.020	0.083	0.028	-0.039	0.020
	(0.061)	(0.076)	(0.070)	(0.067)	(0.080)	(0.076)
N	23,046	10,998	12,027	23,046	10,996	12,027
R-squared	0.059	0.055	0.069	0.078	0.078	0.091
adj. R^2	0.0522	0.0413	0.0569	0.0687	0.0596	0.0741
F	45.08	52.18	28.84	107.1	89.83	55.31

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. Columns (4), (5), and (6) also control for mother tongue, autonomous community of residence, and municipal size (based on population). The table presents the F-test for the joint significance of the independent variables. *** p<0.01, ** p<0.05, * p<0.1

Table 2. OLS Regression

	Dependent variable: Spanish use in the workplace					
	Basic controls			Extended controls		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All sample	Men	Women	All sample	Men	Women
Spanish Proficiency	0.274***	0.275***	0.272***	0.272***	0.272***	0.270***
	(0.014)	(0.014)	(0.025)	(0.014)	(0.015)	(0.026)
Sex	0.013			0.016**		
	(0.008)			(0.007)		
Age	0.006	0.003	0.009	0.006*	0.002	0.008
	(0.004)	(0.004)	(0.006)	(0.003)	(0.004)	(0.005)
Age2	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education				0.004	-0.006	0.018*
				(0.008)	(0.011)	(0.010)
Post-secondary				0.008	0.003	0.019
				(0.010)	(0.013)	(0.015)
Higher education				-0.031**	-0.040**	-0.018
				(0.015)	(0.017)	(0.020)
Constant	0.351***	0.391***	0.329***	0.346***	0.383***	0.344***
	(0.090)	(0.091)	(0.121)	(0.083)	(0.107)	(0.112)
N	17,087	8,628	8,437	17,087	8,626	8,436
R-squared	0.191	0.209	0.179	0.227	0.244	0.222
adj. R^2	0.183	0.195	0.165	0.216	0.225	0.203
F	133.8	178.0	44.74	93.06	96.18	24.31

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. Columns (4), (5), and (6) also control for mother tongue, autonomous community of residence, and municipal size (based on population). The model's sample includes only individuals who were employed at the time of the interview. The table presents the F-test for the joint significance of the independent variables. *** p<0.01, ** p<0.05, * p<0.1

As described in the previous section, we further applied 2SLS estimation to address endogeneity and classification error problems that could bias the OLS estimates. Table 3 shows the results of the first stage regression. The instrument is defined as the interaction between age at arrival and coming from a Spanish-speaking country. We observe that all estimates are highly significant, and the negative sign aligns with theoretical intuition: immigrating to Spain one year later and coming from a non-Spanish-speaking country decreases the probability of acquiring Spanish proficiency by 1.3pp for the overall sample in the basic model. This magnitude is slightly higher for men (1.4pp) compared to women (1.3pp). When all controls are included, the magnitudes are slightly reduced in all cases. As expected, the impact is smaller as the education level increases. Mother tongues also play a significant role in the model. English as mother tongue does not show significant results, but French and Italian are associated with a better acquisition in the probability of acquiring Spanish proficiency, with a more pronounced effect for French (7.4pp) compared to Italian (3.4pp). Interestingly, the effect for Italian appears to be mediated primarily by men. In contrast, having German, Romanian, or Arabic as a mother tongue is linked to a larger decrease in the probability of acquiring Spanish proficiency, particularly for immigrants with Arabic and Romanian as their mother tongues. For the German language, the negative effect on language acquisition is only significant among men.

Table 3. First stage regression

	Dependent variable: Spanish Proficiency					
	Basic controls			Extended controls		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All sample	Men	Women	All sample	Men	Women
Instrument	-0.013***	-0.014***	-0.013***	-0.012***	-0.013***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Sex	-0.016			-0.017		
	(0.016)			(0.014)		
Age	-0.000	-0.003*	0.001	-0.000	-0.002	0.001
	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
Age2	0.000	0.000**	0.000	0.000	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education				0.192***	0.175***	0.202***
				(0.049)	(0.039)	(0.057)
Post-secondary				0.239***	0.228***	0.245***
				(0.060)	(0.049)	(0.068)
Higher education				0.269***	0.256***	0.274***
				(0.066)	(0.057)	(0.072)
Mother tongue (main)						
• English				0.021	0.008	0.040
				(0.030)	(0.028)	(0.036)
• French				0.074***	0.065***	0.084***
				(0.025)	(0.020)	(0.031)
• Italian				0.034**	0.032**	0.040
				(0.015)	(0.015)	(0.037)
• German				-0.036**	-0.047***	-0.034
				(0.016)	(0.014)	(0.024)
• Romanian				-0.133***	-0.147***	-0.106***
				(0.017)	(0.038)	(0.025)
• Arabic				-0.159***	-0.137***	-0.183***
				(0.013)	(0.014)	(0.022)
Constant	0.874***	0.917***	0.851***	0.677***	0.736***	0.646***
	(0.046)	(0.034)	(0.081)	(0.099)	(0.064)	(0.143)
N	25,059	11,072	13,967	25,059	11,070	13,967
R-squared	0.335	0.316	0.366	0.391	0.372	0.423
adj. R^2	0.330	0.306	0.359	0.385	0.360	0.413
F	26.22	25.52	19.44	69.24	103.2	67.58

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. Instrument defined as the interaction of the age at arrival and coming from a Spanish-speaking country. Columns (4), (5), and (6) also control for 33 "other" mother tongue language categories, autonomous community of residence, and municipal size (based on population). The table presents the F-test for the joint significance of the independent variables. *** p<0.01, ** p<0.05, * p<0.1.

The IV results of the second-stage regressions for the labor market outcomes are displayed in the tables 4 and 5. The tables report the F-statistic to test the relevance of the IV. For all the models and tables, weak identification tests by critical values from Stock-Yogo and the rule of thumb (F-stat > 10) reject the null hypothesis, suggesting that the instrument does not appear to be weak. Moreover, we observe that the models yield significantly higher F-test results when all controls are included.

In contrast to the OLS results, when considering endogeneity under 2SLS, the results presented in Table 4 indicate that the estimates of the impact of Spanish proficiency on the probability of employment are not statistically significant in any of the models. Interestingly, the probability of employment shows negative coefficients for Spanish proficiency women. However, these values are not statistically significant and therefore not different from zero. The controls for age work similarly to the OLS models; the probability increases with age but eventually decreases. Additionally, education levels play an important role, indicating a higher probability of overall employment as education increases. The findings suggest that there are no labor selection effects caused by proficiency in the Spanish language, at least for the individuals affected by the instrument. It is important to recall that 2SLS estimates represent local effects on the compliers This implies that given the relevance of the instrument, if there is self-selection into employment, it is not due to the assignment of the instrument, in other words, the instrument itself does not explain the selection into occupation. The findings contrast with previous evidence on the Spanish labor market, suggesting that the linguistic effect of Spanish proficiency may have changed in importance in the productive process and as a component of an individual's human capital for labor market employability.

Table 4 IV regressions

	Dependent variable: Probability of employment					
	Basic controls			Extended controls		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All sample	Men	Women	All sample	Men	Women
Spanish Proficiency	-0.022	0.108	-0.134	-0.044	0.104	-0.192
	(0.087)	(0.093)	(0.127)	(0.090)	(0.092)	(0.141)
Sex	-0.099***			-0.105***		
	(0.019)			(0.019)		
Age	0.029***	0.033***	0.024***	0.028***	0.033***	0.023***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Age2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education				0.028	0.003	0.053
				(0.024)	(0.020)	(0.039)
Post-secondary				0.073**	0.054*	0.095*
				(0.033)	(0.027)	(0.051)
Higher education				0.130***	0.090***	0.170***
				(0.034)	(0.027)	(0.054)
N	23,046	10,998	12,027	23,046	10,996	12,027
R-squared	0.021	0.016	-0.010	0.040	0.037	0.009
adj. R^2	0.0191	0.0125	-0.0133	0.0357	0.0281	-0.000106
F	57.46	67.76	42.68	120.0	135.2	72.63
(p-value)	0.000	0.000	0.000	0.000	0.000	0.000

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. Columns (4), (5), and (6) also control for mother tongue, autonomous community of residence, and municipal size (based on population). The F-test results shown in the table correspond to the F-test for excluded instruments. The p-value associated to the F-test is presented below. *** p<0.01, ** p<0.05, * p<0.1.

On the other hand, Table 5 presents a significant and positive effect for the use of Spanish at work for Spanish proficiency immigrants, with an increase of 87.1pp for the basic model in the overall sample. This effect decreases to 82.8pp in the extended model, once all controls are included, with 83.2pp for men and 77.5pp for women. Under the IV strategy, the magnitudes are significantly higher than in the OLS estimates. In addition, age controls exhibit similar behavior to what was previously described in the OLS estimates. Regarding educational level, we find that the probability of using Spanish at work decreases as education levels increase across all levels, not just in higher education as observed in the OLS results. Additionally, under the IV approach, the magnitudes of the educational effect are higher than those in the OLS results, and it turns significant for women. The estimates show that men have an 18.2pp lower probability of speaking Spanish at work if they have higher education, while women have a 12.2pp lower probability. Based on the findings, we could suggest that among the immigrant population in Spain, high-skilled positions, typically occupied by highly educated individuals, may be more closely associated with the use of languages other than Spanish within the workplace.

Table 5. IV Regressions

	Dependent variable: Spanish use in the workplace					
	Basic controls			Extended controls		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All sample	Men	Women	All sample	Men	Women
Spanish Proficiency	0.871***	0.865***	0.822***	0.828***	0.832***	0.775***
	(0.150)	(0.163)	(0.158)	(0.108)	(0.136)	(0.120)
Sex	0.004			0.011		
	(0.009)			(0.008)		
Age	0.007*	0.004	0.009*	0.006**	0.003	0.009*
	(0.004)	(0.004)	(0.006)	(0.003)	(0.003)	(0.005)
Age2	-0.000	-0.000	-0.000	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Secondary education				-0.086***	-0.106***	-0.052**
				(0.030)	(0.035)	(0.026)
Post-secondary				-0.107***	-0.127***	-0.072***
				(0.032)	(0.043)	(0.024)
Higher education				-0.160***	-0.182***	-0.122***
				(0.035)	(0.046)	(0.029)
N	17,087	8,628	8,437	17,087	8,626	8,436
R-squared	-0.145	-0.158	-0.101	-0.071	-0.086	-0.029
adj. R^2	-0.148	-0.163	-0.106	-0.0775	-0.0992	-0.0420
F	41.20	50.45	26.46	71.23	93.70	35.85
(p-value)	0.000	0.000	0.000	0.000	0.000	0.000

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. Columns (4), (5), and (6) also control for mother tongue, autonomous community of residence, and municipal size (based on population). The model's sample includes only individuals who were employed at the time of the interview. The F-test results shown in the table correspond to the F-test for excluded instruments. The p-value associated to this F-test is presented below. Additionally, the F-test for the difference in the Spanish proficiency coefficients between the extended models by gender is 41.83 with p-value < 1%. *** p<0.01, ** p<0.05, * p<0.1.

Correcting for endogeneity and classification errors, we find that immigrants from non-Spanish-speaking countries who arrive in Spain earlier benefit more from Spanish proficiency in their workplace application, facilitating their integration, compared to the effect indicated by the OLS results. The IV results also reveal heterogeneous effects between genders, indicating a form of job segregation influencing linguistic attitudes in the workplace (as evidenced by the F-test of statistical difference of coefficients between the models by gender). One possible explanation is that male immigrants use Spanish more frequently at work because they have better access to local labor market positions where natives work, compared to female immigrants. This would allow men to integrate more readily into the workplace environment alongside natives and, as a result, use the host language more frequently. However, this is merely a hypothesis that requires further investigation, and we take it as a potential research based on the findings.

6. Robustness checks

To ensure the reliability and validity of our findings, we conducted a series of robustness checks. These checks aim to confirm that our results are not sensitive to various assumptions or potential biases. Specifically, we examined different model specifications, included additional specifications for the Spanish proficiency variable, and tested the stability of our estimates across various subsamples.

Firstly, we modify the identification of the Spanish proficiency variable, which identifies proficiency only in individuals with good skills in all considered language skills (speaking, reading, understanding, and writing). However, in order to measure the effect of language skills on employability, we can argue that for certain jobs, individuals may not require proficiency to write or read, especially low-skilled jobs. For example, positions primarily involving customer service tasks may demand good linguistic skills in the host language but not necessarily written skills. On the other hand, written skills may be especially necessary for higher-skilled jobs. In such cases, how we measure host language proficiency variable could yield different impacts on labor market indicators. Table E in the appendix presents the results of the 2SLS estimations for the employability indicator in which the instrument is used to identify individuals' proficiency in only the "speaking" skills for Panel 1 or "writing" skills for Panel 2. The results of the analysis show no significant difference regarding the measurement of Spanish proficiency when considering all linguistic aspects. Consequently, there are no statistically significant impacts associated with either speaking or writing on the employment probability.

Secondly, Table F in the appendix measures the stability of the results by conducting estimations across different subsamples. Panel 1 and Panel 2 employ subsamples to refine the instrument identification to better diagnose the Spanish language skills of immigrants when they arrive in Spain. Panel 1 shows the results when excluding individuals from non-Spanish-speaking countries but Spanish as their mother tongue. Panel 2 excludes individuals from non-Spanish-speaking countries with a non-Spanish mother tongue but with Hispanic-origin parents. Finally, Panel 3 excludes the bilingual autonomous communities

from the sample, namely: Catalonia, Valencia, Basque Country, and Galicia to avoid the effect of regional languages on the Spanish proficiency levels and their potential impact on the labor market.

In short, Table F shows that the effect on the probability of employment remains non-significant for Spanish proficiency individuals across all subsamples. Regarding the probability of using Spanish at work, both Panel 1 and Panel 2 show a decrease of about 17pp compared to the estimates using the general sample. In Panel 1, Spanish proficiency's effect on workplace Spanish usage decreases from 87.1pp in the overall sample of the basic model to 70.7pp when excluding individuals from non-Spanish-speaking countries but with a Spanish mother tongue. This decline is more pronounced for women (from 82.2pp to 63.5pp) than for men (from 86.5pp to 72.8pp) compared to the baseline estimates. Panel 2 also shows a decrease in the Spanish proficiency's effect on workplace language use from 87.1pp to 68.6pp when excluding individuals from non-Spanish-speaking countries with a non-Spanish mother tongue but with Spain-origin parents, with similar magnitude between sexes. The findings in Panel 1 and Panel 2 indicate that refining the instrument to accurately measure the Spanish proficiency of individuals upon their arrival in Spain leads to a reduction in the estimated impact of Spanish proficiency. Lastly, Panel 3 indicates that Spanish-proficient women have a higher probability of using Spanish at work (85.3pp) compared to when bilingual communities are included (82.2pp). In contrast, for men, the probability decreases from 86.5pp to 80.6pp when bilingual autonomous communities are excluded from the sample. Therefore, we could argue that regional languages do not negatively affect the use of Spanish at work for proficient male immigrants.

Finally, the analysis questions the main assumption of the identification instrument, which establishes that age at arrival effects are the same for both study cohorts: immigrants coming from Spanish-speaking countries and those coming from non-Spanish-speaking countries. We suggest that individuals with English proficiency may have an advantage in finding employment due to the importance of English in the business world and as a lingua franca. This is particularly relevant in countries like Spain, where English proficiency levels among natives are relatively low (according to the latest English Proficiency Index (EF 2023) report, Spain ranks 25th out of 34 European countries in English proficiency). To address this issue, immigrants who only speak English at work are identified and excluded from the estimation. Table G in the appendix shows the results of this analysis. As observed, there are no significant changes in the employment probability when excluding these individuals from the sample. This estimation is also conducted by excluding those individuals as well who frequently (and always) speak English at work, and there is no significant variation.

7. Conclusions

Facing a new environment presents a challenge that is further exacerbated by cultural differences. Immigrants can learn host country languages to develop and use language skills to do many things, including work. These skills can be rewarded in the labor market, and they may invest time and resources in acquiring them. We study the effects of such investment complementing the existing literature in Spain and focusing on an indicator of labor market integration.

Recent data from the ECEPOV-2021 provide an updated and relevant perspective on a changing labor market. Initially, we applied an OLS estimation, assuming that language skills are exogenous to labor market outcomes and classification errors are not significant. The results indicate that Spanish proficiency increases the probability of employment by 5.5pp, with a greater effect as individuals attain higher levels of education. Additionally, the research considers gender differences in labor market outcomes. OLS estimates reveal that the effect on employment probability for women with Spanish proficiency is slightly higher (6.6pp) compared to men with Spanish proficiency (4.2pp). We also observe a significant association between proficiency in Spanish and its use at work among immigrants, with a 27.2pp higher probability compared to non-proficiency individuals, which remains similar in magnitude across genders.

Nonetheless, to address the issues related with classification error and endogeneity, we further developed the analysis using an 2SLS estimation. The research adopts an IV strategy based on the critical period hypothesis of language acquisition, which posits that there is an age range in which younger individuals are particularly efficient at language learning. Additionally, we incorporated the effects of linguistic distance on language acquisition by including fixed effects of mother tongues. Despite the IV methodology, challenges persist in precisely identifying immigrants' language proficiency upon arrival in the host country, potentially introducing bias into the results. Therefore, we conducted sensitivity analyses through estimations with more specific subsamples to better determine Spanish proficiency levels at the age at arrival.

Under the IV strategy, we do not find significant results in the probability of employment for immigrants acquiring Spanish proficiency. These findings suggest that the OLS estimation may be affected by endogeneity issues, meaning that proficiency in Spanish could be correlated with unobserved factors that also influence the probability of employment, or there may be a substantial effect from a classification error, both highlighted in previous literature. The finding contrasts with previous studies that found a significant causal effect of Spanish proficiency on employability but in different contexts and periods. Our findings suggest that there are no labor selection effects in the labor market caused by proficiency in the Spanish language, at least for the individuals affected by the instrument, meaning the instrument itself does not explain selection into occupations. In light of this, the results may suggest that the competitive advantage of language proficiency in terms of employability has weakened in recent years. One possible explanation could be a changing labor market, evidenced by globalisation and digitalization, where languages other than Spanish are not only present but also play significant roles in the Spanish workforce.

Our study significantly contributes by examining Spanish language usage in the workplace as an indicator of integration. Proficiency in Spanish notably increases the likelihood of using the host language at work, with a slightly higher effect for men (83.2pp), but also notable for women (77.5pp). We suggest gender differences may stem from men having better access to local labor markets where natives work, facilitating easier integration alongside locals and more frequent use of the host language. However, this hypothesis requires further investigation. Additionally, our findings indicate that Spanish usage at work reflects genuine integration rather than selective employability criteria. It underscores the adoption of the host language as a key indicator of labor market integration and social assimilation. Moreover, refining language skill assignment shows a 10pp reduction in this effect, highlighting that country of birth alone may not accurately determine host language proficiency upon arrival, potentially biasing estimates upward.

Considering the study's limitations, it is important to note that issues like racial discrimination and segregation into low-skilled sectors, which may disproportionately affect immigrants based on their country of origin, are not fully addressed in the analysis. Including these dimensions would provide a more comprehensive understanding of the factors influencing the employability and integration of immigrants in the workforce. In addition, examining how Spanish proficiency impacts different economic sectors could help identify areas where language training is particularly crucial, providing data for more targeted training programs. Lastly, assessing the sensitivity of the results to other language competencies, such as English, raises the importance of considering other linguistic skills in the individual, which can offer valuable insights for designing more effective integration policies.

Moreover, the data from the ECEPOV-2021 presents some limitations. Firstly, it is situated within a complicated economic context, which may be particularly sensitive for a population at risk of exclusion, such as immigrants, and the findings may not be easily generalizable to other economic contexts. Additionally, since our data predominantly consists of migrants arriving after the age of 18, we could not effectively apply the instrumental strategy proposed by Bleakley & Chin (2004), which focused on childhood immigrants to isolate exogenous migration decisions from the labor market perspective. Furthermore, the sample is cross-sectional, and utilising longitudinal studies to track immigrants over time could provide a more detailed insight into how language proficiency and other factors impact their labor and social trajectory.

Finally, the paper acknowledges the limitation regarding the self-selection bias of immigrants, a factor that could potentially hinder the interpretation of our findings. In our analysis of host language use at work, we focused solely on employed individuals, a group that may exhibit self-selection tendencies. Consequently, our results may be influenced by various unobserved characteristics, such as individual preferences, rather than solely by language proficiency.

References

Aparicio-Fenoll, A., & Di Paolo, A. (2023). Language Economics. In Handbook of Labor, Human Resources and Population Economics (pp. 1-23). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57365-6_411-1

Aparicio-Fenoll, A. (2018). English proficiency and mathematics test scores of immigrant children in the US. Economics of Education Review, 64, 102-113. https://doi.org/10.1016/j.econedurev.2018.04.003

Aoki, Y., & Santiago, L. (2024). Where to live? English proficiency and residential location of UK migrants. Journal of Economic Behavior & Organization, 221, 73-93. https://doi.org/10.1016/j.jebo.2024.03.015

Aoki, Y., & Santiago, L. (2018). Speak better, do better? Education and health of migrants in the UK. Labour Economics, 52, 1-17. https://doi.org/10.1016/j.labeco.2018.03.003

Barone, G., & Mocetti, S. (2011). With a little help from abroad: the effect of low-skilled immigration on the female labour supply. Labour Economics, 18(5), 664-675. https://doi.org/10.1016/j.labeco.2011.01.010

Becker, G. S. (1976). The economic approach to human behavior (Vol. 803). University of Chicago Press.

Beenstock, M., Chiswick, B. R., & Repetto, G. L. (2001). The effect of linguistic distance and country of origin on immigrant language skills: Application to Israel. International Migration, 39(3), 33-60. https://doi.org/10.1111/1468-2435.00155

Bergman, M. E., Watrous-Rodriguez, K. M., & Chalkley, K. M. (2008). Identity and language: Contributions to and consequences of speaking Spanish in the workplace. Hispanic Journal of Behavioral Sciences, 30(1), 40-68. https://doi.org/10.1177/0739986307311255

Berman, E., Lang, K., & Siniver, E. (2003). Language-skill complementarity: returns to immigrant language acquisition. Labour Economics, 10(3), 265-290. https://doi.org/10.1016/S0927-5371(03)00015-0

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. American Economic Journal: Applied Economics, 2(1), 165-192. https://doi.org/10.1257/app.2.1.165

Bleakley, H., & Chin, A. (2008). What holds back the second generation? the intergenerational transmission of language human capital among immigrants. Journal of human resources, 43(2), 267-298. https://doi.org/10.3368/jhr.43.2.267

Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. Review of Economics and Statistics, 86(2), 481-496. https://doi.org/10.1162/003465304323031067

Bloemen, H. (2013). Language proficiency of migrants: the relation with job satisfaction and matching. http://dx.doi.org/10.2139/ssrn.2263642

Budría, S., & Martínez-de-Ibarreta, C. (2021). Education and skill mismatches among immigrants: The impact of host language proficiency. Economics of Education Review, 84, 102145. https://doi.org/10.1016/j.econedurev.2021.102145

Budría, S., Colino, A., & Martínez de Ibarreta, C. (2019). The impact of host language proficiency on employment outcomes among immigrants in Spain. Empirica, 46(4), 625-652. https://doi.org/10.1007/s10663-018-9414-x

Budría, S., Martinez de Ibarreta, C., & Swedberg, P. (2017). The impact of host language proficiency across the immigrants' earning distribution in Spain. IZA Journal of Development and Migration, 7, 1-27. https://doi.org/10.1186/s40176-017-0094-2

Budría, S., & Swedberg, P. (2015). Education and earnings: how immigrants perform across the earnings distribution in Spain.

Budría, S., & Swedberg, P. (2012). The impact of language proficiency on immigrants' earnings in Spain. http://dx.doi.org/10.2139/ssrn.2170645

Carliner, G. (1981). Wage differences by language group and the market for language skills in Canada. Journal of Human Resources, 384-399. https://doi.org/10.2307/145627

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. American Economic Review, 103(2), 690-731. https://doi.org/10.1257/aer.103.2.690

Chiswick, B. R., & Miller, P. W. (2010). Occupational language requirements and the value of English in the US labor market. Journal of Population Economics, 23, 353-372. https://doi.org/10.1007/s00148-008-0230-7

Chiswick, B. R., & Miller, P. W. (2008). A test of the critical period hypothesis for language learning. Journal of Multilingual and Multicultural Development, 29(1), 16-29. https://doi.org/10.2167/jmmd555.0

Chiswick, B. R., & Miller, P. W. (2007). Earnings and occupational attainment: immigrants and the native born. http://dx.doi.org/10.2139/ssrn.978751

Chiswick, B. R., & Miller, P. W. (1999). Language skills and earnings among legalized aliens. Journal of Population Economics, 12, 63-89. https://doi.org/10.1007/s001480050091

Chiswick, B. R. (1998). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. Journal of Population Economics, 11, 253-271. https://doi.org/10.1007/s001480050068

Chiswick, B. R., & Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. Journal of Labor Economics, 13(2), 246-288. https://doi.org/10.1086/298374

Chiswick, B. R. (1991). Speaking, reading, and earnings among low-skilled immigrants. Journal of Labor Economics, 9(2), 149-170. https://doi.org/10.1086/298263

Clarke, A., & Isphording, I. E. (2017). Language barriers and immigrant health. Health Economics, 26(6), 765-778. https://doi.org/10.1002/hec.3358

Davia, M. A., Wang, T., & Gámez, M. (2022). Language proficiency and immigrants' employment outcomes in Spain. RIEM. Revista Internacional de Estudios Migratorios, 12(2), 132-160. https://doi.org/10.25115/riem.v12i2.6322

Di Paolo, A., & Raymond, J. L. (2012). Language knowledge and earnings in Catalonia. Journal of Applied Economics, 15(1), 89-118. https://doi.org/10.1016/S1514-0326(12)60005-1

Dustmann, C., & Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. The Economic Journal, 113(489), 695-717. https://doi.org/10.1111/1468-0297.t01-1-00151

Dustmann, C., & Van Soest, A. (2002). Language and the earnings of immigrants. ILR Review, 55(3), 473-492. https://doi.org/10.1177/001979390205500305

Education First (2023). English Proficiency Index 2023. Retrieved from: https://www.ef.com/wwen/epi/downloads/

Epstein, G. S., & Gang, I. N. (2010). A political economy of the immigrant assimilation: Internal dynamics. In Migration and Culture (Vol. 8, pp. 325-339). Emerald Group Publishing Limited. https://doi.org/10.1108/S1574-8715(2010)0000008019

Froese, F. J., & Peltokorpi, V. (2011). Cultural distance and expatriate job satisfaction. International Journal of Intercultural Relations, 35(1), 49-60. https://doi.org/10.1016/j.ijintrel.2010.10.002

Froese, F. J. (2010). Acculturation experiences in Korea and Japan. Culture & Psychology, 16(3), 333-348. https://doi.org/10.1177/1354067X10371138

Ghio, D., Bratti, M., & Bignami, S. (2023). Linguistic Barriers to Immigrants' Labor Market Integration in Italy. International Migration Review, 57(1), 357-394. https://doi.org/10.1177/0197918322110

Greene, M. L., Way, N., & Pahl, K. (2006). Trajectories of perceived adult and peer discrimination among Black, Latino, and Asian American adolescents: patterns and psychological correlates. Developmental Psychology, 42(2), 218-236. https://doi.org/10.1037/0012-1649.42.2.218

Guven, C., & Islam, A. (2015). Age at migration, language proficiency, and socioeconomic outcomes: evidence from Australia. Demography, 52(2), 513-542. https://doi.org/10.1007/s13524-015-0373-6

Imai, S., Stacey, D., & Warman, C. (2019). From engineer to taxi driver? Language proficiency and the occupational skills of immigrants. Canadian Journal of Economics/Revue canadienne d'économique, 52(3), 914-953. https://doi.org/10.1111/caje.12396

Instituto Nacional de Estadística. (2021). Encuesta de Características Esenciales de la Población y las Viviendas. Metodología 2021. Retrieved from: https://www.ine.es/metodologia/metodologia_ECEPOV_2021.pdf

Isphording, I. E. (2015). What drives the language proficiency of immigrants? IZA World of Labor. https://doi.org/10.15185/izawol.177

Isphording, I. E., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. Journal of Economic Behavior & Organization, 105, 30-50 https://doi.org/10.1016/j.jebo.2014.03.027

Isphording, I. E., & Otten, S. (2013). The costs of Babylon—linguistic distance in applied economics. Review of International Economics, 21(2), 354-369. https://doi.org/10.1111/roie.12041

Isphording, I. E., & Otten, S. (2011). Linguistic distance and the language fluency of immigrants. Ruhr Economic Paper, (274). http://dx.doi.org/10.2139/ssrn.1919474

Lenneberg, E. H. (1967). The biological foundations of language. Hospital Practice, 2(12), 59-67. https://doi.org/10.1080/21548331.1967.11707799

Lønsmann, D. (2014). Linguistic diversity in the international workplace: Language ideologies and processes of exclusion. Multilingua, 33(1/2), 89-116. https://doi.org/10.1515/multi-2014-0005

McManus, W., Gould, W., & Welch, F. (1983). Earnings of Hispanic men: The role of English language proficiency. Journal of Labor Economics, 1(2), 101-130. https://doi.org/10.1086/298006

Miyar-Busto, M., Díaz, F. J. M., & Gutiérrez, R. (2019). Immigrants' educational credentials leading to employment outcomes: The role played by language skills. Revista Internacional de Organizaciones, (23), 167-191. https://doi.org/10.17345/rio23.167-191

Ministerio de Asuntos Económicos y Transformación Digital. (2021). Informe de Situación de la Economía Española 2021. Retrieved from: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/asuntos-economicos/Documents/2021/290721-Informe-de-Situacion-Economia-espanola-2021.pdf

Ministerio de Trabajo y Economía Social. (2022). Observatorio: Seguimiento de indicadores de empleo de la Estrategia Europa 2020/2030 (Septiembre 2022). Secretaría de Estado de Empleo y Economía Social, Subdirección General de Estadística y Análisis Sociolaboral. Retrieved from: https://www.mites.gob.es/ficheros/ministerio/sec_trabajo/analisis_mercado_trabajo/pnr/observatorio/2022/Septiembre/OBSERVATORIO.pdf

Miranda, A., & Zhu, Y. (2013). English deficiency and the native–immigrant wage gap. Economics Letters, 118(1), 38-41. https://doi.org/10.1016/j.econlet.2012.09.007

Nelson, M. (2014). 'You need help as usual, do you?': Joking and swearing for collegiality in a Swedish workplace. Multilingua, 33(1-2), 173-200. https://doi.org/10.1515/multi-2014-0008

OECD. (2008). International Migration Outlook, Annual Report 2008. OECD, Paris. https://doi.org/10.1787/migr_outlook-2008-en

Peltokorpi, V. (2007). Intercultural communication patterns and tactics: Nordic expatriates in Japan. International Business Review, 16(1), 68-82. https://doi.org/10.1016/j.ibusrev.2006.12.001

Rendon, S. (2007). The Catalan premium: language and employment in Catalonia. Journal of Population Economics, 20, 669-686. https://doi.org/10.1007/s00148-005-0048-5

Rodríguez, R. V., & Yepes, G. R. (2018). The impact of the host-country language on international adjustment: Spanish engineers in Germany. Lengua Y Migración/Language and Migration, 10(1), 79-108. http://hdl.handle.net/10017/33781

Scaheffer, P. V., & Bukenya, J. O. (2010). Assimilation of foreigners in former West Germany. International Migration, 52(4), 157-174. https://doi.org/10.1111/j.1468-2435.2010.00617.x

Swedberg, P. (2010). The impact of education and host language skills on the labor market outcomes of immigrants in Spain. Investigaciones de Economia de la Educacion, 5, 798-824.

Toh, S. M., & DeNisi, A. S. (2007). Host country nationals as socializing agents: A social identity approach. Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 28(3), 281-301. https://doi.org/10.1002/job.421

Yao, Y., & Van Ours, J. C. (2015). Language skills and labor market performance of immigrants in the Netherlands. Labour Economics, 34, 76-85. https://doi.org/10.1016/j.labeco.2015.03.005

Zorlu, A., & Hartog, J. (2018). The impact of language on socioeconomic integration of immigrants. http://dx.doi.org/10.2139/ssrn.3170274

Appendices

Table A. Classification of mother tongue languages within the "Other" category into linguistic families

Variable Code	Language family	Languages within the category	N	% of the group
1	Afro-Asiatic	Amazingh, Amharic, Hausa, Hebrew, Maltese, Tigre	58	1.02%
2	Chelja	Chelja	55	0.97%
3	Berber	Berber	148	2.60%
4	Tamazight	Tamazight	94	1.65%
5	Albanian	Albanian	12	0.21%
6	Altaic	Mongolian, Korean, Chinese	583	10.24%
7	Armenian	Armenian	48	0.84%
8	Austronesian	Bikol, Cebuano, Filipino, Ilocano, Indonesian, Laotian, Malagasy	35	0.61%
9	Tai-Kadai	Thai, Vietnamese	54	0.95%
10	Baltic	Latvian, Lithuanian	49	0.86%
11	Bulgarian	Bulgarian	506	8.88%
12	Slavic	Czech, Croatian, Slovak, Slovenian, Macedonian, Belarusian, Bosnian, Swedish, Kashubian, Montenegrin	109	1.91%
13	Polish	Polish	239	4.20%
14	Russian	Russian	502	8.81%
15	Ukrainian	Ukrainian	243	4.27%
16	West Germanic	Afrikaans, Danish, Scottish, Frisian, Welsh, Icelandic, Irish	33	0.58%
17	North Germanic	Luxembourgish, Norwegian, Swedish, Yola	44	0.77%
18	Dutch	Dutch	144	2.53%
19	Hindi	Hindi	80	1.40%
20	Indo-Iranian	Kurdish, Marathi, Nepali, Punjabi, Persian, Romani, Sindhi	75	1.32%
21	Urdu	Urdu, Punjabi	154	2.70%
22	Kartvelian	Georgian	32	0.56%
23	Niger-Congo	Akan, Bambara, Bengali, Bissa, Bubi, Diola, Edo, Ewe, Fang, Fula, Igbo, Kombe, Kwasio, Lingala, Lingala, Mandinga, Mashi, Mossi, Serer, Soninke, Swahili, Twi, Yoruba	270	4.74%
24	Wolof	Wolof	114	2.00%
25	Quechua	Quechua	61	1.07%
26	Romance	Aranese, Arhuaco, Asturian, Breton, Ibicenco, Mallorcan, Moldovan, Neapolitan, Romansh, Sardinian, Sechuran, Sicilian	50	0.88%
27	Portuguese	Portuguese	1,012	17.77%
28	Guaraní	Guarani	140	2.46%
29	Turkic	Kazakh, Turkish, Uzbek	40	0.70%
30	Uralic	Estonian, Finnish, Hungarian	87	1.53%
31	Hellenic	Greek	17	0.30%
32	Japonic	Japanese	27	0.47%
33	Other	Aimara, Malayalam, Tamil, Telugu, Chuj, Quiché, Hmong, Other, Sign Language	580	10.18%
		Total individuals with “other” mother tongue	5695	100%

Note: Languages are classified within each linguistic family based on information from ethnologue.com, a reference source providing information and statistics on the world's living languages. Classification criteria also consider the number of speakers for each language. Languages with more than 10 observations or those presenting classification challenges due to their unique characteristics are placed in separate categories. In cases where individuals indicate multiple languages within the mother tongue "other" category, only the first language listed is considered for simplicity in the study.

Table B. Immigrants by country of birth

Spanish-speaking countries	*N*	% of group	Non-Spanish-speaking countries	*N*	% of group
Colombia	1,931	17.98%	Morocco	3,510	25.01%
Ecuador	1,621	15.09%	Romania	2,403	17.12%
Venezuela	1,596	14.86%	Cyprus	965	6.88%
Argentina	1,131	10.53%	Germany	741	5.28%
Peru	711	6.62%	China	558	3.98%
Bolivia	629	5.86%	Brazil	512	3.65%
Dominican Republic	624	5.81%	Italy	489	3.48%
Cuba	621	5.78%	France	464	3.31%
Uruguay	373	3.47%	Monaco	412	2.94%
Paraguay	370	3.44%	Bulgaria	377	2.69%
Honduras	343	3.19%	Ukraine	317	2.26%
Mexico	217	2.02%	Switzerland	272	1.94%
Chile	188	1.75%	Russia	264	1.88%
Nicaragua	183	1.70%	Albania	227	1.62%
Equatorial Guinea	72	0.67%	Senegal	196	1.40%
El Salvador	71	0.66%	Algeria	172	1.23%
Guatemala	22	0.20%	Pakistan	172	1.23%
Panama	19	0.18%	United Kingdom	151	1.08%
Andorra	12	0.11%	India	137	0.98%
Costa Rica	6	0.06%	Ireland	130	0.93%
Haiti	2	0.02%	Moldova	117	0.83%
Total Spanish-speaking countries	10,742	100%	Nigeria	98	0.70%
			United States of America	97	0.69%
			Philippines	76	0.54%
			Ghana	57	0.41%
			Greece	55	0.39%
			Mali	54	0.38%
			Lithuania	43	0.31%
			Gambia	42	0.30%
			Armenia	40	0.29%
			Czech Republic	35	0.25%
			Australia	35	0.25%
			Sweden	29	0.21%
			Guinea	29	0.21%
			Canada	29	0.21%
			Georgia	27	0.19%
			Mauritania	27	0.19%
			Other European countries	26	0.19%
			Netherlands	24	0.17%
			Bangladesh	24	0.17%
			Guinea-Bissau	23	0.16%
			Ivory Coast	22	0.16%
			Turkey	21	0.15%
			Japan	20	0.14%
			Other	516	3.68%
			Total non-Spanish-speaking countries	14,035	100%

Note: The "Other" category for non-Spanish-speaking countries includes those countries with fewer than 20 individuals per country.

Table C. Descriptive statistics conditional on language skills

	Conditional on Spanish proficiency								
	All sample			Man			Woman		
Variable	Proficiency	No-proficiency	Statistical sign. of difference	Proficiency	No-proficiency	Statistical sign. of difference	Proficiency	No-proficiency	Statistical sign. of difference
Sex									
● Female	0.562	0.538	0.005						
Age	41.242	41.722	0.005	41.071	42.377	0.000	41.375	41.159	0.340
Working	0.757	0.653	0.000	0.796	0.731	0.000	0.723	0.557	0.000
Spanish workplace use	0.827	0.419	0.000	0.814	0.407	0.000	0.839	0.439	0.000
Non-Spanish-speaking country	0.489	0.97	0.000	0.516	0.98	0.000	0.468	0.962	0.000
Region									
● Eastern Europe	0.058	0.082	0.000	0.048	0.08	0.000	0.066	0.083	0.003
● Western Europe	0.267	0.197	0.000	0.277	0.211	0.000	0.259	0.185	0.000
● Northern Africa	0.006	0.014	0.000	0.008	0.016	0.001	0.005	0.011	0.000
● Sub-Saharan Africa	0.109	0.492	0.000	0.136	0.485	0.000	0.089	0.497	0.000
● Latin America	0.529	0.043	0.000	0.497	0.028	0.000	0.555	0.057	0.000
● Asia & Oceania	0.023	0.167	0.000	0.027	0.171	0.000	0.019	0.163	0.000
● Australia & North America	0.007	0.005	0.195	0.006	0.008	0.315	0.008	0.003	0.008
Age at arrival	22.626	26.964	0.000	22.265	27.359	0.000	22.907	26.624	0.000
Educational attainment									
● Higher education	0.316	0.093	0.000	0.284	0.087	0.000	0.34	0.099	0.000
● Post-secondary	0.282	0.14	0.000	0.286	0.137	0.000	0.279	0.142	0.000
● Secondary	0.321	0.349	0.001	0.344	0.383	0.001	0.303	0.319	0.125
● Primary	0.081	0.418	0.000	0.086	0.393	0.000	0.078	0.44	0.000
Mother tongue									
● Spanish	0.641	0.103	0.000	0.625	0.084	0.000	0.654	0.119	0.000
● Co-official	0.014	0.002	0.000	0.015	0.002	0.000	0.013	0.003	0.000
● English	0.037	0.068	0.000	0.038	0.072	0.000	0.037	0.065	0.000
● French	0.045	0.027	0.000	0.048	0.037	0.049	0.043	0.019	0.000
● Italian	0.021	0.01	0.000	0.028	0.012	0.000	0.015	0.009	0.014
● German	0.019	0.016	0.161	0.018	0.014	0.263	0.02	0.017	0.391
● Romanian	0.095	0.099	0.452	0.095	0.102	0.367	0.095	0.097	0.853
● Arabic	0.078	0.397	0.000	0.099	0.376	0.000	0.062	0.415	0.000
● Other	0.151	0.373	0.000	0.141	0.384	0.000	0.158	0.364	0.000

Note: The sample conditional on Spanish proficiency includes 20,996 proficiency individuals, with 9,204 (43.8%) men and 11,792 (56.2%) women, and 4,077 no-proficiency individuals, with 1,884 (46.2%) men and 2,193 (53.8%) women. The subsample conditional on the working population for the variable "Spanish workplace use" includes 14,914 proficiency individuals, with 7,280 (48.8%) men and 7,634 (51.2%) women, and 2,191 non-proficiency individuals, with 1,364 (62.3%) men and 827 (37.7%) women. The table presents the p-values from the t-test, which assess statistical significance of differences in means between samples of proficient and non-proficient immigrants.

Table D. First-stage regression results

	Dependent variable: Spanish proficiency						
	(1) Age at arrival 15	(2) Age at arrival 14	(3) Age at arrival 12	(4) Age at arrival 10	(5) Age at arrival 9	(6) Age at arrival 7	(6) Age at arrival x non-Spanish-speaking country
Instrument	-0.013***	-0.012***	-0.012***	-0.011***	-0.011***	-0.011***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	23,046	23,046	23,046	23,046	23,046	23,046	23,046
F	57.46	54.86	50.13	46.16	44.57	42.04	39.04
(p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: The table shows the coefficient for the identifying instrument on Spanish proficiency for each age at arrival cutoff: 15, 14, 12, 10, 9, and 7, as well as for the non-Spanish-speaking country dummy interacted with age at arrival. The first-stage estimates are based on the model with basic controls and the outcome: probability of employment ("working" variable). The F-test results shown in the table correspond to the F-test for excluded instruments. The p-value associated to the F-test is presented below. The cutoff selection criterion is based on the results of this test. The first regression models include controls for gender, age and its square, and fixed effects for country of birth and age at arrival. The working sample, excluding those engaged in household tasks, consists of 23,063 individuals, with 11,014 men (47.7%) and 12,049 women (52.3%). Tests with different age cutoffs have also been conducted for samples separated by men and women and for the extended control model, all presenting similar results to those reported. Statistics are robust to heteroskedasticity and clustering by country of birth. *** p<0.01, ** p<0.05, * p<0.1.

Table E. IV regressions: Speaking proficiency vs. writing proficiency

	Dependent variable: Probability of employment		
	(1)	(2)	(3)
VARIABLES	All sample	Man	Woman
Panel 1. Speaking proficiency skills			
Speaking Proficiency	-0.030	0.140	-0.186
	(0.116)	(0.121)	(0.179)
Sex	-0.100***		
	(0.019)		
Age	0.029***	0.033***	0.024***
	(0.003)	(0.004)	(0.003)
Age2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
N	23,046	10,998	12,027
R-squared	0.020	0.020	-0.015
adj. R^2	0.0183	0.0161	-0.0186
F	27.07	29.11	24.19
Panel 2. Writing proficiency skills			
Writing proficiency	-0.024	0.115	-0.144
	(0.093)	(0.101)	(0.138)
Sex	-0.099***		
	(0.020)		
Age	0.029***	0.033***	0.024***
	(0.003)	(0.004)	(0.003)
Age2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
N	23,046	10,998	12,027
R-squared	0.021	0.016	-0.010
adj. R^2	0.0191	0.0119	-0.0132
F	55.14	76.22	34.81

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. The F-test results shown in the table correspond to the F-test for excluded instruments. The results are presented in the basic model, but we do not find statistically significant changes when the extended controls are included. *** p<0.01, ** p<0.05, * p<0.1.

Table F. IV regressions: Sample restrictions

	Dependent variable: Probability of employment			Dependent variable: Spanish use in the workplace		
VARIABLES	(1) All sample	(2) Man	(3) Woman	(4) All sample	(5) Man	(6) Woman
Panel 1. Exclusion of individuals from non-Spanish-speaking countries with Spanish as their mother tongue.						
Spanish Proficiency	-0.037	0.080	-0.153	0.707***	0.728***	0.635***
	(0.073)	(0.078)	(0.117)	(0.081)	(0.109)	(0.092)
Sex	-0.100***			0.009		
	(0.021)			(0.008)		
Age	0.030***	0.033***	0.026***	0.004	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
Age2	-0.000***	-0.000***	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	20,141	9,550	10,570	14,864	7,456	7,382
R-squared	0.019	0.016	-0.013	-0.073	-0.094	-0.034
adj. R^2	0.0167	0.0119	-0.0173	-0.0766	-0.101	-0.0403
F	73.54	84.28	56.29	47.19	55.40	31.34
Panel 2. Exclusion of individuals from non-Spanish-speaking countries with a non-Spanish mother tongue but with Hispanic-origin parents.						
Spanish Proficiency	-0.023	0.101	-0.130	0.686***	0.673***	0.668***
	(0.079)	(0.085)	(0.118)	(0.085)	(0.111)	(0.101)
Sex	-0.104***			0.011		
	(0.021)			(0.008)		
Age	0.029***	0.031***	0.025***	0.001	-0.000	0.002
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)
Age2	-0.000***	-0.000***	-0.000***	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	20,932	9,950	10,959	15,395	7,768	7,605
R-squared	0.021	0.014	-0.010	-0.059	-0.058	-0.044
adj. R^2	0.0189	0.00982	-0.0137	-0.0616	-0.0637	-0.0501
F	68.06	93.69	43.79	46.09	63.41	26.16
Panel 3. Exclusion of bilingual autonomous communities.						
Spanish Proficiency	0.024	0.127	-0.056	0.848***	0.806***	0.853***
	(0.071)	(0.085)	(0.130)	(0.133)	(0.142)	(0.150)
Sex	-0.097***			0.012		
	(0.022)			(0.008)		
Age	0.028***	0.033***	0.022***	0.006	0.003	0.008
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)
Age2	-0.000***	-0.000***	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	15,895	7,553	8,320	11,676	5,875	5,783
R-squared	0.024	0.018	0.001	-0.132	-0.116	-0.122
adj. R^2	0.0215	0.0125	-0.00391	-0.137	-0.124	-0.131
F	64.39	84.89	40.91	44.15	57.90	26.09

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. The F-test results shown in the table correspond to the F-test for excluded instruments. The results are presented in the basic model, but we do not find statistically significant changes when the extended controls are included. *** p<0.01, ** p<0.05, * p<0.1.

Table G. IV regressions: Excluding individuals who always use English at work

	Dependent variable: Probability of employment		
	(1)	(2)	(3)
VARIABLES	All sample	Man	Woman
Spanish Proficiency	-0.023	0.113	-0.137
	(0.091)	(0.095)	(0.137)
Sex	-0.101***		
	(0.020)		
Age	0.029***	0.033***	0.024***
	(0.003)	(0.004)	(0.003)
Age2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
N	22,358	10,648	11,688
R-squared	0.022	0.016	-0.010
adj. R^2	0.0197	0.0120	-0.0134
F	55.53	69.05	38.54

Note: Standard errors clustered by country of birth are in parentheses. All models control for age at arrival and country of birth dummies. The F-test results shown in the table correspond to the F-test for excluded instruments. The estimates use a subsample that excludes individuals who report that the language they always use at work is English. The results are presented in the basic model, but we do not find statistically significant changes when the extended controls are included. *** p<0.01, ** p<0.05, * p<0.1.