Journal of International Migration and Integration
https://doi.org/10.1007/s12134-026-01419-5

RESEARCH

From Fluency to Employment: How State and Regional Language Skills Shape Immigrants' Labour Market Outcomes

Josep Ubalde[1] · **Amado Alarcón**[2] · **Cecilio Lapresta-Rey**[1]

Received: 5 December 2025 / Accepted: 6 July 2026
© The Author(s) 2026

Abstract
It has been widely demonstrated that proficiency in the host country's language is important for the labour market integration of immigrants. However, few studies have analysed the role of immigrants' language skills in bilingual contexts. Drawing on representative data from five Spanish autonomous communities, we investigate the differential role of fluency in Spanish—the state language—and in the regional co-official languages (Catalan, Basque, and Galician). The results show that fluency in Spanish is associated with a higher likelihood of being employed, whereas fluency in a regional language shows no measurable association with this outcome. However, when considering access to highly skilled occupations, a different pattern emerges: fluency in two of the three regional languages (Catalan and Basque) is positively associated with skilled employment, while Spanish shows no significant association. These findings suggest a dual function of language skills: Spanish acts as a general gateway to the labour market, while regional languages function as symbolic and context-specific capital, facilitating access to skilled occupations. The article challenges a uniform view of linguistic capital and underscores the relevance of local identity dynamics in shaping migrants' labour trajectories. It concludes by suggesting the value of moving beyond mere course provision and pointing to the importance of promoting both the effective acquisition of language skills by immigrants and their recognition by employers and institutions as part of professional development pathways.

Keywords Labour Market Outcomes · International migrantion · Employment · Skilled employment · Language skills · Regional languages · Spain

Extended author information available on the last page of the article

Published online: 18 July 2026

Introduction

Securing employment is a crucial first step for immigrants to meet basic needs such as housing, food, and healthcare, while also facilitating greater community integration and access to social benefits. More generally, labour market integration, is a key component of immigrant well-being and of broader social cohesion in host societies. One crucial factor in the labour market integration process is the acquisition of the host society's language (De Vroome & van Tubergen, 2010). It is well established that proficiency in the host country's language is strongly associated with immigrants' labour market outcomes—whether in terms of employability, earnings, or occupational status. A considerable body of research has provided evidence in this regard, drawing on diverse methodological approaches and theoretical perspectives (Auer, 2017; Cheng et al., 2021; Chiswick & Miller, 2010; Kanas & Kosyakova, 2023). Substantial evidence also exists for the Spanish labour market—the context in which this study is conducted—particularly with regard to employability (Isphording, 2013; Miyar-Busto et al., 2020; Budría et al., 2019; Davia et al., 2024).

However, the linguistic reality in many host countries is more complex, as the state language often coexists with one or more regional languages (see, for example, the European Charter for Regional or Minority Languages). The academic literature is relatively scarce when it comes to assessing the value of these regional or minority languages in the labour market integration of immigrants. On the one hand, a substantial body of work has focused on immigrants' heritage languages (Pendakur & Pendakur, 2002; Shin & Alba, 2009; Ubalde & Heyman, 2021). These studies, however, address the languages that immigrants bring with them, not the regional languages they find upon arrival. On the other hand, several studies have investigated the labour market returns to regional languages (Drinkwater and O'Leary, 1997; Henley & Jones, 2005; Christofides & Swidinsky, 2010; Quella & Rendon, 2012; Grenier & Nadeau, 2016), but not in relation to international immigrants and labour market integration processes. The advantages that regional languages may confer in the case of immigrants can differ markedly from those experienced by native-born populations, due to factors such as differentiated occupational segmentation, distinct screening or prioritisation mechanisms, and issues of discrimination or linguistic devaluation related to non-native or accented speech.

The question we pose is whether regional language proficiency is associated with immigrants' employment prospects and their chances of holding a highly skilled occupation. The question lies at the intersection of individual decisions, collective outcomes, and public policy: For immigrants, learning a regional language requires a significant investment. Whether they undertake it depends largely on labour market incentives: if the language improves labour market outcomes, it can enhance their economic integration. If not, the opportunity cost may discourage acquisition, potentially limiting their labour market prospects. This individual decision-making process also has broader implications. When regional languages offer economic returns, immigrants' interests align with those

of language communities, contributing to linguistic maintenance. Conversely, the absence of returns may leave immigrants outside these communities, potentially accelerating language shift. Understanding the labour market value of regional languages is therefore relevant for both immigrants' economic opportunities and policies promoting integration and linguistic diversity.

Our study addresses this gap by analysing immigrant populations in five bilingual regions of Spain—Catalonia, Galicia, the Balearic Islands, the Valencian Community, and Euskadi—where Spanish coexists with a regional co-official language (Catalan, Galician, or Basque). The varying levels of linguistic vitality and the different economic structures across these regions allow us to examine how the returns to regional language skills differ across contexts, strengthening the generalizability of our findings.

The analysis reveals that both state and regional language skills play a significant role in immigrants' labour market integration. While fluency in Spanish is clearly associated with increased employability, none of the regional co-official languages have a measurable effect on the likelihood of being employed. However, the reverse holds true for access to skilled occupations. In this case, Spanish fluency is not associated with access to skilled employment, whereas fluency in Catalan (in both Catalonia and the Balearic Islands) and Basque (in Euskadi) shows a positive correlation with the likelihood of attaining high-skilled employment. These results, however, vary across immigrant groups, reflecting differences in their structural position and the symbolic meanings attached to language acquisition.

The remainder of this article proceeds as follows: we begin by outlining the theoretical framework and reviewing the relevant literature; we then present the sociolinguistic context of the five bilingual regions under study; this is followed by the methodological design and the presentation of the results; finally, we conclude with a discussion of the findings.

Theoretical Framework

The Positive Impact of Language Acquisition: Main Explanations

There is extensive support in the literature for the view that immigrants who acquire the host country's language significantly improve their employment prospects. Scholars have also put forward various explanations for why this positive effect occurs, which can broadly be summarised into four key mechanisms: increased productivity, advantages in screening processes, improved access to institutional resources, and the expansion of social capital.

From a human capital perspective, workers' skills are rewarded in the labour market because they represent investments in abilities that enhance productivity and the economic value of firms. Language can be viewed as both a component of work processes—being essential for communication with clients, colleagues, or suppliers—and as an output, such as written or spoken messages. Thus, immigrants who are proficient in the host country's language may become more productive and are therefore more likely to find employment and under better conditions (e.g., higher

wages, occupational status) (Chiswick & Miller, 2010; Auer, 2017; Budría et al., 2019; Cheng et al., 2021; Kanas & Kosyakova, 2023). Language has also been conceptualised as a mechanism for the transfer of human capital in immigrants, as it enables them to mobilise and apply skills previously acquired in their country of origin (Chiswick & Miller, 2012).

Another explanation stems from signalling theory, which holds that educational credentials and other "signals" help employers infer the abilities of potential employees. These signals serve to reduce uncertainty regarding the candidate's capacity and productivity. Proficiency in the host society's language is rewarded not only because it is a skill that enhances productivity, but also because employers interpret it as a proxy for other, less observable traits such as learning capacity, adaptability, perseverance, or commitment (Christofides & Swidinsky, 2010; Grin et al., 2011: 76–77). This interpretation of quantitative findings is also supported from qualitative and ethnographic research, which highlights language requirements are not always neutral indicators of productivity, but may also reflect institutional expectations regarding communicative norms, social belonging, and integration into local linguistic environments (Roberts, 2021; Codó & Garrido, 2014; Van der Worp et al., 2018). Recent research extends signalling theory to account for group-specific returns to language skills. Schmaus (2020) shows that immigrants from groups facing unfavourable attitudes in Germany gain substantially more from improving host-country language proficiency than groups with more positive stereotypes. For these groups, higher language skills reduce the salience of ethnic origin, thereby mitigating discrimination. Conversely, groups with favourable attitudes do not experience such differential returns, suggesting that language can act as a compensatory signal that counteracts negative stereotypes.

Language acquisition also facilitates access to a range of resources that may prove critical for finding employment or advancing in the labour market—for instance, career counselling, country-specific training, qualification recognition, and continuing education. It can also improve access to financial benefits and social assistance, which often represent a crucial threshold for escaping poverty and gaining the stability needed to seek work or start a business (Kaida, 2013).

Finally, social capital theory suggests that immigrants who possess or develop strong networks in the host country are better positioned to improve their employment prospects thanks to the information and support these connections provide (Portes, 1995). The theory holds that immigrants who build ties both within their own communities and with members of the native-born population tend to access better job opportunities. Intra-community connections can offer job information and referrals, while ties with natives may open doors to more stable and better-paid employment. In this sense, acquiring the host society's language becomes a crucial tool for immigrants seeking to expand their networks with host-country nationals (Hoehne & Michalowski, 2016) and to access community language resources that can facilitate employment (Nam et al., 2025).

We draw on these theoretical mechanisms to motivate our analysis and interpret our findings. It is not the aim of this article to directly test which specific mechanism is at work. Instead, we use this theoretical framework to generate hypotheses (see below) and later interpret the observed patterns in light of existing theory.

What Happens when Two Languages Coexist within the Same Territory? Contrasting Hypotheses

It is well established that proficiency in the host country's language plays a key role in immigrants' labour market integration. Less clear, however, is what occurs when two native languages coexist within the same territory. The vitality of each language within a given region or country may be a critical factor in answering this question. Broadly speaking, language vitality depends on the number of speakers and those who understand the language (Grin, 1990). This is not solely determined by local demographics in areas where a specific language is predominantly spoken, but must also be considered within the broader global context (De Swaan, 1993). More comprehensive definitions of vitality incorporate additional dimensions such as the status of the speaker community and the level of institutional support it receives (Giles et al., 1977). Status refers to the economic, social, and symbolic position of a linguistic group in society and contributes to the visibility and perceived value of its language. Institutional support encompasses backing from entities such as the education system, media, and government.

We argue that the vitality of a language influences the four key channels through which language competence leads to improved labour market outcomes. The greater the vitality, the greater its potential to enhance productivity, strengthen signalling effects in recruitment processes, increase access to relevant resources, and expand social networks. In cases where there is a pronounced disparity in language vitality within a region, the majority language tends to assume all functions that may generate economic benefits for new speakers. This is particularly true in contexts where the minority language lacks institutional support, such as the absence of revitalisation policies. However, even when the vitality of the minority language is limited, it may still offer unique opportunities and resources. We might consider jobs or businesses in which communication in the minority language is essential for effective performance. Likewise, there may be tightly knit communities where minority language use is deeply embedded, requiring immigrants to acquire the language in order to build relationships. Additionally, certain resources or opportunities may be available exclusively to speakers of the minority language.

The academic literature does not provide a straightforward answer to these possible scenarios. We can summarise the main findings in terms of the relative status of the languages analysed and their associated labour market returns. Studies conducted in the United States show that bilingualism, when it involves maintaining a heritage language, tends to offer no advantage or may even be detrimental for immigrants (Pendakur & Pendakur, 2002; Shin & Alba, 2009; Alarcon et al., 2014; Author et al., 2021). Regardless of the vitality of these heritage languages in the host society, the overwhelming dominance of English in the global context leaves little scope for minority languages to yield benefits. These studies also suggest that immigrants' languages may act as ethnic signals, potentially leading to discrimination, while in some cases being informally used by employers without corresponding economic returns (Alarcón et al., 2014).

In Catalonia (Spain), where both Catalan and Spanish hold official status, several studies have been conducted. Despite Catalan's lower vitality relative

to Spanish, existing evidence suggests that proficiency in Catalan yields benefits in terms of employability, occupational status, and earnings (Di Paolo & Raymond, 2012; Quella & Rendon, 2012; Rendon, 2007). Similarly, research in Wales shows that Welsh speakers experience lower unemployment rates and higher wages than individuals who speak only English (Drinkwater & O'Leary, 1997; Henley & Jones, 2005).

The Canadian case offers a valuable comparative perspective, but the literature presents divergent findings depending on the target population, the region of Canada, and the time horizon. Studies focusing on the general population highlight the distinction between language knowledge and language use at work. Christofides and Swidinsky (2010) show that English yields substantial returns in Quebec, the French-speaking region; in contrast, French in the rest of Canada yields more limited returns that are not associated with use at work (Christofides & Swidinsky, 2010). Grenier and Nadeau (2016) confirm this asymmetry in Montreal: Francophones benefit substantially from using English at work, while Anglophones do not gain from using French—reflecting English's status as a global lingua franca. Studies focusing specifically on immigrant populations provide a more nuanced picture for our question. Grondin (2005), tracking recent immigrants over four years, finds that English proficiency improves wages and skilled job outcomes nationally, whereas French shows no clear effect. However, Fleury et al. (2025), following two cohorts over up to ten years, show that French proficiency in Quebec was a significant asset for both employment and income, and its value has increased over time.

In short, the academic literature presents mixed evidence: while in some countries or regions minority language proficiency appears to confer labour market advantages, in others it shows no significant effect. Crucially, except for Grondin (2005) and Fleury et al. (2025) in Canada, none of the studies reviewed focus specifically on immigrant populations, which makes it more challenging to formulate a well-grounded hypothesis in this area. Consequently, we propose two competing hypotheses concerning the labour market returns to bilingualism among immigrants, specifically in relation to native languages in the host society and their impact on employability and access to skilled occupations.

Brief Contextual Note: Five Bilingual Autonomous Communities in Spain

There are five bilingual regions (autonomous communities) in Spain where Spanish and a regional language hold co-official status throughout the entire territory (this excludes, for example, the region of Navarre, where Basque, the regional language, is only official in part of the territory). Below, we briefly outline the sociolinguistic context in each region, followed by a summary table presenting their relative linguistic vitality.

Balearic Islands

The sociolinguistic evolution of the Balearic Islands has been characterised by the progressive replacement of Catalan (Balearic variety) by Spanish, particularly since the eighteenth century, a process that intensified in the twentieth century due to migration and tourism. Today, both languages are co-official; however, Spanish dominates in urban and tourist areas, while Catalan remains more present in rural zones. Demographically, there is asymmetric bilingualism: although most residents understand Catalan, its use in social settings is relatively low. Despite institutional protection (notably in education and public administration), Catalan enjoys lower social status, with Spanish prevailing in public life and economic activity (Bretxa et al., 2019). The islands' economy is heavily dependent on hospitality and tourism (see Appendix 2), sectors where public-facing service roles are concentrated. The predominance of Spanish in tourist areas may limit actual returns to Catalan.

Catalonia

Catalonia represents the most developed example of bilingualism in Spain, where Catalan—despite heavy repression during the Franco era—has been successfully revitalised through post-1980s normalisation policies. Catalan is now widely used in administration, education, culture, and the media, and the school system—where Catalan is the main language of instruction—has been praised for promoting integration and high levels of bilingual competence (Vila et al., 2017). Demographic changes, driven by large-scale internal migration in the twentieth century and more recent international immigration, have diversified linguistic practices. In metropolitan areas such as Barcelona, Spanish tends to dominate everyday interactions, especially among young people and migrant groups. Even so, Catalan retains strong institutional backing, a dynamic cultural sphere and broad social legitimacy. The preference for Catalan has been linked to upward mobility among regional elites, with a strong regional nationalist identity (Block, 2015). Key debates today concern how to balance individual linguistic rights with collective planning, ensure Catalan's presence in digital environments, and maintain intergenerational transmission (Sendra & Vila, 2019). The region has a diversified economy (see Appendix 2) with a sizeable industrial base and a large public sector, where Catalan proficiency is often a formal requirement for employment in public administration, education, and healthcare.

Valencian Community

The linguistic evolution of the Valencian Community has been marked by the historical predominance of Valencian (a variety of Catalan) until the seventeenth century, followed by its gradual decline in favour of Spanish—particularly between the eighteenth and twentieth centuries and most notably during the Franco dictatorship. With the restoration of democracy, Valencian regained its co-official status in 1982, and efforts were made to promote its use in education, public administration, and

cultural domains. Spanish is the language usually spoken by approximately 75% of the population, while around 25% predominantly use Valencian, especially in rural and inland areas. Both languages are co-official, but Spanish holds greater presence in urban settings, the media, and commercial life. Valencian retains a strong identity value in historically Valencian-speaking rural areas. It is incorporated into the education system (via bilingual models), regional and local administration, and receives cultural subsidies, though its effective use in public life remains limited (Bretxa et al., 2019; Hernàndez i Dobon, 2016). The region's economic structure combines agriculture, industry, and tourism, with a noticeable concentration of public-facing skilled occupations (see Appendix 2).

Galicia

Its sociolinguistic trajectory has shifted from a situation where Galician was traditionally associated with rural and family life, and Spanish with urban and institutional spheres, to a more normalised presence of Galician following its recognition as a co-official language in 1981. Currently, around 56% of the population speak Galician, though its use is significantly lower in urban areas. Spanish, historically regarded as more prestigious, continues to dominate in formal and media contexts. Galician receives legal and educational support from institutions, but its social status remains lower, and its use has declined among younger generations. Both languages coexist in administration and education, although Spanish continues to have greater visibility in the media and cultural life (O'Rourke & Ramallo, 2015; Instituto Galego de Estadística, 2023). Its economy is more reliant on the primary sector than that of Catalonia and the Basque Country (see Appendix 2). Galician is less often a requirement for employment in public administration, education, and healthcare.

Euskadi

Euskadi (the Basque Country) presents a unique bilingual context, as Basque (Euskera) is an unrelated, non-Indo-European language with historically uneven transmission. By the late twentieth century, especially during the Franco regime, Basque was severely weakened in urban and industrial areas. Since the 1980s, however, strong revitalisation policies—supported by robust regional institutions—have significantly expanded Basque's presence in education, media and cultural life. The school system offers three models (Spanish-medium A, bilingual B and Basque-medium D), with Model D becoming increasingly dominant (Government of the Basque Country, 2025). This has produced new generations with high competence in Basque, although proficiency does not always translate into regular everyday use. Linguistic behaviour remains geographically uneven: Basque is strongest in Gipuzkoa and parts of Bizkaia, while Spanish predominates in the Bilbao metropolitan area and some coastal zones. The region benefits from extensive institutional support, active cultural production, and a widespread network of adult Basque language learning centres. Current challenges centre on converting passive or school-based knowledge into active use, expanding Basque-speaking networks in urban contexts,

and adapting revitalisation strategies to a diverse and highly mobile population (Madariaga, 2019; Eustat, 2021). Euskadi has been a centre of industrialisation in Spain (see Appendix 2) and proficiency in Basque is a formal requirement for many public-sector positions.

Although a comprehensive analysis of each regional language's vitality across these regions would require a more detailed examination, Table 1 provides a qualitative assessment of their linguistic vitality (see also the detailed breakdown of each dimension in Appendix 1). The categories are based on the framework developed by Giles et al. (1977), which conceptualises vitality as a combination of status, demography, and institutional support. Status and institutional support are assessed drawing on the sociolinguistic literature on language normalisation and public presence in each region (see above). Demographic indicators are derived from official statistical sources. The resulting categories are qualitative summaries synthesising this information, closely aligning with the earlier summary developed by Ros et al. (1987).

Methodology

Data

Our analysis draws on data from the 2014 and 2021 special ad hoc modules of the Spanish Labour Force Survey (LFS). These modules are conducted in parallel with the main LFS, using one-sixth of the quarterly LFS sample. The sampling design of the modules therefore follows that of the main LFS. The 2014 and 2021 modules enable the examination of immigrants' labour market situations (in terms of employability and occupational level) and provide detailed information on respondents' language fluency in five bilingual regions. This data is unique in allowing for

Table 1 Linguistic vitality of the regional co-official languages in the five bilingual Autonomous Communities in Spain

Region (Language)	Status	Demography	Support	Overall vitality
Catalonia (Catalan)	High	Medium–High	Medium–High	Medium–High
Euskadi (Basque)	Medium	Low	Medium	Medium
Valencian Community (Catalan)	Medium–Low	Medium	Low	Medium–Low
Balearic Islands (Catalan)	Medium–Low	Medium–High	Low	Medium–Low
Galicia (Galician)	Medium–Low	Medium	Low	Medium–Low

Source: Authors' own elaboration based on the sociolinguistic literature cited in the "Brief contextual note: Five bilingual autonomous communities in Spain" section and on official statistical sources (Idescat, Eustat, IGE, Generalitat Valenciana, IBESTAT). See Appendix 1 for a detailed breakdown of each dimension

an analysis of the labour market advantages associated with bilingualism in the regional languages of the host societies. By using both survey waves in our analysis, we benefit from a larger dataset and a broader observational period. Given the seven-year gap between waves and the LFS's rotating panel design (households are interviewed for six consecutive quarters and then exit), there is no systematic overlap of participants across waves. In this article, we focus on the immigrant population having the following characteristics: individuals born abroad, of working age, and economically active. The size of this pooled subsample (2014 and 2021) across the five bilingual regions is $N = 5323$ ($n_{Catalonia} = 1815$, $n_{Balearic} = 547$, $n_{Valencia} = 1249$, $n_{Galicia} = 1235$, $n_{Euskadi} = 477$).

Variables

The outcomes: We employ two distinct labour market outcomes as dependent variables: employment status and working in a highly skilled occupation. The first is a binary variable with two categories: employed (1) and not employed (0). This classification is provided by the Spanish National Statistics Institute (INE), based on the following definition: "individuals aged 16 or over who, during the reference week, worked for at least one hour in exchange for remuneration in cash or kind, or who were temporarily absent from their job due to illness, holiday, etc. Employed individuals are subdivided into self-employed (employers, own-account workers without employees, and independent workers) and employees (in either the public or private sectors)". This employment measure groups together stable full-time employment with marginal or precarious work. Accordingly, our employability findings capture access to any form of employment, not job quality.

This limitation is partially offset by our skilled employment outcome, which is typically associated with better working conditions. It is also operationalised as a binary variable: working in a highly skilled occupation (1) versus not doing so (0). We follow the Eurostat classification of highly skilled occupations, which includes the following categories (Eurostat, 2023: 83): managers, professionals, and technicians and associate professionals (ISCO-08 major groups 1–3). Although these broad occupational categories encompass considerable heterogeneity, they are robustly associated with higher wages and job quality (Eurofound, 2017). Our findings should therefore be interpreted as capturing a broad—yet meaningful—proxy for higher-quality labour market integration.

The predictors: Our key independent variables relate to oral language fluency in both the state language (Spanish) and the regional languages of the five autonomous communities—namely, Catalan, Basque, and Galician. Note that Catalan is a co-official language in three of the five regions. Specifically, we focus on self-assessed oral proficiency, as this is the only language dimension available in the survey. The Labour Force Survey's ad hoc module provides descriptors for self-assessment of oral proficiency based on the Common European Framework of Reference for Languages (CEFR) developed by the Council of Europe (see Appendix 3). We harmonised the response scales across the two survey years into a four-level ordinal scale, ranging from low to high oral fluency: "no ability

or beginner level", "intermediate level", "advanced level", and "native-like fluency". We merged the two lowest categories from the 2021 survey ("beginner" and "no skill") to match the 2014 specification, where they are presented under the same response category. These linguistic fluency variables are analysed both as ordinal variables (1–4) and as a set of dummy variables, enabling us to capture not only the linear pattern associated with oral proficiency but also how each fluency level relates to the outcomes.

The use of self-assessed measures may introduce misclassification error, as individuals may inaccurately report their true level of language proficiency. Dustmann and van Soest (2002) show that such measurement error can significantly downward-bias simple regression estimates of the effect of language fluency on labour market outcomes. Nonetheless, our measurement approach relies on respondents' self-evaluation based on clear, activity-based "can-do" descriptors, indicating specific tasks the respondent can perform in the target language. While this cannot fully eliminate the risk of overreporting or underreporting, it is expected to mitigate bias to a meaningful extent.

The covariates: The analysis controls for a comprehensive set of sociodemographic characteristics, selected based on previous literature examining the effects of language proficiency on employability and occupational status. Specifically, we include all available covariates that are likely to be jointly correlated with both language skills and labour market outcomes. These include: current age (in years) and age squared, age at arrival (note that these variables together captures how long an individual was in the country), dummy variables for educational attainment (low, medium, high), dummy variables for marital status (married, divorced/widowed vs single), dummy variables for province of residence within each bilingual region (up to four dummy variables), gender (male vs female), and country of birth dummies, grouped into nine geographical regions (Europe, Africa, North America, Central America and the Caribbean, South America, East Asia, West Asia, South and Southwest Asia, and Oceania). Finally, we account for socio-economic background by including the number of years of schooling completed by both parents (mother and father).

Basic descriptive statistics for the main variables are provided in Table 2, for the full sample and disaggregated by region. Appendix 4 offers a detailed summary of all variables included in the analysis. Focusing on our main variables of interest, we observe that the average employment rate among working-age immigrants is 55.6%, compared to 63.4% among the native-born population. Regarding employment in highly skilled occupations, 20.9% of immigrants are in such positions, markedly lower than the 36.6% observed among natives. While these rates vary somewhat across regions, the gap between immigrants and natives remains consistent. In terms of oral language fluency, the average score on the 1–4 scale for the state language (Spanish) is 3.07, indicating an advanced level of fluency. In contrast, the average fluency score for regional languages is just 1.53, falling below the intermediate level. Again, there is some regional variation, but the predominance of the regional language is a constant across all five regions.

Table 2 Descriptive statistics — proportions for dichotomous variables; mean (and standard deviation) for continuous variables

	All five regions	Balearic Islands	Catalonia	Valencian Community	Galicia	Euskadi
Labour market outcomes						
Employed	55.6	56.7	58.4	51.1	53.7	56.6
Highly skilled occupation	20.9	18.2	22.3	19.8	24.6	14.5
Language Fluency Variables						
Regional lang. fluency scale [ranged 1–4]	1.53(1.0)	1.52(1.0)	1.59(0.9)	1.31(0.8)	2.36(1.2)	1.13 (0.7)
Spanish fluency scale [ranged 1–4]	3.07(1.1)	3.09 (1.0)	2.99(1.1)	3.05(1.0)	3.45(0.8)	3.23 (0.9)
Fluency in the regional language:						
No skill/beginner	60.2	60.5	53.1	74.5	31.8	81.6
Intermediate level	18.3	15.7	23.2	12.7	23.0	6.6
Advanced level	11.4	12,8	15.5	6.0	18.7	3.0
Native level	5.5	5.4	4.0	3.2	2.6	2.3
Fluency in Spanish:						
No skill/beginner	10.0	8.8	11.7	10.2	4.1	5.8
Intermediate level	16.4	19.0	15.9	18.0	10.2	15.9
Advanced level	26.2	24.4	27.2	26.5	22.3	26.1
Native level	46.5	47.3	43.6	44.8	63.3	51.9

Data source: Spanish Labour Force Survey, Immigrant Module (2014 and 2021), weighted sample. Percentages may not add up to 100 due to item non-response. Standard deviations of the quantitative variables are reported in parentheses

Data Analysis

We rely on linear probability models to analyse the dichotomous indicators of labour market integration — namely, employment status and occupational attainment (Long, 1997). For each outcome we first provide unconditional estimates, and we then provide the results controlling for the full set of covariates. As we used the pooled data for both sample waves (2014 and 2021) we account for time effects by including an indicator for the year in which individuals participated in the survey. This allows us to control for any variation due to the specific timing of the survey. Given the inherent heteroskedasticity of this kind of models, we apply heteroskedasticity-consistent standard errors to all estimates.

For the models predicting highly skilled occupational attainment, we additionally include in the equation the inverse Mills ratios to account for differential selection into employment (Heckman, 1979). The subsample of individuals who succeed in securing employment may be selectively composed along multiple dimensions, including unobservable traits such as motivation. By incorporating the Mill's ratios into the equations, we adjust for this non-random selection. Mill's ratios are calculated for each individual based on a first-stage probit model estimating the probability of being employed. The selection equation includes the same set of regressors as the equation of interest plus the variables marital status and a dummy indicating whether a child aged 10 or younger is currently living in the responding household as the exclusion restriction variable.

In an observational study such as this, there is the possibility that unobserved factors affecting both labour market performance and language skills may upwardly bias the estimates. Our strategy consists of controlling for as many sources of heterogeneity as possible, based on the information available in the survey, with particular attention to variables identified as relevant in previous research (see the "Variables" section on covariates). Naturally, we lack information on some variables that may also be correlated with both outcomes, such as motivation or innate ability. To assess the sensitivity of statistically significant estimates to unobserved heterogeneity, we apply the approach proposed by Frank (2000) (see the "Results" section).

Results

Main Regression Results: State vs. Regional Language Fluency

Table 3 summarises the estimated coefficients for the linguistic fluency variables, where the control variables are added sequentially to the regression model in each column. Columns 1 and 2 present the raw estimates of language fluency. The coefficients indicate a positive correlation of Spanish fluency on the probability of being employed, whereas no statistically significant correlation is found for regional language fluency. The opposite pattern emerges when the outcome is employment in highly skilled occupations: while fluency in regional languages shows a significant positive correlation, this is not the case for Spanish fluency.

Table 3 Results for LPM regressions of labour market outcomes (employed and employed in a skilled occupation) on the ordinal scales of regional language and Spanish fluency

	(1) Employed	(2) Skilled	(3) Employed	(4) Skilled	(5) Employed	(6) Skilled
Regional language fluency [ranged 1–4]	−0.010 (0.024)	0.063*** (0.007)	0.006 (0.008)	0.021* (0.009)	0.008 (0.008)	0.024*** (0.009)
Spanish fluency [ranged 1–4]	0.072*** (0.007)	−0.006 (0.013)	0.032*** (0.010)	0.007 (0.015)	0.043*** (0.010)	0.003 (0.016)
Educational level (ref: low level)						
Medium level			0.040** (0.016)	0.086*** (0.023)	0.046** (0.017)	0.053* (0.024)
High level			0.113*** (0.017)	0.517*** (0.036)	0.135*** (0.019)	0.452*** (0.019)
Age			0.073*** (0.003)	0.055* (0.022)	0.070*** (0.040)	0.049* (0.022)
Age squared			−0.001*** (<0.001)	−0.001* (<0.001)	−0.001*** (<0.001)	−0.001* (<0.001)
Age at arrival			−0.004*** (0.001)	−0.006*** (0.001)	−0.004*** (0.001)	−0.006*** (0.001)
Gender (ref: men)			−0.115*** (0.013)	−0.127** (0.034)	−0.115*** (0.014)	−0.104** (0.032)
Region of origin (ref: Europe)						
Africa			−0.185*** (0.021)	−0.205*** (0.060)	−0.183*** (0.023)	−0.144* (0.056)
North America			−0.029 (0.084)	0.286** (0.099)	0.021 (0.089)	0.341*** (0.096)
Central America and the Caribbean			0.024 (0.028)	0.012 (0.034)	0.041 (0.031)	0.024 (0.036)
South America			−0.004 (0.018)	−0.024 (0.022)	−0.015 (0.020)	−0.028 (0.022)
East Asia			0.149** (0.046)	0.08 (0.066)	0.142** (0.053)	0.035 (0.069)
Western Asia			−0.017 (0.122)	0.241 (0.140)	−0.084 (0.137)	0.165 (0.159)
South and Southwest Asia			−0.036 (0.038)	−0.11* (0.049)	−0.066 (0.041)	−0.116* (0.054)
Oceania			0.039 (0.160)	0.462* (0.187)	0.048 (0.185)	0.491* (0.197)
Marital status (ref: single)						
Married			−0.021 (0.016)		−0.024 (0.018)	
Widowed			−0.046 (0.051)		−0.074 (0.054)	
Separated/divorced			0.011 (0.026)		0.028 (0.028)	
Child aged 10 or younger			0.002 (0.014)		0.011 (0.015)	
Father's education (years)					−0.002 (0.003)	0.007* (0.003)
Mother's education (years)					−0.005 (0.003)	0.008* (0.003)
N	5307	2965	5234	2965	4589	2553

Table 3 (continued)

Data source: Spanish Labour Force Survey, Immigrant Module (2014 and 2021). All models include survey wave and province of residence. Robust standard errors in parentheses

$^{*}\ p<0.05,\ ^{**}\ p<0.01,\ ^{***}\ p<0.001$

When all sociodemographic variables are included in the model (columns 3 and 4), the previously significant coefficients are substantially reduced in both cases, suggesting that a considerable part of the observed correlation is driven by individual, geographic, and educational differences. Nonetheless, the same pattern persists: while Spanish fluency is associated with a higher probability of being employed, it shows no significant association with the likelihood of holding a higher-skilled position—a pattern that is inverted for regional language fluency. Thus, after applying full controls, we conclude that Spanish fluency is positively associated with employment opportunities, corresponding to an increase of 3.2 percentage points in the probability of being employed for each additional fluency level. In contrast, fluency in regional languages increases the likelihood of being in a highly skilled occupation by an average of 2.1 percentage points per level. Finally, the inclusion of parental education variables (columns 5 and 6) slightly raises these estimates, to 4.3 and 2.4 percentage points, respectively.

In the methodological section of the article ("Data Analysis" section), we noted that these models assume the absence of uncontrolled variables. We measure how much omitted variable bias is necessary to invalidate inference by means of the ITCV and RIR (Frank, 2000; Frank et al., 2013). Regarding the regional languages, the ITCV is 0.121, meaning that a confounder would need to produce a combined correlation of at least 0.121 with both language fluency and skilled employment in order to explain away the estimated coefficient. This represents a level of confounding of moderate magnitude. When compared to the observed covariates in the model, such a threshold would require an omitted factor with a strength comparable to that of the most powerful predictors—such as high educational attainment or certain regions of origin. The RIR tells us how many cases in a study would need to be replaced with cases showing no effect to make a statistically significant result become non-significant. To nullify the estimate of regional language fluency on skilled employability using the threshold of statistical significance (with null hypothesis = 0 and alpha = 0.05), 27% of the estimate of 0.024 would have to be due to bias. This implies that to nullify the inference one would expect to have to replace almost one third of the sample (677 cases) with cases for which the effect is 0. The statistically significant estimate of Spanish fluency on employability appears to be even more robust, as to nullify it, 77.4% of the estimate of 0.043 would have to be due to bias. Moreover, the correlation with a confounding variable should be at least 0.148.

Table 4 shows the estimate for each level of language fluency, based on the models in columns (3) and (4) of Table 3, which include sociodemographic controls (but exclude parental education variables, as their inclusion substantially reduces the sample size without significantly altering the estimates). With regard to employability (column 1), we observe that even an intermediate level of Spanish fluency

Table 4 LPM estimates for language fluency (categorical) on labour market outcomes (employment and employment in a highly skilled occupation)

	(1)	(2)
	Employed	Skilled
Regional language fluency (ref: No skill/beginner)		
Intermediate level	0.011	0.050*
	(0.018)	(0.022)
Advanced level	0.021	0.055*
	(0.022)	(0.022)
Native level	0.029	0.068*
	(0.027)	(0.033)
Spanish fluency (ref: No skill/beginner)		
Intermediate level	0.097***	−0.020
	(0.027)	(0.051)
Advanced level	0.157***	0.069
	(0.029)	(0.061)
Native level	0.120***	0.004
	(0.032)	(0.056)
N	5234	2965

Data source: Spanish Labour Force Survey, Immigrant Module (2014 and 2021). Both models include the same set of covariates as those in columns (3) and (4) of Table 3. Robust standard errors in parentheses

*$p<0.05$, **$p<0.01$, ***$p<0.001$

increases the probability of employment by 9.7% percentage points compared to those with beginner or no proficiency. The increase in percentage points associated with advanced fluency is notable, 15.7, while the change associated with native-level proficiency is 12, suggesting diminishing returns at the highest level. Turning to employment in highly skilled occupations (column 2), we find that an intermediate level of regional language proficiency raises the probability by 5 percentage points relative to those with no or beginner-level skills, while the estimated difference is 5.5 for advanced proficiency and 6.8 for native-level speakers. To contextualize these findings: an immigrant with native-level fluency in Spanish has a 12-percentage point higher probability of being employed compared to an immigrant with no proficiency—an estimate comparable to that of higher education (11.3 percentage points). For regional languages, the estimated difference in skilled employment probability between native-level fluency and no proficiency is 6.8 percentage points, close in magnitude to the estimate for medium-level education (8.6 percentage points).

Subgroup Analysis

The preceding results help address the question of whether fluency in regional languages is expected to be associated with immigrant's labour market outcomes in bilingual contexts. However, the impact of regional language proficiency may vary depending on the characteristics and linguistic vitality of each region. To explore

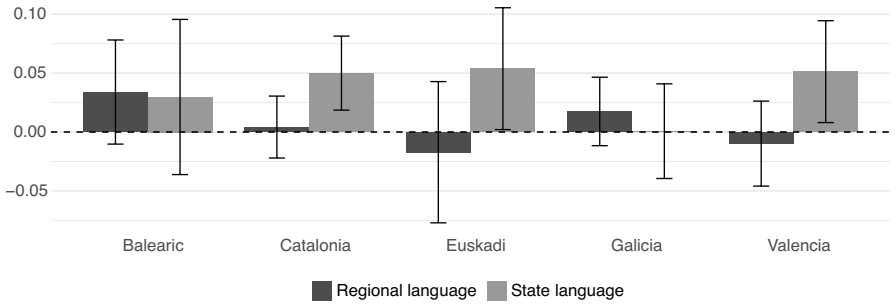


Fig. 1 LPM estimates for language fluency on employment status, by region (Comunidades Autónomas), Note: Data source: Spanish Labour Force Survey, immigrant module (2014 and 2021). Bar heights indicate estimated coefficients; error bars represent 95% confidence intervals. All five regional models include the same set of covariates as those used in columns (3) and (4) of Table 3

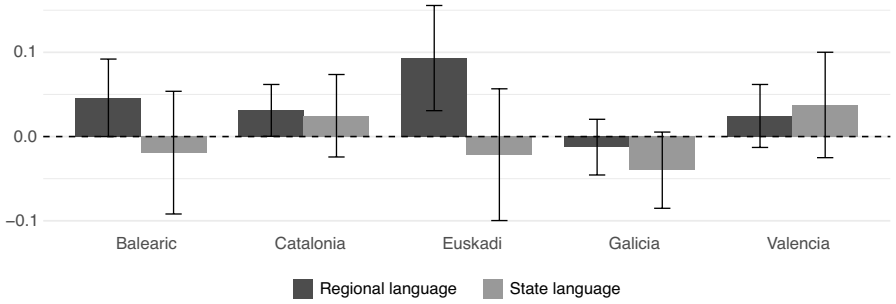


Fig. 2 LPM estimates for language fluency on skilled employment, by region (Comunidades Autónomas), Note: Data source: Spanish Labour Force Survey, immigrant module (2014 and 2021). Bar heights indicate estimated coefficients; error bars represent 95% confidence intervals. All five regional models include the same set of covariates as those used in columns (3) and (4) of Table 3

this heterogeneity, we conducted a stratified analysis based on models (3) and (4) for each of the five bilingual regions. The results are presented in Figs. 1 and 2 (see also Appendix 5), with the outcomes being "employed" and "employed in a highly skilled occupation", respectively. Regarding employability (Fig. 1), we observe that the coefficients for regional language proficiency are not statistically significant in any of the regions. In contrast, Spanish proficiency shows a significant association with employment in Catalonia, Euskadi, and Valencia. As for employment in skilled occupations (Fig. 2), the pattern is reversed: Spanish fluency does not exhibit a significant coefficient in any of the Autonomous Communities, whereas regional language fluency shows a significant association with skilled employment in Catalonia (Catalan), the Balearic Islands (Catalan), and Euskadi (Basque). This association is particularly pronounced in the Euskadi, reaching an increase of 9 percentage points, while in Catalonia and the Balearic Islands it is more modest (3 and 5 pp, respectively).

Table 5 presents the results of the same model for various subsamples: region of origin (Western countries, Africa, Latin America, Asia) and gender (men, women). While we generally observe results similar to those found in the full sample, some variations also emerge. Regarding region of origin, the association between fluency in the regional language and access to skilled employment is statistically significant for individuals from Latin America and Africa, whereas no significant association is observed for those from Western or Asian countries. Spanish proficiency, on the other hand, shows a statistically significant association for immigrants from Western and African countries, but not for those from Latin America or Asia. As for gender, Spanish fluency is significantly associated with employability for both men and women. Regional language fluency is significantly associated with skilled employment for men, but not for women.

Finally, we examined whether the positive association between language skills and skilled employment varies across economic sectors. Table 6 show the results of a model including interaction terms between language fluency and sector (note that we employ interaction terms rather than a simple subgroup analysis, as the latter is not feasible with the Heckman model). Although the coefficients do not reach statistical significance at the conventional level of 5%, the pattern of results is suggestive: the positive association of regional language fluency seems to be specifically concentrated in the public and personal services sector ($p = 0.09$), which includes public administration, education, healthcare, and social services. This finding aligns with the interpretation that regional languages function as symbolic and institutional capital in sectors where interaction with the local population is central and where language requirements are often formalised. No such pattern was observed for other sectors (the primary sector, industry, construction, commerce or market services), though these results should be interpreted with caution given the limited statistical power of the interaction analysis.

Tables 5 to 6. Subgroup analysis tables.

Discussion

This study addresses a gap in the literature on immigrant labour market integration in bilingual contexts by analysing the distinct value of fluency in both the state and regional languages of the host region. Drawing on representative data from five bilingual autonomous communities in Spain, our results indicate that while proficiency in Spanish is positively associated with the likelihood of being employed, fluency in regional languages shows no significant correlation with this labour market outcome. However, the opposite holds true when examining access to highly skilled occupations: regional languages—specifically Catalan in Catalonia and the Balearic Islands and Basque in Euskadi—are positively associated with this outcome, whereas Spanish shows no such association.

These findings suggest a dual role for language skills: the state language, as the majority and dominant language, plays a key role in facilitating extensive integration—that is, general access to employment and likely faster labour market entry. It may serve as a gateway to a broad range of job opportunities. This result aligns

Table 5 LPM estimates for language fluency on labour market outcomes, by region of origin and gender

	Europe		Africa		Latin America		Asia		Woman		Man	
	(1) Employed	(2) Skilled	(3) Employed	(4) Skilled	(5) Employed	(6) Skilled	(7) Employed	(8) Skilled	(9) Employed	(10) Skilled	(11) Employed	(12) Skilled
Regional language fluency [ranged 1–4]	−0.001 (0.013)	0.006 (0.017)	0.019 (0.024)	0.039* (0.019)	0.016 (0.011)	0.031* (0.013)	−0.060 (0.042)	−0.017 (0.051)	−0.002 (0.011)	0.015 (0.011)	0.012 (0.011)	0.030* (0.012)
Spanish fluency [ranged 1–4]	0.031* (0.013)	0.050 (0.027)	0.054* (0.023)	−0.003 (0.024)	−0.004 (0.021)	−0.031 (0.021)	0.018 (0.037)	0.004 (0.036)	0.067*** (0.010)	−0.025 (0.018)	0.037*** (0.011)	0.016 (0.015)
N	1781	1079	864	309	2331	1395	293	159	2887	1487	2420	1478

Data source: Spanish Labour Force Survey, Immigrant Module (2014 and 2021). Both models include the same set of covariates as those in columns (3) and (4) of Table 2, including survey year and controls for province of residence. Robust standard errors in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 6 LPM estimates for language fluency on skilled employment, by economic sector

	Skilled employment
Regional language fluency [ranged 1–4]	0.011 (0.016)
Economic sector (ref: Commerce)	
Primary sector	0.002 (0.082)
Industry	−0.003 (0.045)
Construction	−0.030 (0.045)
Market services	0.054 (0.037)
Public, personal services	0.070 (0.039)
Regional language fluency in Primary sector	0.048 (0.045)
Regional language fluency in Construction	0.051 (0.048)
Regional language fluency in Industry	0.052 (0.046)
Regional language fluency in Market services	0.052 (0.044)
Regional language fluency in Public, personal services	0.078 (0.044)
N	2965

Data source: Spanish Labour Force Survey, Immigrant Module (2014 and 2021). Both models include the same set of covariates as those in columns (3) and (4) of Table 2, including survey wave and controls for province of residence. Robust standard errors in parentheses

with previous research examining the integrative role of host-country language proficiency (Isphording, 2013; Auer, 2017; Cheng et al., 2021; Davia et al., 2024). However, our study also underscores the potential of regional languages to enable intensive integration, characterised by access to highly skilled positions. By simultaneously accounting for the state and regional languages, our models offer a more nuanced understanding than previous research, showing that coexisting languages in a given territory may both play important roles in shaping immigrant integration trajectories.

Our findings are consistent with human capital theory, which holds that language learning enhances productivity and transferability, thereby improving employment prospects (Chiswick & Miller, 2010). They also suggest that not all languages enhance productivity in the same way: while Spanish facilitates general labour market access, the regional co-official language may act as a form of specific capital valued only within certain occupational segments. A possible interpretation of this finding is the signalling hypothesis (Christofides & Swidinsky, 2010; Schmaus, 2020):

Spanish proficiency may serve as a minimum threshold for entering the labour market, while regional language proficiency—costlier and less widespread among immigrants—may function as a signal of deeper integration, local commitment, and adaptability (see also Roberts, 2021, on linguistic gatekeeping in job interviews). These attributes are particularly valued in skilled or public sector jobs, such as those in healthcare, education, public administration, or direct citizen services.

The positive association between regional language fluency and skilled employment is statistically robust and substantively meaningful, with an average gain of 2.4 percentage points per fluency level—a magnitude close to that of medium-level education. Our sensitivity analyses suggest that only substantial biases would be sufficient to invalidate the average effect of regional language fluency. This is a relevant finding in ongoing debates about the economic value of bilingualism for migrants in contexts where regional languages do not dominate all social spheres. Our findings suggest that bilingualism can offer specific advantages (i.e., access to skilled jobs), particularly in contexts where regional languages retain some vitality. This evidence is in line with previous studies showing that proficiency in regional languages such as Welsh, Catalan or French in bilingual territories can yield labour market benefits (Di Paolo & Raymond, 2012; Drinkwater & O'Leary, 1997; Fleury et al., 2025; Henley & Jones, 2005; Quella & Rendon, 2012). Our study extends this evidence by focusing specifically on immigrant labour market integration in different bilingual regions.

These interpretations are consistent with ethnographic research showing that regional language requirements are not neutral indicators of productivity but may reflect institutional expectations about communicative norms and social belonging. In the Catalan context, Codó and Garrido (2014) have documented how migrant support organisations increasingly framed Catalan proficiency as a prerequisite for labour market access. In the Basque Country, Van der Worp et al. (2018) have shown how the promotion of Basque in internationally operating companies positions the language fluency as a valued asset in professional settings. Together, these studies reinforce our interpretation that regional language fluency functions as a signal of local integration and institutional belonging.

However, we find important variation across regions and immigrant subgroups in the benefits associated with regional language proficiency. Among the five bilingual communities examined, statistically significant associations are observed in Catalonia (Catalan), the Balearic Islands (Catalan), and Euskadi (Basque), while no such associations are found in Valencia (Catalan) or Galicia (Galician). These differentiated results can be understood by examining the intersection of three factors: sectoral economic structure, institutional language requirements, and the sectoral concentration of the effects. Catalonia and Euskadi exhibit greater vitality than Galicia or Valencia. Moreover, public and personal services sector, where the correlation between regional language and skilled employment tend to be higher, account for a substantial share of employment and have strong institutional language requirements (proficiency in Catalan or Basque is often a formal requirement for public-sector positions). Our findings also suggest that even in regions with moderate to low vitality—such as the Balearic Islands—regional languages may offer integration-related benefits for immigrants. This apparent puzzle can be explained by the region's

economic structure: the islands have the highest share of services in their GVA, with hospitality (i.e. tourism) dominating. In a labour market characterised by high seasonal turnover, proficiency in Catalan may signal stability or commitment—qualities valued by employers seeking reliable workers in a context of transient employment.

The absence of significant associations in Galicia and the Valencian Community is equally revealing. Both regions have medium–low linguistic vitality. Specifically: In Galicia, the regional language has a weaker institutional presence in the labour market, where the primary sector carries strong weight—a sector in which language skills have the lowest impact—while the public sector, where regional language skills yield the highest returns, has fewer formal language requirements. In the Valencian Community, although the service sector is large, institutional support for Valencian is also weaker, and language requirements are less strictly enforced. In both regions, the identity dimension associated with the language is less pronounced than in Catalonia or the Basque Country. Thus, co-official status alone does not guarantee labour market returns.

We also observe that the association between regional language proficiency and skilled employment is strongest for Latin American and African migrants. In contrast, migrants from Asian or Western countries do not experience significant impacts. This heterogeneity suggests that the labour market value of regional language skills is moderated by both the social position of each group and the symbolic meanings attached to language learning (Schmaus, 2020). For Latin American immigrants, who are native Spanish speakers, learning a regional language may function as a signal of integration and commitment—a symbolic gesture particularly valued in skilled occupations involving public-facing roles (education, healthcare, public administration). For African immigrants, who tend to receive the most negative attitudes from native populations, regional language fluency may be particularly appreciated by employers, especially in roles involving close interaction with the local population.

In contrast, the null effects for Asian immigrants reflect a different mechanism. According to the LFS of Spain (2014 and 2021 waves), 61% of Asian workers are concentrated in hospitality and commerce—sectors where they predominantly operate within ethnic enclave economies (Valenzuela-Garcia et al., 2017). In these contexts, access to employment, information, and resources is mediated by dense co-ethnic networks, and heritage languages constitute the relevant linguistic capital. Even when Asian immigrants acquire regional language skills, these may not translate into labour market advantages because employers, customers, and job referrals are accessed within the ethnic community. For Western immigrants, the absence of associations may reflect their concentration in multinational or English-speaking professional niches where the added value of local languages is diminished.

While our study offers sound and novel evidence on the differentiated role of language proficiency in bilingual contexts, certain limitations must be acknowledged. First, language proficiency is measured through self-assessment. Although, as noted, the questionnaire includes detailed descriptions to aid accurate reporting, some degree of misclassification may persist. Second, although we conducted a range of robustness checks to support our conclusions—including the inclusion of multiple controls, sensitivity analyses, and comparisons across different geographical

areas—the cross-sectional survey design cannot fully rule out the possibility of bias arising from unobserved confounding variables.

A particularly important issue in our study is simultaneity. Our cross-sectional data cannot distinguish between two directions: language proficiency improving employment, or employment improving language proficiency. While both are plausible, natural experiments (e.g., Auer, 2017; Kanas & Kosyakova, 2023) provide strong evidence that language proficiency causally affects employment outcomes. We also argue that the alternative direction is likely less quantitatively important. Among immigrants, most intensive language learning occurs prior to entering the labour market, driven by instrumental motivations (e.g., finding a job). Moreover, access to publicly funded language courses is typically targeted at immigrants seeking employment. We therefore consider it likely that a genuine association from language to employment exists, potentially alongside a feedback loop in the opposite direction.

While we have demonstrated the differential relevance of state and regional languages for immigrant labour market integration, future research should empirically explore the mechanisms through which these languages generate employment benefits. Moreover, a detailed analysis of differences between first- and second-generation immigrants would be valuable; however, given their qualitatively distinct educational pathways, labour market trajectories, and identity formation processes, we leave this question for future research. Finally, although we compare five regions with varying levels of linguistic vitality, our findings are strictly generalisable only to the Spanish context. Broader generalisation would require data from other multilingual settings with diverse sociolinguistic landscapes and institutional frameworks.

Conclusion

This article provides empirical evidence on an underexplored issue: the specific value of regional co-official languages in the labour market integration process of immigrants beyond the state-wide language. Drawing on representative data from five Spanish autonomous communities, our findings suggest that fluency in regional languages may play a role in the labour market integration of immigrants. More specifically, Spanish fluency is significantly associated with a higher likelihood of being employed, whereas regional languages show no meaningful association with this indicator. In contrast, proficiency in Catalan and Basque is associated with a greater probability of accessing skilled occupations.

These findings suggest that knowledge of Spanish is a necessary condition for general access to the labour market, whereas regional languages may function as a form of symbolic and cultural capital, particularly valued in specific occupational niches and institutional contexts. We argue that the identity dimension in these regions—largely neglected in previous research on language and labour market integration—may be especially salient in bilingual environments. Our results challenge a homogeneous view of linguistic capital and suggest that regional language proficiency may operate also as symbolic capital or a signal of valuable integration in certain sectors.

Our results also carry potential implications for language policy. The findings suggest the value of moving beyond generic, functional literacy courses and considering the development of sector-specific programmes that link regional language proficiency to skilled occupations—particularly in public administration, education, and healthcare. However, it is important to recognise that language training alone is unlikely to be sufficient. Any such initiatives should be accompanied by efforts to address structural barriers for groups concentrated in ethnic enclave economies, where language training alone may yield limited labour market returns. This would require, among other measures, facilitating access to professional networks and promoting mobility beyond co-ethnic niches. Furthermore, the symbolic value of bilingualism should be acknowledged, and policies could promote regional language skills as professional assets in job placement services, while encouraging employers to value these competencies in skilled occupations. Ultimately, the effectiveness of these measures will also depend on broader structural factors, including credential recognition, sectoral access, and efforts to counter discrimination.

Appendix 1

Linguistic vitality of the regional co-official languages in the five bilingual Autonomous Communities in Spain (detailed breakdown)

Region	Status	Demography	Support
Catalonia (Catalan)	**High**: co-official status, widespread institutional and social use; historically linked to upward mobility	**Medium–High**: 80.4% of the population aged 15 + are able to speak Catalan; for 36.4% Catalan is the language usually spoken	**Medium–High**: compulsory in education, use in public administration, public media
Euskadi (Basque)	**Medium**: co-official status, but lower social use; historically associated with regional elites and industrialisation	**Low**: 43.3% of the population aged 2 + are able to speak Basque; 19.9% of the population aged 16 + use Basque as much as or more than Spanish at home	**Medium**: strong institutional support, but limited use in administration and media
Valencian Community (Catalan)	**Medium–Low**: co-official status, but weaker institutional presence; historically associated with rural areas and lower social classes	**Medium**: 56.4% of the population aged 15 + are able to speak Valencian; 31.1% use only or mainly Catalan at home	**Low**: limited use in education (bilingual models), administration, and media
Balearic Islands (Catalan)	**Medium–Low**: co-official status, but weaker institutional presence; historically associated with rural areas	**Medium–high**: 80.5% of the population aged 15 + are able to speak Catalan; for 36,8% Catalan is the language usually spoken	**Low**: use in education, but limited in administration and media
Galicia (Galician)	**Medium–Low**: co-official status, but low social use in formal settings; associated with rural and family life	**Medium**: 83.4% of the population aged 5 + can speak Galician well or very well; 24% speak Galician always	**Low**: weak institutional support, non-universal use in education

The assessments for each dimension are based on the sociolinguistic literature cited in this section and on official statistical sources from each autonomous community, specifically

Idescat (Institut d'Estadística de Catalunya). (2018). Enquesta d'usos lingüístics de la població 2018. Barcelona: Idescat

Eustat (Instituto Vasco de Estadística). (2021). Estadística sociolingüística del euskera 2021. Vitoria-Gasteiz: Eustat

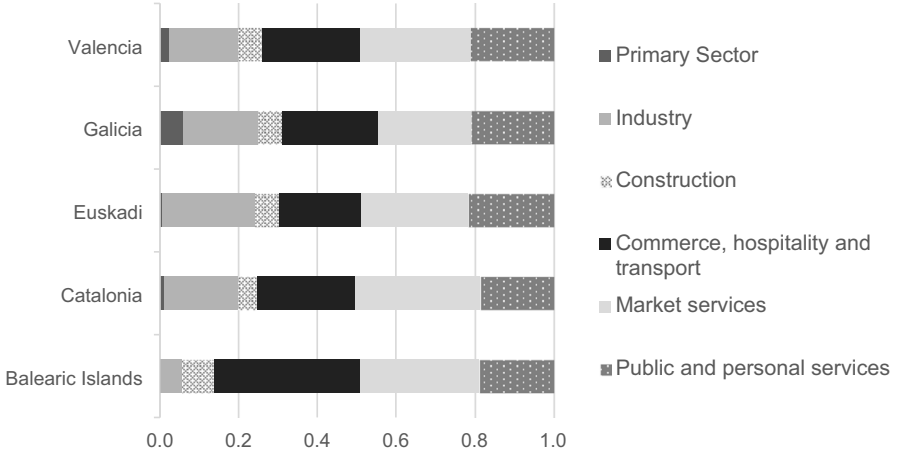
IGE (Instituto Galego de Estadística). (2021). Enquisa estrutural a fogares. Coñecemento e uso do galego. Santiago de Compostela: IGE

Generalitat Valenciana. (2020). Enquesta d'usos lingüístics de la població de la Comunitat Valenciana 2020. València: Generalitat Valenciana

IBESTAT (Institut d'Estadística de les Illes Balears). (2019). Enquesta d'usos lingüístics a les Illes Balears 2019. Palma: IBESTAT

Appendix 2

Percentage of each economic sector in the total Gross Value Added of each Autonomous Community



Appendix 3

Questions and linguistic descriptors from the questionnaire of modules 21 and 14 of the Spanish Labour Force Survey (example referring to regional languages).

How would you rate your spoken proficiency in Catalan/Valencian/Galician/Basque?

5 – *It is your mother tongue or you speak it at a native level*
4 – *Advanced level*
3 – *Intermediate level*
2 – *Beginner level*
1 – *No or almost no language skills*

"Mother tongue" refers to the first language learned at home during childhood and still understood by the individual at the time of the survey. The term "mother tongue" should not be interpreted as necessarily being the language of the mother. In bilingual households, the father's language may be dominant, in the sense that it is the main language used for communication. It is also possible that both the mother's and the father's languages are used; in that case, the individual is considered to have more than one mother tongue. In some countries or contexts, the mother tongue may be referred to as the "first language."

Code 2: Advanced level is to be marked if the respondent believes that they are currently able to express themselves spontaneously and fluently in Spanish. That is, they can articulate ideas and opinions precisely and interact skilfully with other speakers. (Equivalent to level C1 of the Common European Framework of Reference for Languages.)

Code 3: Intermediate level is to be marked if the respondent believes that they are currently able to cope effectively with most situations likely to arise while travelling in areas where Spanish is spoken. They can engage in unprepared conversations on topics that are familiar, of personal interest, or related to everyday life (e.g., family, hobbies, work, travel, and current events). (Equivalent to level B1 of the Common European Framework of Reference.)

Code 4: Beginner level is to be marked if the respondent believes they are currently able to ask and answer simple questions in Spanish on immediate needs or familiar topics, provided the person they are speaking with is willing to repeat or rephrase statements and assist in communicating their intended message. (Equivalent to level A1 of the Common European Framework of Reference.)

Code 5: No or almost no language skills is to be marked if the respondent currently does not speak Spanish at all, or only understands and can use a few words or basic phrases. This category is intended to distinguish basic users (level A1) from individuals with extremely limited or no language ability.

Appendix 4

Descriptive statistics —shares for dichotomous variables, mean (SD) for continuous variables

	All five regions	Balearic Islands	Catalonia	Valencia	Galicia	Euskadi
Labour market outcomes						
Employment status (employed)	55.6	56.7	58.4	51.1	53.7	56.6
Occupation (highly skilled)	20.9	18.2	22.3	19.8	24.6	14.5
Language Fluency Variables						
Regional lang. fluency scale	1.53(1.0)	1.52(1.0)	1.59(0.9)	1.31(0,8)	2.36(1.2)	1.13 (0,7)
Spanish fluency scale	3.07(1.1)	3.09 (1.0)	2.99(1.1)	3.05(1.0)	3.45(0.8)	3.23 (0.9)
Fluency in the regional language:						
No skill/beginner	60.2	60.5	53.1	74.5	31.8	81.6
Intermediate level	18.3	15.7	23.2	12.7	23.0	6.6
Advanced level	11.4	12,8	15.5	6.0	18.7	3.0
Native level	5.5	5.4	4.0	3.2	2.6	2.3

	All five regions	Balearic Islands	Catalonia	Valencia	Galicia	Euskadi
Fluency in Spanish:						
No skill/beginner	10.0	8.8	11.7	10.2	4.1	5.8
Intermediate level	16.4	19.0	15.9	18.0	10.2	15.9
Advanced level	26.2	24.4	27.2	26.5	22.3	26.1
Native level	46.5	47.3	43.6	44.8	63.3	51.9
Sociodemographic variables						
Gender (Women)	52.6	53.9	51.2	52.5	56.7	53.7
Low education	41.2	39.2	44.2	37.3	40.0	42.7
Medium education	30.5	38.1	25.5	35.5	30.4	31.4
High education	28.3	22.7	30.2	27.3	29.6	25.9
Age	39.5(12.1)	39.2(12.6)	38.8(11.7)	40.3(12.5)	40.9(12.3)	39.7(11.5)
Age at arrival	25.2(12.1)	24.7(11.4)	25.5(11.4)	26.2(12.6)	20.7(14.1)	25.1(11.4)
Marital status:						
Single	37.4	42.5	35.8	37.1	39.0	40.3
Married	53.0	49.2	55.3	52.4	49.0	50.4
Divorced	1.3	0.9	1.2	1.0	2.7	2.2
Widowed	8.2	7.4	7.6	9.5	9.3	7.1
Region of origin						
Europe	46.4	32.7	23.5	41.7	40.0	28.3
Africa	18.2	18.0	23.6	12.9	5.4	19.7
North America	0.5	0.2	0.5	0.4	0.8	0.4
Central America and the Caribbean	6.6	3.5	7.6	4.4	7.5	14.2
South America	35.3	40.5	33.6	34.8	43.7	31.3
East Asia	2.3	1.2	2.9	2.2	1.0	2.5
West Asia	0.3	0.09	0.5	0.03	0.5	0.2
South and Southwest Asia	5.1	3.8	7.7	3.3	0.6	3.2
Oceania	0.1	0.0	0.0	0.1	0.2	0.0

Pooled sample from the EAPS immigrant module (2014 and 2021)

Appendix 5

Subgroup analysis (full results)

Table 9 LPM estimates for language fluency on labour market outcomes, by region (CCAA)

	Balearic Islands		Catalonia		Valencian Community		Galicia		Euskadi	
	Employed	Skilled	Employed	Skilled	Employed	Skilled	Employed	Skilled	Employed	Skilled
Regional language fluency [ranged 1–4]	0.034 (0.023)	0.045* (0.024)	0.004 (0.013)	0.031* (0.016)	-0.010 (0.018)	0.024 (0.019)	0.017 (0.015)	-0.013 (0.017)	-0.017 (0.031)	0.093** (0.032)
Spanish fluency [ranged 1–4]	0.03 (0.034)	-0.019 (0.037)	0.050** (0.016)	0.025 (0.025)	0.051* (0.022)	0.038 (0.032)	0.001 (0.021)	-0.04 (0.023)	0.054* (0.026)	-0.024 (0.040)
N	544	313	1811	1035	1241	656	1232	682	474	274

Author Contribution Josep Ubalde: Writing – original draft, Writing – review & editing, Visualization, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. Amado Alarcón: Writing – review & editing, Funding acquisition. Cecilio Lapresta: Writing – review & editing.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by Grant PID2021-122575NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by the European Regional Development Fund (ERDF), A way of making Europe, and by Project 5QM23 funded by the Office for Development and Cooperation (Oficina de Desenvolupament i Cooperació) of the University of Lleida.

Data Availability The study uses the microdata from the 2014 and 2021 special ad hoc modules of the Spanish Labour Force Survey (LFS). The microdata are available from the Spanish National Statistics Institute (Instituto Nacional de Estadística, INE) upon request and subject to its access requirements.

Declarations

Conflict of interest The authors declare no competing interests.

Clinical Trial Number Not applicable.

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

References

Alarcon, A., Di Paolo, A., Heyman, J., & Morales, M. C. (2014). Returns to Spanish-English bilingualism in the new information economy: The health and criminal justice sectors in the Texas border and Dallas-Tarrant counties. In R. M. Callahan & P. C. Gándara (Eds.), *The bilingual advantage: Language, literacy and the US labor market* (pp. 138–158). Multilingual Matters.

Auer, D. (2017). Language roulette – The effect of random placement on refugees' labour market integration. *Journal of Ethnic and Migration Studies, 44*(3), 341–362. https://doi.org/10.1080/1369183X.2017.1304208

Block, D. (2015). Social class in applied linguistics. *Annual Review of Applied Linguistics, 35*, 1–19.

Bretxa, V., Sorolla, N., Torres-Pla, J., Torrijos, A., & Vila i Moreno, F. X. (2019). *Els usos lingüístics als territoris de llengua catalana*. Generalitat de Catalunya. Generalitat de Catalunya. Departament de Cultura. Direcció General de Política Lingüística. https://llengua.gencat.cat/web/.content/documents/publicacions/btpl/arxius/27-btpl-usos-linguistics-territoris.pdf

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

Cheng, Z., Wang, B. Z., Jiang, Z., Taksa, L., & Tani, M. (2021). English skills and early labour market integration: Evidence from humanitarian migrants in Australia. *International Migration*. https://doi.org/10.1111/imig.12889

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/s00148-008-0211-3

Chiswick, B. R., & Miller, P. W. (2012). Negative and positive assimilation, skill transferability, and linguistic distance. *Journal of Human Capital, 6*(1), 35–55. https://doi.org/10.1086/663581

Christofides, L. N., & Swidinsky, R. (2010). The economic returns to the knowledge and use of a second language. *Canadian Public Policy, 36*(2), 137–158. https://doi.org/10.3138/cpp.36.2.137

Codó, E., & Garrido, M. R. (2014). Shifting discourses of migrant incorporation at a time of crisis: Understanding the articulation of language and labour in the Catalan non-governmental sector. *International Journal of Multilingualism, 11*(4), 389–408. https://doi.org/10.1080/14790718.2014.944529

Davia, M. A., Wang, T., & Gámez, M. (2024). Language proficiency and immigrants' employment outcomes in Spain. *Revista Internacional De Estudios Migratorios, 12*(2), 132–160. https://doi.org/10.25115/riem.v12i2.5234

De Swaan, A. (1993). The emergent world language system: An introduction. *International Political Science Review, 14*(3), 219–226. https://doi.org/10.1177/019251219301400302

De Vroome, T., & van Tubergen, F. (2010). The employment experience of refugees in the Netherlands. *International Migration Review, 44*(2), 376–403. https://doi.org/10.1111/j.1747-7379.2010.00810.x

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)60005-9

Drinkwater, S. J., & O'Leary, N. C. (1997). Unemployment in wales: Does language matter? *Regional Studies, 31*(6), 583–591. https://doi.org/10.1080/00343409750131712

Dustmann, C., & van Soest, A. (2002). Language and earnings of immigrants. *Industrial and Labor Relations Review, 55*(3), 473–492. https://doi.org/10.2307/3270613

Eurofound. (2017). *Sixth European Working Conditions Survey – Overview report (2017 update)*. Publications Office of the European Union. (see file:///C:/Users/t4801079/Downloads/ef1634en.pdf)

Eurostat. (2023). *Eurostat regional yearbook 2023 edition*. https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-ha-23-001

Eustat (Instituto Vasco de Estadística). (2021). Estadística sociolingüística del euskera2021. Vitoria-Gasteiz: Eustat

Fleury, C., Piché, V., & Bélanger, D. (2025). Language and immigration: Knowledge of both official languages as an asset on the job market. *Recherches Sociographiques, 66*(1), 89–118.

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research, 29*(2), 147–194.

Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis, 35*(4), 437–460.

Giles, H., Bourhis, R. Y., & Taylor, D. M. (1977). Towards a theory of language in ethnic group relations. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 307–348). Academic Press.

Government of the Basque Country. (2025). *Concepts and definitions: Linguistic models (enrolment).* Department of Education. https://www.euskadi.eus/informacion/directorio-de-centros-conceptos-y-definiciones/web01-a2hestat/es/?utm_source=chatgpt.com

Grenier, G., & Nadeau, S. (2016). English as the lingua franca and the economic value of other languages: The case of the language of work in the Montreal labor market. In M. Gazzola & B.-A. Wickström (Eds.), *The economics of language policy* (pp. 192–214). The MIT Press.

Grin, F. (1990). The economic approach to minority languages. *Journal of Multilingual and Multicultural Development, 11*(1–2), 153–173. https://doi.org/10.1080/01434632.1990.9994393

Grin, F., Sfreddo, C., & Vaillancourt, F. (2011). *The economics of the multilingual workplace*. Taylor & Francis.

Grondin, C. (2005). Knowledge of official languages among new immigrants: How important is it in the labour market? (Cat. no. 89–624-XIE). Statistics Canada. https://doi.org/10.25318/89-624-x2007000-eng

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153–161. https://doi.org/10.2307/1912352

Henley, A., & Jones, R. E. (2005). Earnings and linguistic proficiency in a bilingual economy. *The Manchester School, 73*(3), 300–320. https://doi.org/10.1111/j.1467-9957.2005.00452.x

Hernàndez i Dobon, F. J. (2016). *El tio Canya ha mort: Notes sobre la mecànica sociolingüística del valencià*. Fundació Nexe.

Hoehne, J., & Michalowski, I. (2016). Long-term effects of language course timing on language acquisition and social contacts: Turkish and Moroccan immigrants in Western Europe. *International Migration Review, 50*(1), 133–162. https://doi.org/10.1111/imre.12130

Instituto Galego de Estadística (2023). Encuesta estructural a hogares. Conocimiento y uso del Gallego. https://www.ige.gal/web/mostrar_actividade_estatistica.jsp?idioma=es&codigo=0206004&num_pag=1

Isphording, I. E. (2013). Returns to foreign language skills of immigrants in Spain. *Labour, 27*(4), 443–461. https://doi.org/10.1111/labr.12020

Kaida, L. (2013). Do host country education and language training help recent immigrants exit poverty? *Social Science Research, 42*(3), 726–741. https://doi.org/10.1016/j.ssresearch.2013.01.004

Kanas, A., & Kosyakova, Y. (2023). Greater local supply of language courses improves refugees' labor market integration. *European Societies, 25*(1), 1–36.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications.

Madariaga, J. M. (2019). The Basque Country. In C. Lapresta-Rey & À. Huguet (Eds.), *Multilingualism in European language education* (pp. 98–110). Multilingual Matters.

Miyar-Busto, M., Mato Díaz, F. J., & Gutiérrez, R. (2020). 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

Nam, Y., Richards-Desai, S., & Abraha, A. (2025). Individual English proficiency, community language resources, and immigrants' employment in Western New York: Evidence from the American Community Survey. *International Migration Review, 59*(3), 1657–1677.

O'Rourke, B., & Ramallo, F. (2015). Neofalantes as an active minority: Understanding language practices and motivations for change amongst new speakers of Galician. *International Journal of the Sociology of Language, 2015*(231), 147–165.

Pendakur, K., & Pendakur, R. (2002). Language as both human capital and ethnicity. *International Migration Review, 36*(1), 147–177.

Portes, A. (1995). Economic sociology and the sociology of immigration: A conceptual overview. In A. Portes (Ed.), *The economic sociology of immigration: Essays on networks, ethnicity, and entrepreneurship* (pp. 1–41). Russell Sage Foundation.

Quella, N., & Rendon, S. (2012). Occupational selection in multilingual labor markets: The case of Catalonia. *International Journal of Manpower, 33*(8), 918–937. https://doi.org/10.1108/01437721211280380

Rendon, S. (2007). The Catalan premium: Language and employment in Catalonia. *Journal of Population Economics, 20*(3), 669–686.

Roberts, C. (2021). *Linguistic penalties and the job interview*. University of Toronto Press.

Ros, M., Cano, J. I., & Huici, C. (1987). Language and intergroup perception in Spain. *Journal of Language and Social Psychology, 6*(3–4), 243–259. https://doi.org/10.1177/0261927X8700600308

Schmaus, M. (2020). Ethnic differences in labour market outcomes—The role of language-based discrimination. *European Sociological Review, 36*(1), 82–103.

Sendra, M., & Vila, F. X. (2019). Catalonia. In C. Lapresta-Rey & À. Huguet (Eds.), *Multilingualism in European language education* (pp. 111–124). Multilingual Matters.

Shin, H. J., & Alba, R. (2009). The economic value of bilingualism for Asians and Hispanics. *Sociological Forum, 24*(2), 254–275. https://doi.org/10.1111/j.1573-7861.2009.01099.x

Ubalde, J., & Heyman, J. (2021). Is there a bilingual advantage? the effects of non-English language skills on occupational attainment among immigrants and natives. *International Journal of the Sociology of Language*, 2021(270), 19–38. https://doi.org/10.1515/ijsl-2020-0014

Valenzuela-Garcia, H., Parella, S., & Güell, B. (2017). Revisiting the ethnic enclave economy: Resilient adaptation of small businesses in times of crisis in Spain. *International Journal of Anthropology and Ethnology, 1*(5), 1–13. https://doi.org/10.1186/s41257-017-0008-5

Van der Worp, K., Cenoz, J., & Gorter, D. (2018). Multilingual professionals in internationally operating companies: Tensions in their linguistic repertoire? *Multilingua, 37*(4), 353–375.

Vila, F. X., Lasagabaster, D., & Ramallo, F. (2017). Bilingual education in the autonomous regions of Spain. *Bilingual and multilingual education* (pp. 505–517). Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Josep Ubalde[1] · Amado Alarcón[2] · Cecilio Lapresta-Rey[1]

✉ Josep Ubalde
josep.ubalde@udl.cat

Amado Alarcón
amado.alarcon@urv.cat

Cecilio Lapresta-Rey
cecilio.lapresta@udl.cat

[1] University of Lleida, Lleida, Spain

[2] Rovira i Virgili University, Tarragona, Spain