



<https://doi.org/10.25115/riem.v12i2.6322>

ISSN: 2173-1950

Language Proficiency and Immigrants' Employment Outcomes in Spain

María A. Davia¹, Wang Ting^{2*}, Matías Gámez³

Abstract: This paper estimates returns on proficiency in Spanish as regards employability and occupational prestige in first-generation immigrants. Because of occupational segregation in the Spanish labour market amongst workers from different regions of the world, we distinguish between proficiency in Spanish in general and acquisition of Spanish amongst immigrants whose native language is not Spanish. Our empirical approach allows us to consider two issues that may lead to an underestimation of returns on skills in the host language in occupational prestige. First, given that occupations are observed only for employed individuals and the share of employed immigrants is not only relatively low in Spain but also unequal across regions of origin and levels of Spanish, we control for sample selection bias and dissociate employment and occupational prestige amongst employed immigrants. Second, we also take into account potential endogeneity problems which may arise for different reasons and downward bias returns on language acquisition. As a result, our findings nuance previous evidence on the topic while confirming that returns on an immigrant's level of language skills in Spain follow a pattern that is different from those in other large destination countries.

¹ Facultad de CC Económicas y Empresariales, Universidad de Castilla-La Mancha, Plaza de la Universidad s/n (02071) Albacete, Spain. MAngeles.Davia@uclm.es.

² University International College, Macau University of Science and Technology, Avenida Wailong, Taipa, Macau, China. twang@must.edu.mo.

³ Facultad de CC Económicas y Empresariales, Universidad de Castilla-La Mancha, Plaza de la Universidad s/n (02071) Albacete, Spain. matias.gamez@uclm.es.

* Correspondence author.

Keywords: language skills, endogeneity, immigrants' employability, occupational prestige

Dominio del español y logros laborales de los inmigrantes en España

Resumen: Este trabajo estima los rendimientos al dominio del español en inmigrantes de primera generación en términos de empleabilidad y prestigio ocupacional. La segregación ocupacional en el mercado de trabajo español entre trabajadores de distintas regiones del mundo aconseja la distinción entre dominio del español en general y su nivel de adquisición en inmigrantes de otras lenguas maternas. Nuestra estrategia empírica nos permite considerar dos cuestiones que pueden causar una infraestimación del rendimiento del dominio de la lengua en el país de destino. En primer lugar, el prestigio ocupacional es observable solo en personas ocupadas. Ante la relativamente baja tasa de empleo en España y su desigual distribución por regiones de origen y niveles de conocimiento del español, controlamos por el posible sesgo de selección muestral, disociando el rendimiento en empleo en sí y prestigio ocupacional en inmigrantes con empleo. En segundo lugar, tenemos en cuenta el riesgo de endogeneidad de que, por diversas razones, puede conducir a una infraestimación del rendimiento del dominio de la lengua en el país de destino. De este modo, nuestros resultados ofrecen algunos matices a la evidencia existente, confirmando además que los rendimientos al dominio por los inmigrantes de la lengua de destino siguen un patrón diferente al de otros grandes países receptores.

Palabras Clave: Dominio de la lengua, endogeneidad, empleabilidad de los inmigrantes, prestigio ocupacional.

1. Introduction

Language skills, i.e., the level of understanding, speaking, reading and writing in the host country language, represent a very important type of human capital for international migrant workers. Their determinants and returns have been widely explored in many developed immigrant-receiving countries such as the United States, the United Kingdom (UK), Germany, the Netherlands and Canada, among others. In Spain, the evidence is scarcer than in other receiving countries because Spain has become an important immigration country only in recent decades.

This evidence contributes to the literature on determinants of and returns on language proficiency among immigrants in Spain (see a survey on this literature in Budría & Martínez de Ibarreta, 2021a). We think it is relevant because of the global importance of Spanish as a spoken language (Chiswick & Miller, 2015) and because of the peculiar linguistic profile of immigrants in Spain, nearly half of whom come from developing Spanish-speaking countries.

Immigrants from Latin America share ethnocultural and linguistic affinities with Spain because of colonial links. Initially, this should lead to better integration into the host society and the labour market. However, evidence from other traditional receiving countries with former colonies like the UK, France and Portugal shows different patterns in the impact of linguistic background on labour market outcomes for immigrants from former colonies (Bouju & Edel, 2018). In the UK, language proficiency helps to improve employment outcomes (Dustmann & Fabbri, 2003), particularly for immigrants from so-called Old Commonwealth countries as designated by Lessard-Phillips (2018) namely, Canada, Australia, New Zealand and South Africa: —which suggests a pattern based on nationality (Demireva, 2011; Dustmann & Fabbri, 2005). African French-speaking immigrants in France find it difficult to exploit social and cultural capital because of ethnic and racial discrimination (Thomas, 2013). Brazilian immigrants in Portugal are found in service industries like sales, restaurants or other leisure-related sectors and are quickly stereotyped (Sardinha, 2018). If linguistic background made Latin American immigrants a privileged group in the Spanish labour market, the returns on Spanish proficiency would definitely be high, but this group is in fact segregated into low-productivity sectors⁴ (Rodríguez-Planas & Nollenberger, 2016). Given that they account

⁴In Spain, EU-15 immigrants are employed in the same sectors as natives, but other immigrant workers suffer from high levels of segregation. Latin American immigrants, immigrants from other European countries and African immigrants have traditionally been concentrated in construction, food preparation and serving and domestic services (Gonzalez & Ortega, 2011; Sanromá et al., 2009).

for a large proportion of the immigrant population, the average returns on Spanish proficiency seem low.

The above-mentioned occupational and sector segregation based on nationality affects returns on Spanish proficiency and makes it necessary to distinguish between Spanish proficiency in general and the acquisition of Spanish in particular. Spanish proficiency refers to both immigrants whose mother tongue is Spanish and those who have learnt Spanish in their country of origin or upon arriving in Spain. Spanish acquisition is exclusive to those whose mother tongue was not Spanish and needed to invest in language human capital. In this paper, we are particularly interested in the returns on that investment.

Most of the evidence about returns on immigrants' language skills, in both Spain and other host countries, is measured exclusively with wages. We offer here a complementary approach based on a variable that is correlated with wages but comprises the nuance of prestige, namely, occupational attainment. We measure it by means of the *New International Socio-Economic Index* [ISEI], constructed by Ganzeboom (2010) on the basis of the *International Standard Classification of Occupation 2008* [ISCO-08].

The high levels of immigrants' labour market participation and employment rates in most of the main host countries (the US, the UK, Germany, etc.) may explain why analyses of labour market outcomes often ignore the fact that wages (among other outcomes) are observable only amongst employed individuals. However, in the Spanish labour market, this nuance should not be left out: given the prevailing employment rates for both immigrants and natives, being in paid employment is already a relevant outcome in itself that might well depend on language skills. Our empirical strategy will take this into account by controlling for potential sample selection.

The analysis of returns on language acquisition can also suffer from endogeneity if language acquisition depends on non-observable features—as is often the case—that are positively correlated with labour market outcomes. More able individuals may be better at language learning, and this ability may explain their labour market outcomes rather than their language skills. Under this scenario, significant and positive statistical links observed between language acquisition and the adopted measure of success in the labour market would be (upward) biased (see Budría & Martínez de Ibarreta (2021b)). Our multivariate strategy (simultaneous estimation of both language acquisition and labour market outcomes) will address these challenges.

So far, the evidence on these issues in Spain is widely based on the NIS-2007 (National Immigration Survey), a cross-sectional one-time survey conducted just before the outbreak of the economic crisis in 2008. Evidence with that dataset showed that knowledge of Spanish contributed to immigrants' access to employment (Gutiérrez et al., 2010; Isphording, 2013; Swedberg, 2010) and their level of income, especially for non-Hispanic immigrants (Budría & Swedberg, 2015; Swedberg, 2010). In this paper, we explore the 2014 Spanish Labour Force Survey (LFS) ad hoc module, "Labour market situation of migrants and their immediate descendants".

This dataset portrays the aftermath of the Great Recession after peak unemployment rates in 2013. This might imply different returns on language skills than the ones observed in the National Immigration Survey (NIS) 2007. In an economic downturn, employers are expected to retain their most qualified (and least replaceable) workers, job queues will lengthen and employers will be able to choose a selection of the most qualified applicants. The resulting competition for jobs should increase returns on all sorts of human capital amongst immigrants, language skills among them. Accordingly, a good level of Spanish should be more relevant for immigrants' employability than during an economic boom. This is one of the scenarios tested in Cebolla-Boado et al. (2015) as regards the evolution of returns on education to reduce unemployment risks in both native and immigrant adults in Spain. However, the crisis increased the employability gap between immigrants and natives and returns on education fell amongst immigrants as compared with natives during the economic downturn. The relative disadvantage became more pronounced for both Africans and Latin Americans than for Eastern Europeans. In this sense, instead of increasing, returns on language skills might indeed have fallen during the recession: the more fluent in Spanish immigrants were, the more they were likely to compete with natives in certain occupations.

To sum up, this paper contributes to existing evidence on returns on language skills in Spain in several ways. First, it addresses two relevant non-monetary labour outcomes: access to employment and occupational attainment or prestige. Second, it analyses these outcomes in the aftermath of the Great Recession. Third, it deals with two problems that may bias previous evidence, namely, sample selection and endogeneity.

2. Spain as a receiving country of migrants

Spain has become an important immigration country only in the past three decades. Traditionally an emigrant country since the colonial era, by the mid-20th century, Spain

was a labour-exporting country, with France, Germany, Switzerland and the United Kingdom the main destinations (Cornelius, 2004). As of the late 1970s, socio-political changes and the oil crisis in developed countries caused the reduction of emigration from Spain and the increase of migration inflows. At first, most of the immigrants were Spaniards returning to their native country, but as of the 1980s, the inflow of foreign residents in Spain became more and more intensive. The number of foreign-born residents peaked in 2007: 5,249,993 people (estimates for January 1st), i.e., 11.61% of the total resident population in Spain (Fernández Páez, 2012).

The severe depression from 2008 to 2013 reduced labour market opportunities for immigrants, with foreign residents more vulnerable than Spanish natives (Gil-Alonso & Vidal-Coso, 2015). As a result, many immigrants returned to their home countries or searched for alternative destinations and outflows exceeded inflows (Domingo & Blanes, 2015; Romero-Valiente, 2003) up to 2015. By January 1st, 2014, the size of the immigrant population in Spain had fallen to 4,677,059, i.e., 10% of the full resident population, and continued to shrink throughout 2014 (Callejo & Fuentes, 2015).

The economic recession affected immigrants in different ways and employment rates varied widely by regions of origin: EU-15 immigrants maintained a very stable employment rate, while immigrants from other European countries and Latin Americans suffered a notable fall. The unemployment rate of Latin American immigrants skyrocketed from less than 10% in 2007 to 30-40% during the period 2008-2012, similar to immigrants from Eastern European countries and far worse than immigrants from the EU-15. African immigrants registered the lowest employment rate before, during and even after the Great Recession (Mooi-Reci & Muñoz-Comet, 2016; Rodríguez-Planas & Nollenberger, 2016).

Since 2014, the economic cycle has changed substantially in Spain with the post-2014 recovery and the recession caused by the lockdown for COVID-19. Unfortunately, at the moment of writing this document, there are not (to our knowledge) more recent datasets with equivalent information about immigrants' language skills.

3. Determinants and returns on proficiency in the host language

3.1. Determinants: a mix of colonial ties and the three E's

As explained in the Introduction, a large share of the immigrants in Spain come from former colonies, mostly from Latin America, and therefore are native Spanish speakers⁵. From now on, we will refer to them as Hispanic. We will take them into account at a first stage of our analysis, where we will describe returns on Spanish proficiency for all immigrants, and this group will be singled out because of its members' full command of Spanish. Other immigrants learn Spanish either at school in their home countries or upon arrival in Spain. For them, language acquisition is an investment in an important type of human capital. In the second stage of our analysis, both the determinants of the effective level of Spanish acquisition and its returns will be estimated.

The degree of mastery of the host country language—when different from the immigrants' mother tongue—usually results from several factors, widely documented in the economic literature and clustered in the so-called *three E's* (Chiswick, 2009): exposure to the language of the destination country, efficiency in learning and economic incentives of immigrants to acquire knowledge of the local language.

Exposure to the host language is usually measured in years of residence in the destination country, marital status and co-residence with parents or children and marriage upon arrival, especially to a native spouse (Chiswick et al., 2005; Chiswick & Miller, 2015).

Efficiency in learning the local language is widely shown to decline with age at the moment 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. Consistently, international evidence on immigrants' language skills finds that early arrivers are more fluent in the destination language (Budría et al., 2017; Chiswick & Miller, 2002; Miranda & Zhu, 2013; Yao & Ours, 2015; Zorlu & Hartog, 2018). Efficiency 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). Finally, the linguistic distance between the

⁵This is not always the case, as some people in those countries belong to minority ethnic groups where Spanish is not spoken at home but acquired at school. Those people are not identifiable in our dataset, and we expect them to be less prone to migrate to Spain than native Spanish speakers.

native and host languages also influences efficiency in the acquisition of the host language (Isphording & Otten, 2011).

The main economic incentives for learning the host country language are immigrants' employment status and earnings and the extent to which they respond to proficiency in the host language. They are expected to increase immigrants' willingness to stay in the destination country, which explains the inclination of immigrants to invest in learning the host language and other types of human capital specific to the host country as well (Zorlu & Hartog, 2018).

Most of the evidence on the determinants of language acquisition amongst immigrants with no native Spanish-speaking background stems from the NIS-2007 (see Budría et al., 2019, 2021a; Budría & Swedberg, 2015; Gutiérrez et al., 2010; Swedberg, 2010) and corroborates the main findings in other destination languages. Namely, those who married a native Spanish speaker or live with school-aged children at home (Exposure) proved to be more proficient in Spanish. Similarly, those who arrived at an earlier age, those who have a higher educational level and those whose mother tongue is less distant from Spanish (Efficiency) also had a greater command of Spanish. Finally, employed immigrants (Economic incentives) also possessed better levels of Spanish.

3.2. Returns on proficiency in the host language: beyond monetary outcomes

Much of the international evidence on returns on proficiency in the host language assesses its impact on immigrants' earnings. There is a large body of evidence for this in the main receiving countries such as the UK (Dustmann & Fabbri, 2003; Miranda & Zhu, 2013), the United States (Bleakley & Chin, 2004; Chiswick & Miller, 2002), Canada (Chiswick & Miller, 1988), Australia (Chiswick et al., 2005), Germany (Dustmann & van Soest, 2002), and the Netherlands (Bloemen, 2013). In the survey conducted in Chiswick & Miller (2015), the authors concluded that a good command of the host language contributes to increased earnings ranging from 5% to 35% amongst male immigrants.

Wages are a good proxy for job quality and labour market integration and are particularly versatile dependent variables for statistical multivariate analyses of returns on language proficiency and language acquisition. Still, they may suffer from two shortcomings: first, the level of non-negligible underreporting in surveys; and second, the inconsistency of wages in capturing socioeconomic status: some mid-qualified jobs may offer higher wages than others because of the relative scarcity of a skilled labour force or to compensate for hard working conditions. Non-monetary outcomes, such as

occupational attainment, may show more nuanced returns than wages, allowing the explicit observation of job quality or subjective wellbeing⁶. They are also characterized by lower shares with data that is missing or misreported.

The evidence for non-monetary returns on language skills in labour market outcomes is sparser than the evidence for wages and covers different dependent variables, from employment to job satisfaction and overqualification risk reduction. Fluency in the destination language is a relevant determinant of employment probabilities (Dustmann & Fabbri (2003) in the UK, Aldashev et al. (2009) and Yao & Ours (2015) in Germany). Language proficiency also proves to be relevant for overqualification risks (Greene et al., 2006 and reducing overqualification or educational mismatch rates in both pre- and post-immigration jobs (Budría et al., 2021b; Imai, Stacey, & Warman, 2019;)). Language proficiency is also positively related to satisfaction with the type of job and professional status (Bloemen, 2013).

As for other host countries, empirical evidence for Spain shows that immigrants' language skills increase their employability (Budría et al., 2017; Budría et al., 2019; Miyar-Busto et al., 2019). It also contributes to job security by garnering more permanent contracts (Gutierrez et al., 2010; Martínez-Pastor, 2014). Still, its impact on income is limited (Isphording, 2015; Swedberg, 2010). The author explains such apparently counterintuitive results by the inclusion of Hispanic immigrants in the analysis. Segmentation and occupational segregation in the Spanish labour market drive many native Spanish-speaking immigrants from Latin American countries to low- and middle-skill occupations, some of which require a high level of Spanish. However, proficiency in Spanish does not guarantee access to white-collar jobs (Budría and Ibarreta, 2021a) and can even diminish the possibility of landing one (Budría et al., 2019). White-collar jobs require high technical training which demands abilities other than language skills. Immigrants from OECD countries exemplify this point, occupying highly qualified, well-paid positions though they do not always have a good command of Spanish.

Hispanic immigrants' fluency in Spanish is not subject to the three *E*'s, which in their case may not explain their level of proficiency. Moreover, their presence in the analysis may bias returns on language skills because of unobserved heterogeneity; they speak the host language and are also culturally very close to native Spanish citizens. What is more,

⁶Moreover, wages are not available in the LFS, so it is not possible to test whether our results are applicable to monetary and non-monetary outcomes.

because of the segmentation in the Spanish labour market, immigrants from Latin America tend to concentrate in middle- and low-qualified occupations that require language skills but not formal education. Therefore, in our multivariate analysis, we will distinguish between Hispanic and non-Hispanic immigrants and will devote particular attention to the latter group.

4. Data: The 2014 LFS ad hoc module

In order to identify efficiency in the acquisition of Spanish. Linguistic distance stems from the Automatic Similarity Judgement Program (ASJP), developed in a research project from the German Max Plank Institute for Evolutionary Anthropology and described in Brown et al. (2008) and Isphording and Otten (2013). This powerful matrix of linguistic distances has already been widely used in international evidence (Adsera & Pytliková, 2015; Bloemen, 2013; Isphording & Otten, 2011, 2014, among others).

As regards occupational status, we link the two-digit occupation codes with their corresponding values in the *New International Socio-Economic Index* [ISEI]. First created in Ganzeboom, De Graaf and Treiman (1992), the ISEI-2010 covers 42 countries. It is used to measure socio-economic job status and is highly correlated with earnings and job qualification requirements, contributing to our understanding of the relationship between education and income from a whole array of occupations. The ISEI values of our sample range from 14 (assigned to agricultural assistants and cleaners, for example) to 69 (for members of executive teams and directors). We standardize it to ease interpretation of the results.

Finally, in order to take into account the local labour market conditions when predicting the immigrant's probability of being in paid employment, we "plug" regions⁷ and quarter- specific unemployment rates in 2014 (obtained from the whole Spanish LFS sample) into each observation.

Our analysis has two steps: it identifies first the probability of being employed and then the occupational status for those in paid employment. This entails using one sample of first-generation immigrants (5,787 observations with no missing information in relevant explanatory variables; 3,530 are non-Hispanic immigrants) and another of employed first-generation immigrants who report valid information on the predictors of

⁷ Regions are identified here with autonomous communities, NUTs 2 in Eurostat terminology.

occupational status (3,036 observations, 1,782 of which correspond to non-Hispanic immigrants). We keep second generation immigrants latter aside from the analysis because their assimilation process is expected to differ⁸ from first generation migrants.

In the LFS-2014, proficiency in Spanish is self-reported. All immigrant interviewees were asked to self-assess their level⁹. The wording of the question reads as follows: My level of spoken Spanish is

- 1. Spanish is my mother tongue or I speak it like a native*
- 2. Advanced level*
- 3. Intermediate level*
- 4. Beginner or without skills*
- 0. Do not know¹⁰.*

We split category 1 into two groups: (1a) Hispanic (Spanish is my mother tongue, for immigrants born in Spanish-speaking countries or whose parents were born in Spanish-speaking countries), which accounts for 40% of first-generation immigrants and (1b) non-Hispanic who speak Spanish like natives do (17% of all first-generation immigrants and nearly 29% of non-Hispanic immigrants).

Other features of the different samples (also displayed in Table A.1) are as follows: one third came from other European countries (either EU-15, other EU-28 or non-EU European nations) and 15.5% were born in Maghreb countries. A bit more than 50% were women and about one fourth had higher education. Around 39.5% of the migration was work-related.

Within the subsample of non-Hispanic immigrants, nearly one third did not live with a spouse or partner, and about 40% did not live with children of their own, probably because of their relatively young age (on average, 40 years old). In the sub-sample of employed immigrants (regardless of their mother tongue), the average time spent as a

⁸ Second generation migrants were born in Spain from at least one non Spanish-born parent and are expected to be either monolingual in Spanish or bilingual in Spanish and their parents' mother tongue (Medvedeva & Portes, 2017) as well as to receive all their schooling in the Spanish education system. This entails learning both Spanish as well as other co-official languages in the region of residence.

⁹ Unfortunately, the question is formulated differently from previous studies, which hinders comparability of the results. Self-assessed language skills usually rank from "very well" to "not at all". In the NIS-2007, the variable that indicates language skills holds values ranging from "very well" to "needs improving" (Budría & Swedberg 2015, Budría, Ibarreta & Swedberg, 2017 and Isphording 2014). Other research stresses difficulties with the host language. Yao & van Ours (2015) address difficulties with Dutch, while Miranda and Zhu (2013) use deficiencies in English to proxy poor language skills.

¹⁰ This category is so small (0.47 %) that we omit it in the analysis.

resident in Spain¹¹ was 17 years, and the average tenure in the present job position was six years. Nearly 40% worked in very small firms, about one fourth worked part-time and 16% were self-employed.

An initial description of returns on language skills is shown in Table 1. It displays employment rates and the average values of the ISEI index amongst those in paid employment according to the level of Spanish proficiency and country of origin.

Table 1 - Language skills, country of origin and employment outcomes

	Employment rate	ISEI index
Proficiency in Spanish		
Hispanic	56.3	31.2
Non-hispanic who speaks Spanish like a native	55.1	41.9
Advanced level	53.0	34.9
Intermediate level	50.5	27.2
Beginner or no skills	35.4	29.7
Country of origin		
UE-15/EFTA	60.6	44.6
New EU-28 member states	55.1	28.5
Other European countries	47.0	32.1
Latin American countries	56.2	31.1
Maghreb	31.2	28.5
Other African countries	37.8	29.2
Asia	69.6	29.7
North America and Oceania	68.8	54.8
Total	51.2	33.7
Number of observations	5,729	3,004

Source: 2014 Spanish LFS ad hoc module.

Employment rates increase with the level of Spanish: they are 20 percentage points higher for those who report a native level than for “beginners”. Occupational status does not clearly respond to the level of Spanish, though: non-native Spanish speakers with a good or very good command of Spanish report higher ISEI levels than native Spanish speakers. The distribution of ISEI seems to be related to migrants’ level of development

¹¹ This variable was computed as the difference between age at the interview and age on arrival.

in the country of origin, with those from the EU-15, North America and Oceania reporting the highest levels.

According to the aforementioned trends we would expect to find clear returns on fluency in Spanish in terms of employment rates but not for occupational status. This result would be in line with part of the literature (such as Budría et al., 2018; Isphording, 2015; Miyar et al., 2019; Swedberg, 2010) and consistent with the segmentation of the Spanish labour market across nationalities, occupations and activities explained in Section 3.2.

5. Empirical strategy

5.1 Endogeneity and sample selection in the analysis of returns on language skills

The information displayed in Table 1 shows that higher levels of Spanish are characterized by a higher likelihood of being employed, while the association with occupational prestige is not clear. To properly measure returns on proficiency in Spanish, we adopt a multivariate regression analysis framework where labour market outcomes (in our case, being in paid employment and occupational prestige (ISEI index)) are explained by host country language skills in the presence of a large set of additional variables. The coefficients (parameter estimates) of the variables that denote language skills would measure returns on fluency in Spanish.

The problem is that the estimated returns on host country language skills may be biased because of endogeneity problems that may arise for several reasons (Yao & Ours, 2015): first, unobserved features that affect both labour market performance and language skills (such as motivation and innate ability); second, reverse causality, i.e., labour market performance reversely contributes to proficiency in the host language; third, potential measurement errors in self-reported language skills. Unobserved heterogeneity and reverse causality would lead to an upward bias in the parameter estimate of returns on language skills, while measurement errors would instead lead to a downward bias. Moreover, because occupational prestige is observed only for employed immigrants, who are not a random sub-set of the population, its analysis would be additionally biased if some unobserved features affected both immigrants' employment prospects and occupational prestige. We address the two sources of bias (endogeneity and sample selection) with our multivariate strategy. In doing so, we expect to achieve an unbiased estimate of returns on language skills amongst first-generation immigrants.

There are two strategies for addressing endogeneity in the estimation of returns on language proficiency: (a) substituting language skills with instrumental variables¹² (IV)—observed variables that affect language acquisition but are not related to unobservable ability and labour market outcomes (Isphording, 2015); (b) bivariate analysis, where labour market outcomes are estimated at the same time as language proficiency, allowing for potential correlation between non-observables that affect labour market outcomes and the likelihood of being proficient in Spanish.

The most common IVs in the literature are linguistic distance, age on arrival (Isphording, 2015) and the interaction of these two IVs (as in Isphording & Otten (2014)). They prove particularly useful when the dependent variable is continuous (such as wages), but valid instruments are often difficult to find and do not work equally well in all datasets. The strategy we follow here—a simultaneous equations approach—has two advantages over IVs: it allows us to work simultaneously with discrete (the probability of being in paid employment) and continuous (ISEI index) dependent variables and to correct for two sources of bias (endogeneity and sample selection) at the same time.

5.2 Multivariate empirical strategy

The strategy adopted here has two steps. First, we consider language skills (L) to be exogenous to labour market outcomes ($ISEI$ and E) in a two-equation system where we estimate occupational prestige ($ISEI$) in all first-generation immigrants, conditional on employment (E) to address only potential sample selection. In a second step, we deal with potential endogeneity in language skills as well. This estimation covers only non-Hispanic immigrants, i.e., those who acquired Spanish in a process featured by the three E 's described in Section 3.1. This process will be tested in the third equation of the system.

The first two-equation system may be described as follows:

$$\begin{aligned}
 ISEI_i &= \gamma_1' L_i + \beta_1' X_{1i} + \varepsilon_{1i}; & ISEI_i \text{ is observed if } E_i = 1 & \quad (1) \\
 E_i^* &= \gamma_2' L_i + \beta_2' X_{2i} + \varepsilon_{2i}; & E_i = 1 \text{ if } E_i^* \geq 0 \text{ and } 0 \text{ otherwise} &
 \end{aligned}$$

where E_i^* is the latent propensity to work and $ISEI$ is observed only when E_i takes the value 1, i.e., if the individual is employed at the time of the interview. L_i is a categorical variable

¹²Examples of these studies are Chiswick and Miller (1995, 1998); Dustmann and van Soest (2002); Bleakley and Chin (2004, 2010); Miranda and Zhu (2013) and Budría and Swedberg (2015).

that captures Spanish proficiency, contributing to predicting both ISEI and employment. It takes five values (from “beginner/no skills” to “speaks it like a native” and “Hispanic”). The sets of parameters γ_1 and γ_2 represent the effect of language proficiency (i.e., its returns) on occupational prestige (ISEI levels) and the likelihood of employment, respectively.

The two sets of additional covariates (X_1, X_2) are intended to detect heterogeneity across immigrants and their jobs, contributing to predicting their occupational status and their proneness to work. Both X_1 and X_2 include control for region of origin fixed effects, so γ_1 and γ_2 capture only language skills, net from cultural or institutional peculiarities that influence the speed in labour market insertion. Additional control variables in X_1 are personal features (gender, age when education finished, number of years of residence in Spain) and job features (full- or part-time job, tenure in years, professional status and type of contract). Finally, X_2 includes personal and family-related characteristics (gender, number of children and their interaction, level of education and main reason for migrating) and a region-specific unemployment rate (see Budría & Ibarreta (2021b) for a discussion about the convenience of including context variables in the analysis).

Errors ($\varepsilon_1, \varepsilon_2$) are clustered across home country codes. They are normally distributed with mean 0, variances σ_1 and σ_2 and covariance σ_{12} . Should they be mutually correlated because of unobservable factors that determine both labour market outcomes, the only off-diagonal element of the variance-covariance matrix (Σ), ρ_{12} , would be significant. Additionally, the equation system also allows us to estimate lambda (the coefficient of the inverse Mills ratio (IMR) which, if statistically significant, proves the presence of sample selection bias and the adequacy of the deployed strategy).

As explained above, the estimates of returns on language skills (γ_1 and γ_2) could be downward-biased because of endogeneity due to reverse causality, unobserved heterogeneity or measurement errors. This is why, in the second step, we work on a subsample of non-Hispanic immigrants whose proneness to work, occupational status (ISEI) and level of Spanish (this time, from “beginner/no skills” to “speaks like a native”) we can estimate. The equation system may be expressed as follows:

$$\begin{aligned}
 ISEI_i &= \gamma_1' L_i + \beta_1' X_{1i} + \varepsilon_1; & ISEI_i \text{ is observed if } E_i &= 1 & (2) \\
 E_i^* &= \gamma_2' L_i + \beta_2' X_{2i} + \varepsilon_2; & E_i &= 1 \text{ if } E_i^* \geq 0 \text{ and } 0 \text{ otherwise} \\
 L_i^* &= \beta_3' X_3 + \varepsilon_3;
 \end{aligned}$$

Where X_3 is the vector of explanatory variables that contribute to predicting language skills. They are related to the three *Es* explained in Section 3.1: exposure to the host language, efficiency in skills acquisitions and economic incentives.

Exposure is captured by variables that reflect interviewees' co-residence with a spouse (whether they are Spanish born, a native Spanish-speaking immigrant or a non-native Spanish-speaking immigrant) or their own children (and whether they are school-age). Efficiency is measured through (a) education attainment, also potentially correlated with pre-migration exposure to Spanish; (b) age on arrival and (c) linguistic distance. The latter two are expected to be negatively correlated with language skills, but part of the effect of age on arrival and linguistic distance may be related to non-language factors and correlated with the measurement error in Spanish proficiency. For instance, arriving early in the destination country increases the likelihood of studying in the host country and, therefore, full recognition of the immigrant's qualification. Similarly, linguistic distance may reflect cultural similarities which will ease the integration of the immigrant upon arrival in the host country. That is why we control for a dichotomous variable that takes value 1 if the immigrant comes from an EU-15 or EFTA (European Free Trade Area) country, North America or Oceania. The aim is to identify countries which are at least as developed as Spain. In such countries, the official language might be relatively distant from Spanish, but immigrants from those countries may be culturally closer to Spanish-born citizens than immigrants from other countries where the official language is not as distant from Spanish. For their part, economic incentives are captured through the reason for migrating.

β_3 is the vector of parameters to be estimated in the third equation of the system. Since language skills are measured with a Likert scale, they will be estimated via an ordered probit model (see Greene, 2018, for further details).

Finally, ε_1 , ε_2 and ε_3 , the error terms, are assumed to follow a joint normal distribution and are clustered across home country codes. Correlations across error terms would be observed if the off-diagonal elements of the variance-covariance matrix (Σ) (ρ_{12} , ρ_{13} and ρ_{23}) were statistically different from zero. In this case, the coefficients from separate equations would be biased (ρ_{12} would pinpoint sample selection bias, while ρ_{13} and ρ_{23} would identify endogeneity bias), and the simultaneous equations estimates would be preferred to estimating the three equations separately.

In order to estimate the models explained above, we deploy the *Stata conditional mixed process estimator* (cmp) command, developed in Roodman (2011).

6. Results

In Table 2, the coefficients (expressed as marginal effects) for three different specifications are displayed. Models I and II consist of a linear regression for occupational prestige (*ISEI*) and a binary probit model that estimates the likelihood of being in employment at the time of the interview (*E*). Model I (first two columns) is estimated for the whole sample of first-generation immigrants to single out the coefficients for the Hispanic immigrants' category. Model II (next two columns) applies only to non-Hispanic immigrants to obtain the relevant coefficients for language acquisition before controlling for endogeneity. In Model III (last three columns), a third equation is added to predict non-Hispanic immigrants' level of Spanish in order to correct for potential endogeneity in the measurement of returns on command of Spanish.

At the bottom of the table, the lambda coefficient is shown to be significantly different from zero in all specifications, and the correlation across errors (ρ_{13} and ρ_{23}) is significantly different from zero in Model III, which supports the choice of the multi-equation strategy to address both sample selection and endogeneity.

6.1 Returns on language skills

A command of Spanish implies a higher likelihood of being in employment, even in the presence of non-language effects linked to language proficiency. Still, and in line with previous evidence and the descriptive analysis displayed in Table 1, it does not seem to influence ISEI scores in either the overall sample (all first-generation immigrants) or the sub-sample of non-Hispanic immigrants. Nevertheless, when the potential bias in language skills due to endogeneity is corrected for, returns on language skills become significant and positive for non-Hispanic immigrants. This would confirm the underestimation (downward bias) of the impact of language skills that is usually revealed in the literature (see Isphording (2015), among others). That bias is likely due to unobservable ability that affects labour outcomes of immigrants. It affects language returns also in terms of employability. In Model III, these returns are shown to be much larger. At the same time, when endogeneity is corrected for in Model III, differences in the ISEI levels by regions of origin become less significant.

6.2 Other determinants of labour market outcomes

With regard to the two sets of control variables¹³ (X_1 and X_2), women do not face lower occupational status than men; age and education attainment increase the likelihood of being in paid employment. In the presence of control for macroeconomic conditions (the prevalent unemployment rate in the region of residence) and human capital (language skills and education attainment), immigrants from North Africa (Maghreb) still register the lowest employability levels; only Asian workers are more likely to work than immigrants from EU-15 countries.

The ISEI is positively correlated with education attainment (proxied by the age when the interviewee left/finished education) and with tenure in the current position but, interestingly, not with time spent as a resident in Spain. Its connection with the region of origin becomes nearly non-significant in Model III.

¹³ The coefficients for other covariates are not shown for reasons of space but are available to the interested reader upon request. They show that job status is higher for self-employed workers and lower amongst part-timers, *ceteris paribus*. Also, migrants are more likely to be found in higher-ranked positions when they work in firms or work centers larger than 10 employees.

Table 2: Results of multivariate simultaneous models (marginal effects)

	Model I (All immigrants)		Model II (Non-Hispanic)		Model III (Non-Hispanic)		
	ISEI	Empl.	ISEI	Empl.	ISEI	Empl.	Spanish*
Spanish proficiency (ref: beginner/none)							
Intermediate level	-0.020	0.109	-0.019	0.104	0.072	0.175	
Advanced level	0.040	0.105	0.051	0.098	0.220	0.235	
Speaks like a native	-0.001	0.159	0.045	0.147	0.312	0.346	
Hispanic immigrant	0.005	<i>0.114</i>	-	-	-	-	
Female (ref: male)	-0.050	-0.013	-0.026	-0.037	-0.031	-0.034	
Years of residence in Spain	0.002		-0.001		-0.003		
Age when education ended	0.014		0.013		0.011		
Education (ref: no schooling)							
Primary education		0.042		<i>0.064</i>		0.031	0.046
Lower secondary		0.035		0.058		0.006	0.097
Upper secondary		<i>0.076</i>		0.090		0.018	0.168
Higher education		0.221		0.239		0.155	0.213
Reasons for migrating (ref: work-related)							
Family-related migration		-0.085		-0.067		-0.089	0.007
Education-related migration		0.027		0.015		-0.016	0.028
Refugee or other motives		<i>-0.072</i>		-0.057		-0.055	-0.021
Regional unemployment rate		-0.481		-0.332		-0.336	
Age on arrival							-0.010
Linguistic distance (standardised)							-0.118
Cultural proximity: from EU-15 /EFTA, N. America, Oceania							0.223
Region of origin (ref: EU-15)							
Other EU-28 countries	-0.142	-0.015	-0.167	-0.010	-0.103	0.052	
Other European countries	<i>-0.093</i>	-0.085	-0.113	-0.078	-0.063	-0.012	
Latin American countries	-0.110	0.016	-0.124	0.021	<i>-0.079</i>	0.055	
Maghreb	-0.104	-0.166	-0.108	-0.168	-0.033	-0.113	
Other African countries	-0.103	<i>-0.100</i>	-0.116	<i>-0.097</i>	-0.064	-0.055	
Asia	-0.138	<i>0.159</i>	-0.151	<i>0.159</i>	-0.094	0.205	
North America and Oceania	0.156	0.020	0.138	0.089	0.131	0.096	
Coefficients printed in bold when $p < 0.01$, in italics when $p < 0.05$. * Marginal effects are estimated on the likelihood of the highest value of the dependent variable (i.e., speaking Spanish like a native).							
The full set of coefficients and their standard errors are not shown for reasons of space but are available upon request. Other control variables in vectors X_1 , X_2 and X_3 are explained in Section 4.2.							
Source: 2014 Spanish LFS ad hoc module.							

Table 2: Results of multivariate simultaneous models (marginal effects) (Cont.)

	Model I (All immigrants)		Model II (Non-Hispanic)		Model III (Non-Hispanic)		
	ISEI	Empl.	ISEI	Empl.	ISEI	Empl.	Spanish*
Spouse (ref: none)							
Spanish-born spouse							0.076
Native Hispanic foreign-born spouse							0.099
Not Hispanic foreign-born spouse							-0.099
Children (ref: none)							
Living with children not school-age							0.014
Living with school-age children							0.010
Chi2 ($p > 0$)	43,016 (0.000)		11886 (0.000)		63505 (0.000)		
-2 Log-likelihood	-3,221		-1,892		-5,073		
N clusters	102		96		96		
Number of obs	3,004	5729	1762	3493	1762	3493	3411
Lambda (s.e.)	-1.414 (0.079)		-1.433 (0.120)		-1.400 (0.091)		
	ρ_{12}		ρ_{12}		ρ_{12}	ρ_{13}	ρ_{23}
ρ_{jk} (correlations across errors)	<i>-0.581</i>		-0.632		<i>-0.529</i>	-0.381	-0.187
	(0.226)		(0.326)		(0.247)	(0.084)	(0.054)
Coefficients printed in bold when $p < 0.01$, in italics when $p < 0.05$. * Marginal effects are estimated on the likelihood of the highest value of the dependent variable (i.e., speaking Spanish like a native).							
The full set of coefficients and their standard errors are not shown for reasons of space but are available upon request. Other control variables in vectors X_1 , X_2 and X_3 are explained in Section 4.2.							
Source: 2014 Spanish LFS ad hoc module.							

6.3 Understanding migrants' proficiency in Spanish

Some of the determinants of language skills acquisition among non-native Spanish speakers are displayed¹⁴ in the last column in Table 2. They are intended to capture the three *E*'s in the relevant literature. First, variables that proxy *exposure*, such as having a Spanish-born or native Spanish-speaking spouse or partner, are related to higher levels of Spanish, and living with a non-Hispanic partner would be related to lower levels of Spanish than having no partner at all. Once this is controlled for, living with one's children does not seem to be correlated with fluency in Spanish.

As regards features related to learning *efficiency*, and in line with previous literature, educational attainment is positively related to the level of Spanish. Both linguistic distance between migrants' mother tongue and Spanish and age on arrival are negatively

¹⁴ The coefficients are expressed as marginal effects computed over the likelihood of category 4 ("speaking Spanish like a native"). That is, they express how a marginal change in one unit would affect the likelihood of having acquired the highest level of Spanish. For instance, non-Hispanic immigrants with primary education are five percentage points more likely to speak Spanish like a native than migrants with no schooling at all.

correlated with proficiency in Spanish. Finally, in the presence of the aforementioned variables the reasons for migrating (*economic incentives*) do not seem to be correlated with higher levels of Spanish.

7. Conclusions

In this paper, we study returns on both immigrants' language proficiency and language acquisition in Spain in the late phase of a severe economic crisis. The former includes Hispanic immigrants, who are native Spanish speakers, while the latter refers exclusively to those who learnt Spanish in their country of origin or, in most cases, once in Spain.

Immigrants' integration into the Spanish labour market is measured herein terms of the likelihood of employment and occupational attainment (measured via the ISEI index) for those who are in paid employment. This strategy allows us to nuance the estimation of occupational prestige with control for potential selection bias, as in a labour market with severe problems creating and, particularly, retaining employment, part of the return on language skills may reside merely in being employed.

The occupational segregation in the Spanish labour market across regions of origin pushes many Hispanic immigrants towards middle- or low-skilled occupations in personal services and hospitality with rather low levels of prestige where speaking the language is crucial. As a result, language skills seem to be unrelated to occupational prestige. This is why we turn our attention towards non-Hispanic immigrants who had to acquire the language and whose language skills may be complementary to other types of human capital. For these immigrants, we also consider potential endogeneity problems that may bias the measurement of returns on proficiency in Spanish in our empirical strategy. In line with previous evidence for pre-crisis periods, returns to language skills are clear only for non-Hispanic immigrants and, particularly, when the most appropriate tools are deployed to measure them.

The contribution of non-Hispanic immigrants' language skills to job quality challenges the idea that labour market segmentation and occupational segregation across nationalities in Spain inevitably assign immigrants to low-skilled jobs. Therefore, we may indeed conclude that the acquisition of proficiency in Spanish contributes to success in the labour market and should therefore be an important part of integration policies addressed to immigrants mainly from developing countries. Instead, for highly qualified immigrants from the most developed regions of the world, cultural similarities and pre-

migration exposure to the language contributes to overcoming the linguistic distance between their mother tongue and Spanish.

References

- Adsera, A., & Pytliková, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586), F49-F81.
- Aldashev, A., Gernandt, J., & Thomsen, S. L. (2009). Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection. *Labor Economics*, 16(3), 330-341.
- Bleakley, H., & Chin, A. (2004). Language skills and earnings: evidence from childhood immigrants. *Review of Economics & Statistics*, 86(2), 481-496.
- Bloemen, H. (2013). *Language Proficiency of Migrants: The relation with Job satisfaction and Matching*, IZA Discussion Paper No. 7366.
- Bouju, A., & Edel, A. (2018). *Similar but different: Inclusion and exclusion of immigrant communities sharing similar cultural background with their host societies*, Population Europe Discussion Paper No. 8.
- Brown, C. H., Holman, E. W., Wichmann, S., & Velupillai, V. (2008). Automated classification of the world' s languages: a description of the method and preliminary results. *STUF-Language Typology and Universals Sprachtypologie und Universalienforschung*, 61(4), 285-308.
- Budría, S., & Martínez de Ibarreta, C. (2021a). El español como elemento integrador en el mercado laboral. In J. L. García Delgado (Dir.) *El Español, Lengua Internacional: Proyección y Economía* (pp. 85-102). Civitas.
- Budría, S., & Martínez de Ibarreta, C. (2021b). Education and skill mismatches among immigrants: The impact of host language proficiency. *Economics of Education Review*, 84, 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, 625-652.
- Budría, S., Ibarreta, C. M. D., & Swedberg, P. (2017). The impact of host language proficiency across the immigrants' earning distribution in Spain. *IZA Journal of Development & Migration*, 7(1):12.

- Budría, S. & Swedberg, P. (2015). The impact of language proficiency on immigrants' earnings. *Revista de Economía Aplicada*, 23(67), 63-91.
- Callejo, M. B., & Fuentes, F. J. M. (2015). Precarización y vulnerabilidad de la población inmigrante en la España en crisis. *El papel del Estado de bienestar. Panorama Social*, 22, 139-151.
- Cebolla-Boado, H, Miyar-Busto, M. & J. Muñoz-Comet (2015). Is the Spanish Recession Increasing Inequality? Male Migrant-native Differences in Educational Returns Against Unemployment, *Journal of Ethnic and Migration Studies*, 41(5), 710-728.
- Chiswick, B. R. (1998). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel, *Journal of Population Economics*, 11(2), 253-271.
- Chiswick, B. R., & Miller, P. W. (1988). Earnings in Canada: the roles of immigrant generation, French ethnicity, and language. *Research in Population Economics*, 6, 183-228.
- Chiswick, B. R., & Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13(2), 246-288.
- Chiswick, B. R., & Miller, P. W. (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics*, 15(1), 31-57.
- Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2005). A longitudinal analysis of immigrant occupational mobility: a test of the immigrant assimilation hypothesis. *International Migration Review*, 39(2), 332-353.
- Chiswick, B.R. (2009). The economics of language for immigrants: An introduction and overview. In J. Sook Lee, R.W. Rumberger, & T.G. Wiley (Eds.), *The education of language minority immigrants in the United States* (pp. 72-91). Multilingual Matters.
- Chiswick, B. R., & Miller, P. W. (2015). International migration and the economics of language. In B. R. Chiswick, & P. W. Miller, (Eds.) *Handbook of the Economics of International Migration* (Vol. 1) (pp. 211-269). North-Holland.
- Cornelius, W. A. (2004). Spain: the uneasy transition from labor exporter to labor importer. In J. Hollifield, P. L. Martin, & P. Orrenius (Eds.). *Controlling immigration: A global perspective* 2nd ed. (pp. 387-429). Stanford University Press.
- Demireva, N. (2011). New Migrants in the UK: Employment Patterns and Occupational Attainment. *Journal of Ethnic and Migration Studies*, 37(4), 637-655.

- Domingo, A., & Blanes, A. (2015). Inmigración y emigración en España: estado de la cuestión y perspectivas de futuro. *Anuario CIDOB de la Inmigración*, 91-122.
- Dustmann, C. & Fabbri, F. (2005). Immigrants in the British Labour Market. *Fiscal Studies*, 26(4), 423-470.
- Dustmann, C. & Fabbri. (2003). Language Proficiency and Labor Market Performance of Immigrants in the UK, *The Economic Journal*, 489, 695-717.
- Dustmann, C. & Van Soest, A. (2002). Language and the Earnings of Immigrants, *Industrial & Labor Relations Review*, 55(3), 473-492.
- Fernández Páez, P. E. (2012). *Influencia de los flujos migratorios contemporáneos en España, tesis doctoral*. Universidad de Santiago de Compostela.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56.
- Ganzeboom, H. B. (2010). A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007. In *Annual Conference of International Social Survey Programme*, Lisbon (Vol. 1).
- Gil-Alonso, F., & Vidal-Coso, E. (2015). Inmigrantes Extranjeros en el mercado de trabajo español: ¿más resilientes o más vulnerables al impacto de la crisis?. *Migraciones*, 37, 97-123.
- Gonzalez, L., & Ortega, F. (2011). How do very open economies adjust to large immigration flows? Evidence from Spanish regions. *Labour Economics*, 18(1), 57-70.
- 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.
- Greene, W.H. (2018). *Econometric Analysis (8th ed.)*. Prentice Hall.
- Gutiérrez R., Mato, F.J. & Miyar M. (2010). Lengua e inmigración: conocimiento del español e integración de los inmigrantes, *Revista del Ministerio de Trabajo e Inmigración*, 85, 47-64.
- Imai, S., Stacey, D., & Warman, C. (2019). From engineer to taxi driver? Language proficiency and the occupational skills of immigrants. *Canadian Journal of Economics*, 52(3), 914-953.

- Isphording, I. E. (2013). Returns to Foreign Language Skills of Immigrants in Spain. *Labour*, 27(4), 443-461.
- Isphording, I. E. (2015). Language and labor market success. In J.D. Wright, (Ed. in Chief) *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed. (pp. 260-265). Elsevier.
- Isphording, I. E. & Otten S. (2011). *Linguistic Distance and the Language Fluency of Immigrants*, Ruhr Economic Papers No. 274.
- Isphording, I. E., & Otten, S. (2013). The Costs of Babylon—Linguistic Distance in Applied Economics. *Review of International Economics*, 21(2), 354-369.
- Isphording, I. E. & Otten, S. (2014). *Linguistic Barriers in the Destination Language Acquisition of Immigrants*, IZA Discussion Paper No. 8090.
- Lessard-Phillips, L. (2018). Beyond shared background? The Inclusion of Commonwealth Citizens in the UK in Education and the Labour Market. In A. Boujou, & E. Andreas (Eds.), *Similar but Different Inclusion and Exclusion of Immigrant Communities Sharing Similar Cultural Backgrounds with Their Host Societies* <https://population-europe.eu/file/3055/download?token=HkCtz-r6>
- Martínez-Pastor, J.I. (2014). Fixed-Term Contracts: Does Nationality Matter?, *Journal of Ethnic and Migration Studies*, 40(5), 814-828.
- Medvedeva, M., & Portes, A. (2017). Immigrant Bilingualism in Spain: An Asset or a Liability?. *International Migration Review*, 51(3), 632-666.
- Miranda, A. & Zhu, Y. (2013). English Deficiency and the Native-Immigrant Wage Gap, *Economics Letters*, 118(1), 38-41.
- Miyar-Busto, M., Mato, F. J., & Gutiérrez, R. (2019). Immigrants' educational credentials leading to employment outcomes: The role played by language skills. *Revista Internacional de Organizaciones*, 23, 167-191.
- Mooi-Reci, I., & Muñoz-Comet, J. (2016). The Great Recession and the immigrant–native gap in job loss in the Spanish labour market. *European Sociological Review*, 32(6), 730-751.
- Rodríguez-Planas, N. & Nollenberger, N. (2016). Labor market integration of new immigrants in Spain. *IZA Journal of Labor Policy* 5, 4.
- Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, 11(2), 159-206.

- Romero-Valiente, J.M. (2003). Migraciones, In Arroyo Pérez (Coord.): *Tendencias demográficas durante el siglo XX en España* (pp.209-253). Instituto Nacional de Estadística.
- Sanromá, Esteban & Ramos, Raul & Simón, Hipólito. (2009). Immigrant Wages in the Spanish Labour Market: Does the Origin of Human Capital Matter?. *SSRN Electronic Journal*. 10.2139/ssrn.1402504.
- Sardinha, J. (2018). Patterns of Integration, Cultural Negotiations and the Roles of Brazilian Associations in Portugal (Late-1980s to the Present) (pp. 25-29). *Similar but Different: Inclusion and Exclusion of Immigrant Communities Sharing Similar Cultural Backgrounds with Their Host Societies*.
- Swedberg, P. (2010). The impact of education and host language skills on the labor market outcomes of immigrants in Spain. *Investigaciones de Economía de la Educación*, 5, 798-824.
- Thomas, D. (2013). *Africa and France: Postcolonial Cultures, Migration, and Racism*. Indiana University Press.
- Yao, Y., & Ours, J. C. V. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labor Economics*, 34, 76-85.
- Zorlu, A., & Hartog, J. (2018). *The Impact of Language on Socioeconomic Integration of Immigrants*, IZA Discussion Paper No. 11485.

Table A.1. Description of variables deployed in the multivariate analysis

	Overall	Non-Hispanic	Overall (employed)	Non-Hispanic (employed)
Labour market outcomes	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
In employment	0.524	0.507		
ISEI (s.d.)			0.356(0.279)	0.377 (0.278)
Spanish proficiency				
Beginner or without skills	0.061	0.101	0.039	0.066
Intermediate level	0.144	0.237	0.134	0.228
Advanced level	0.230	0.376	0.224	0.381
Speaks Spanish like a native	0.174	0.286	0.190	0.325
Hispanic immigrant	0.390	-----	0.413	-----
Personal features				
Male	0.455	0.481	0.488	0.542
Female	0.545	0.519	0.512	0.458
Years of residence in Spain			16.77 (12.34)	18.89 (13.43)
Age at the time of the interview	39.29 (11.83)	40.00 (11.56)		
Age on arrival		22.56 (12.86)		
Education attainment				
Age when education ended			19.90 (6.19)	19.66 (6.23)
Illiterate	0.057	0.081		
Primary education	0.116	0.120		
Lower secondary	0.267	0.259		
Upper secondary	0.313	0.285		
Higher education	0.248	0.255		
Reason for migration				
Work-related	0.395	0.381		
Family reasons	0.507	0.528		
Education	0.023	0.023		
Refugee and other reasons	0.076	0.068		
Region unempl. rate (S.D.)	0.238 (0.057)	0.239 (0.056)		
No. of observations	5729	3493	3004	1762

Table A.1. Variables deployed in the multivariate analysis (Cont.)

	Overall	Non-Hispanic	Overall (employed)	Non-Hispanic (employed)
Linguistic distance (S.D.)		0.821 (0.210)		
Cultural distance: from EU-15 country or EFTA, etc.		0.297		
Region of origin				
EU-15 country or EFTA	0.178	0.292	0.207	0.354
New EU-28 member states	0.135	0.222	0.139	0.237
Other European countries	0.028	0.047	0.027	0.047
Latin American countries	0.434	0.078	0.459	0.085
Maghreb	0.154	0.253	0.090	0.154
Other African countries	0.024	0.040	0.018	0.031
Asia	0.039	0.064	0.050	0.086
North America and Oceania	0.007	0.005	0.009	0.007
Job features				
Tenure, in years (S.D.)			5.98 (6.75)	6.57 (7.28)
Firm > 10 employees			0.381	0.380
Part-time worker			0.240	0.220
Self-employed or employer			0.164	0.196
Employee - permanent			0.565	0.540
Employee - temporary			0.271	0.264
Household composition				
Number of children	1.036 (1.145)	1.103 (1.145)		
Does not live with a spouse		0.317		
Spanish-born spouse		0.271		
Native Hispanic foreign-born		0.039		
Non-Hispanic foreign-born		0.373		
Not living with children		0.413		
Not school-age children		0.257		
School-age children		0.330		
No. of observations	5,729	3,493	3,004	1,762

Source: 2014 Spanish LFS ad hoc module.