Language proficiency and immigrants' labor market outcomes in post-crisis Spain

Davia, María A. and Wang, Ting and Gámez, Matías

Universidad de Castilla-La Mancha, Macau University of Science and Technology, Universidad de Castilla-La Mancha

2 July 2019

Language proficiency and immigrants' labor market outcomes in post-crisis Spain

María A. Davia
Facultad de CC Económicas y Empresariales, Universidad de Castilla-La Mancha, Plaza de la Universidad s/n (02071) Albacete, Spain. (ORCID 0000-0001-9551-6111)

Wang Ting
University International College, Macau University of Science and Technology, Avenida Wailong, Taipa, Macau, China. (ORCID 0000-0002-9742-3928)

(Tel: +853 8897 1976, email: twang@must.edu.mo)

Matías Gámez
Facultad de CC Económicas y Empresariales, Universidad de Castilla-La Mancha, Plaza de la Universidad s/n (02071) Albacete, Spain. (ORCID 0000-0002-6772-9301)

Abstract

This paper analyses the impact of Spanish proficiency on first generation immigrants' labor market outcomes, based on the Labor Force Survey 2014 ad hoc module on the "Labor market situation of migrants and their immediate descendants". A very high level of proficiency in Spanish is found to enhance immigrants' employability, particularly for non Spanish-speaking immigrants. The impact increases when potential endogeneity in language skills is addressed via IV variables. Still, proficiency in Spanish does not help to get higher ranked occupations, measured via ISEI (International Socio-Economic Index) – and language skills neither contribute to explain occupational status, nor are endogenous to it, even after control for sample selection. The first result confirms the downward bias of the impact of the language proficiency on employment probabilities when the endogeneity problem is not accounted while the second responds to the particular occupational segregation in Spain amongst workers from different areas of the world.

Keywords: Spanish Proficiency, immigrants, labor market outcomes, IV regressions
JEL-codes: J15-J24

1. Introduction

Language skills, ie, the level of understanding, speaking, reading and writing in the host country language, are a particularly important type of human capital for international migrant workers. Studies on the determinants of immigrants' language proficiency and its impact on the labor outcomes, especially earnings, are abundant for the main traditional immigrant-receiving countries like the United States, UK, Germany, Netherlands and Canada, among others. The evidence for Spain is more recent and scarcer since it was transformed from an emigrant-sending country into one of the most relevant immigration countries in the world relatively recently, thanks to the economic and social development of Spain since the 80s of last century, together with relatively friendly migration policies. The analysis of the determinants of language proficiency and its returns among immigrants in Spain is of great interest because Spanish is widely spoken

around the world (Chiswick & Miller, 2015) as well as the peculiarities in the linguistic profile of immigrants in Spain.

Proficiency in host language

The degree of immigrants' mastery in the host country's language usually results from several factors, 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 the knowledge of the local language. They are widely documented in the international literature.

Exposure is usually measured through years of residence in the destination, the composition of the immigrants´ household, namely, marital status and co-residence with parents and/or children, marriage upon arrival to the destination country, especially to a native spouse, among others (Chiswick & Repetto, 2001; Chiswick, Lee, & Miller, 2004; Kulkarni & Hu, 2014; Chiswick & Miller, 2007a, 2007b, 2015).

As for the efficiency in learning the new local language, empirical evidence shows that it declines with age at migration. Young arrivers have much better abilities to learn a new language (Bleakley & Chin, 2004; Miranda & Zhu, 2013; Budría, Ibarreta & Swedberg, 2017). Also immigrants' educational attainment is positively correlated with the learning efficiency, as it is potentially correlated with before–migration exposure to the host language (Chiswick & Miller, 1995, 2007a, 2007b; Isphording & Otten, 2014). Moreover, linguistic distance between the origin and host language makes learning more difficult, reducing efficiency in language skills acquisition: regardless the exposure to a language, the distance between the immigrant's mother tongue and the host country's one will hinder the learning process (Isphording & Otten, 2011).

Concerning economic incentives to learn the host country language, employment status and earnings are expected to increase migrants' willingness to stay in the destination country and thus their inclination to invest in the host country specific human capital, in this case, the language (Zorlu & Hartog, 2018).

Evidence on the determinants of fluency in the host language amongst immigrants in Spain from the National Immigration Survey 2007 (NIS 2007) corroborates the existing one in other receiving countries. Namely, it has been shown to depend on age upon arrival in Spain, marital status, presence of children at home and their level of Spanish, geographical origin, educational attainment, elapsed duration of stay, plans to remain in Spain and, amongst those in employment, occupation and economic sector at the moment of the interview (Gutierrez, Mato & Miyar, 2010; Swedberg, 2010; Ipshording, 2013b; Budría & Swedberg, 2015).

Returns to proficiency in the host language

Most previous empirical studies show the positive impact of language skills on immigrants' earnings in the main receiving countries such as the UK (Dustmann & Fabbri, 2003; Miranda & Zhu, 2013), United States (Chiswick & Miller, 2001; Bleakley & Chin, 2004), Canada (Chiswick & Miller, 1988, 2003), Australia (Chiswick, Lee & Miller, 2005), Germany (Dustmann, 1994; Dustmann & van Soest, 2001, 2002), the Netherlands (Bloemen, 2013) and Israel (Chiswick & Repetto, 2001; Berman, Lang & Siniver, 2003). Many of them are surveyed in Chiswick & Miller (2015), who concluded that a good domain in host language contributes to increased earnings ranging from 5% to 35%

amongst male migrants.

There is also spare evidence on non-monetary returns to language skills on labor market outcomes such as higher job satisfaction and a lower risk of over-qualification (see Bloemen (2013) for an analysis of such types of returns to skills in Dutch). Aldashev et al. (2009) conclude that German proficiency is a significant factor to access employment for immigrants in Germany. Dustmann and Fabbri (2003) analyze the case of UK, and find about 20% higher employment probability for immigrants with good English.

In Spain, migrants' dominance of Spanish has been shown to positively influence success both in the labor market - especially on employability and job quality - and in social integration, although its effect on the latter is weaker (Gutierrez, Mato & Miyar, 2010). Similarly, in Swedberg (2010) and Ipshording (2013b) the mastery of Spanish is shown to have a significant and positive effect on access to employment, but not on income. The latter, initially puzzling result is explained by the inclusion of Hispanic immigrants in the analysis. When they are excluded from the sample, a significant and positive return to proficiency in Spanish (between 23.3% and 27%) on monthly income arises (Swedberg, 2010). Similar results are found in Budría & Swedberg (2015): the return to Spanish language on wages is about 20% in average but varies across educational attainments, ranging from nearly null for low-educated immigrants to around 50 % for the highly educated. And in Budría, Ibarreta, & Swedberg (2017): Spanish language proficiency increases immigrants earning by 17.2% in average but at the top quantile of the earning distribution, the size of the impact reaches 30%. Spanish speaking immigrants tend to come from mid-developed country and are prone to work in low and mid-skills occupations (some of which require a high level of Spanish) while a non-negligible share of OECD migrants tend to hold the most qualified, best paid positions.

Most of the evidence on these issues is based on the National Immigration Survey (NIS) (Gutierrez, Mato & Miyar, 2010; Swedberg, 2010; Ipshording, 2013b; Budría & Swedberg, 2015). The NIS is a cross-sectional one-time survey held in 2007, just before the outbreak of the economic crisis in 2008. At that moment the number of foreign-born residents reached its peak values, 5,249,993 people (estimates for January the first), ie, 11.61% of the total resident population in Spain (Fernández Páez, 2012).

This paper analyses the impact of Spanish proficiency on the labor outcomes, based on the 2014 Spanish Labor Force Survey (LFS) ad hoc module on the "Labor market situation of migrants and their immediate descendants". It therefore portraits a fully different context than the one featured by NIS-2007: the severe depression from 2008 to 2013 reduced labor market opportunities for migrants, with foreign residents being more vulnerable than Spanish native workers in the labor market (Gil-Alonso & Vidal-Coso, 2015). At the first trimester of 2013 the Spanish Labor Force Survey registered peak unemployment rates (39.16% amongst immigrants; 24.98% for Spanish-born; 26.94% in average). As a consequence, a reversion in migration flows took place, and plenty of immigrants returned to their home countries or searched for alternative destinations. According to the Spanish National Statistics Institute, by the 1st of January of 2014, the size of the migrant population in Spain was reduced to 4,677,059, i.e., 10% of the full resident population. It went on shrinking during 2014 up to 4,447,852 by the 1st of January 2015, representing the 9.5% of the total population (Callejo & Fuentes, 2015). At the same time, foreign-born residents unemployment rate was still very high (23.78% in average, 22.42% amongst Spanish natives and 33.65% amongst immigrants), very much

beyond the minimum values registered in the second trimester in 2007 (7.93% in average, 12% amongst immigrants and 7.29% in Spanish born).

This paper contributes to existing literature in various ways: first, as just mentioned, we study the situation of immigrants in Spain in the context of the beginning of the recovery from the Great Recession. We aim to establish whether the impact of language skills on immigrants labor outcomes is still relevant in a slack labor market. Second, we study the impact of Spanish proficiency on labor outcomes different from earnings, namely, access to employment and its quality. The latter is measured via the *New International Socio-Economic Index* [ISEI] adapted to the *International Standard Classification of Occupation 2008* [ISCO-08] in Ganzeboom (2010). ISEI is used to measure job status as it is highly correlated with earnings (non-available in the LFS) and job qualification requirements. Third, we address the potential endogeneity problems in language proficiency when explaining labor outcomes: we deploy the arrival age combined with, first, the linguistic distance between Spanish and the migrants' mother tongues and, second, being born in a Spanish-speaking country. The household composition is also used to instrument the Spanish proficiency.

Our paper is organized as follows. In the next section, we present our data base and describe our sample. In Section 3 we describe our estimation strategies with the IV approach and define some relevant instruments. In Sections 4 and 5 we analyze the results of our multivariate models and conclude.

2. Data description

The 2014 LFS ad hoc module and the sample

Ad hoc European LFS modules are a set of additional questions to the sixth interviews of the Labor Force Survey covering a different topic each year. In 2014 the module was devoted to the "Labor market situation of migrants and their immediate descendants". Questions were addressed to all interviewees aged between 16 and 64. We identify 5,787 first-generation immigrants and 1,211 second-generation immigrants. The latter were born in Spain from at least one non Spanish-born parents and are expected to be monolingual in Spanish, or bilingual in Spanish and their parents' mother tongue (Medvedeva & Portes, 2017). Given they are native speakers, educated in the Spanish education system, and their assimilation process to Spain differs from the one experienced being first generation migrants, we keep them aside from the analysis.

The 2014 ad hoc module questionnaire covers socio-demographic features, self-reported linguistic proficiency in Spanish and labor outcomes, entailing both employment status and job (professional status, type of contract, occupation -ISCO-2008 codes at 2 digits-, supervisory roles) and employer characteristics (size of the company, sector of activity, etc). Some relevant information about the migration process is also available, such as reasons to migrate, arrival age (from which the elapsed period of residence is computed) and whether the migrant had a job offer or employment contract in Spain before migration. Moreover, being a household survey, it is very straightforward to detect also co-resident spouses and children, as well as their place of birth, their elapsed time in Spain and their Spanish proficiency. These variables are usually used to measure the level of exposure to the language.

As regards efficiency in the learning of Spanish, we have added the linguistic distance between the mother tongue of the immigrant and Spanish to the data-set thanks to the detailed home country[1] classifier provided in the data-set. We introduce the linguistic distance from the *Automatic Similarity Judgement Program* (ASJP) developed in a research project from the German Max Plank Institute for Evolutionary Anthropology, described in Brown et al (2008) and Isphording & Otten (2013). This powerful matrix of linguistic distances is already widely used (Isphording (2013a, 2013b) and Isphording & Otten (2011, 2014) to analyze Spain, Germany and the United States; Bloemen (2013) deploys it when studying the Netherlands and Adsera & Pytliková (2015), Adsera & Ferrer (2015) for different OECD destination countries and Canada). It is widely used as instrumental variable, given the negative correlation between linguistic distance with the home language and proficiency in host language found in the abovementioned empirical studies. It is therefore a proxy for language learning efficiency and expected to be unrelated to labor market outcomes.

We also add two additional variables to the data-set in order to describe the quality of the job. The first one is ISEI. We encode all occupations in Module 2014 with the *New International Socio-Economic Index* [ISEI] based on the *International Standard Classification of Occupation 2008* [ISCO-08] constructed by Ganzeboom (2010). It was firstly created in Ganzeboom, De Graaf & Treiman (1992) and explains the relationship between education and income of people from different occupations across 42 countries in the New ISEI. It is used as a measure of socioeconomic status. The ISEI values of our sample varies from 14 (agricultural assistants, cleaners) to 69 (members of executive team, directors). High scores in ISEI indicate higher qualification in the employment and, generally, higher income as well.

There is some evidence that certain high ranked occupations are also quite demanding as regards host language skills. This means that language skills may be relevant to get a job in highly ranked occupations because they increase workers' options to get language-intensive jobs. In order to control for this eventuality we have "plugged" as well via the 2-digit ISCO-2008 occupation identifier the occupation-specific share of workers who reported dealing all the time with people different from co-workers (clients, providers, users, pupils, patients…) in their job, built from the Spanish sub-sample of the 2015 *European Working Conditions Survey*, by Eurofound. We label it as the "dealing with people" index. Interestingly enough, the piecewise correlations between both ISEI and high command in Spanish and the one between the "dealing with people" index and high command in Spanish are quite low, though significant (0.1 and 0.06, respectively), while the "dealing with people" indicator is highly correlated with ISEI (Spearman piecewise correlation is 0.46).

Finally, in order to control for the local labor market conditions in the models prediction employment probabilities, we also add the region-specific unemployment rate for every quarter of 2014 to the data-set, obtained from the Spanish National Statistical Office.

Table A.1. (in the Appendix) displays the main features of the two samples deployed in our analysis: in order to study the probability of having a job, we observe all the 5,787 first generation immigrants of the data-set with no missing data in the explanatory

[1] In the case of countries with more than one official language, like Belgium, Switzerland or Canada (among others) the language distance is the average of the one for the official languages, weighted by the share of population who speaks each of them according to the most recent census.

variables in the multivariate models. From them, 3,036 are working in the moment of the interview and report valid information on the explanatory variables deployed in the estimations. For this latter sub-sample we predict the occupational status via ISEI index, conditional to being in employment.

Around 50% (55% in the employed sub-sample) of the simple report a native level of Spanish, with about 40% of migrants born in Spanish-speaking countries. Around 50% of the sample members were women and one fourth held university degree, most of which come from EU15, EFTA and North America or Oceania, which account for 15% of the overall sample. Given that the migration process is relatively recent in Spain, in 2014 the average elapsed time of residence in Spain was 14 years. Accordingly, as many migrants were still young in 2014, about one third did not report living with a spouse or partner and about 40% did not live with children. The most common reason for migration, comprising around 40% of the sample, was working. Amongst migrant workers, 16% are self-employed and 10% report a supervisory position. Two thirds of first-generation migrants worked in low/mid qualified services.

Spanish proficiency and labor market outcomes

Immigrants are asked to provide a self-assertion of their Spanish proficiency in the ad hoc module. The wording of the question differs from many previous studies[2] is as follows:

Knowledge of spoken Spanish:

1. Mother tongue or speak it as a native
2. Advanced level
3. Intermediate level
4. Beginner or without skills
0. Do not know.

We recode the answers into a binary variable that takes value 1 if the immigrant chooses the first option and "0" otherwise. Around half of first-generation immigrants are very fluent in Spanish: either they are Spanish speakers or they speak Spanish as natives would do (Table 1); nearly 26% self-assess their level as "advanced", and 16,4 and 7.8 % report "intermediate" and "beginner" levels, respectively. This distribution very much differs among those born in Spanish speaking countries, 94% of which report Spanish as their native language and those who were not born in Spanish speaking countries, and one fifth self-assess their level of Spanish as "native language" or equivalent to native level.

******* Table 1 about here ********

As regards the two labor market outcomes explored in this paper, employment rates are, in average, higher for non Spanish speakers, but widely vary along levels of command in Spanish: they are 20 percentage points higher in those who report a native level than in those who self-label as beginners. In the overall sample, ISEI average levels do not differ between those who self-report advanced or native levels of Spanish, though; non-native

[2] Self-assessed language skills usually rank from "very well" to "not at all". In NIS-2007 the variable indicating language skills holds values ranging from "very well" to "needs improving" (Budría & Swedberg 2015, Budría, Ibarreta & Swedberg, 2017 and Isphording 2014). Other pieces of research stress difficulties with the host language (Yao & van Ours (2015) address difficulties with Dutch language while Miranda and Zhu (2013) take deficiencies in English to proxy poor language skills).

Spanish speakers work, in average, in occupations ranked higher than native Spanish speakers – and again there is an increasing trend in ISEI along the distribution of Spanish proficiency for non-native speakers.

3. Empirical strategy

Instrumental variables (IV)

In Yao & van Ours (2015) three arguments are provided for the need to validate the assumption of exogeneity of language skills and absence of measurement errors underlying the OLS estimates: Firstly, unobserved heterogeneity may be correlated with both labor market performance and language skills. For instance, more motivated immigrants will make a greater effort to learn the host language and, at the same time, they are more likely to have better labor market outcomes. Secondly, reverse causality, i.e., labor market performance reversely contributes to the proficiency in the host language. Thirdly, self-reported information suffers from measurement errors. Unobserved heterogeneity and reverse causality lead to an upward bias in the parameter estimate of the language effects while measurement errors lead to a downward bias.
Instrumental variables are widely used to address the endogeneity between language proficiency and labor outcomes as earnings. Instruments are observed variables that affect the language acquisition but are not related to the unobservable ability and labor market outcomes (Isphording, 2015). Finding a good instrument is really difficult. Most of the common ones in the literature are proxies for exposure to the host language, but other strategies are also followed: in Chiswick & Miller (1995, 1998), IVs are dummy variables for overseas marriage, plus number and age of children at household and minority concentration measures. The availability of panel data allows Dustmann & van Soest (2002) to deploy leads and lags of language skills and father's education as IVs. Bleakley & Chin (2004, 2010) and Miranda & Zhu (2013) use the interaction of age at arrival and dummy for born in non-English speaking countries as IVs. Budría & Swedberg (2015) use also early arrival age, co-residence with children fluent in Spanish and planning to stay in Spain for next five years as IVs. In all the above-mentioned evidence, IV estimates for the impact of language ability on wages largely over-exceed uncorrected OLS ones.

In this paper, we combine three variables to build instrumental variables aimed to correct the possible endogeneity problems in the returns to host language proficiency: linguistic distance between Spanish and immigrants' mother tongue, age at arrival and being born in a Spanish – speaking country.

The first one – linguistic distance - is commonly used in the existing empirical studies, not only for traditional immigrants receiving countries but also for Spain (Bleakley & Chin, 2004 & 2010; Swedberg, 2010; Miranda & Zhu, 2013; Yao & Ours, 2015; Budría & Swedberg, 2015; Budría, Ibarreta, & Swedberg, 2017). This is a relatively new instrumental variable in the literature, already used in studies for Spain (Isphording, 2015). It requires assuming that the linguistic distance is a significant variable to the language proficiency and unrelated with the labor market outcomes, or only affect the latter one through the language proficiency.

Since age at arrival in the destination negatively influences language learning efficiency, young arrivers are more fluent in the destination language (Chiswick & Miller, 2009;

Miranda & Zhu, 2013; Yao & Ours, 2015; Budría, Ibarreta & Swedberg, 2017; Zorlu & Hartog, 2018). Linguistic distance is highly significant to the language learning, but it hardly means any difference for those who arrived at a young age at the destination country. This is why linguistic distance is sometimes interacted with early arrival dummies (as in Isphording (2014) and Isphording and Otten (2013)). We also use the interaction of early arrival age and linguistic distance as an instrument.

Figure 1 shows the average values of the regression-based predicted probability of a very high level of command in Spanish along the distribution of linguistic distance[3] ranked into deciles to make it more intuitive. Most immigrants arriving in Spain at a very early age[4] (immigrants arriving before the age of 4 are labeled in Figure 1 as "Under 4"; otherwise, they are "Over 3" , i.e., they were elder than 3 at the moment of arrival) report the highest level of Spanish proficiency regardless the distance between their mother tongues and Spanish. The only exception is decile 8, where the most common language is Arab and the most common nationality is Morocco. This group is not only quantitatively large (they account for 12 % of the first generation migrants sub-sample) but also live in communities where some members may well have limited need to speak Spanish on daily basis. Amongst those who arrived in Spain at a later age, we have found a very different profile in those who come from a set of countries we have labeled as "highly developed" (they include the EU-15, EFTA – European Free Trade Area - countries, North America and Oceania) and those who do not. The latter group self-report a decreasing trend in the share of self-reported native (or its equivalent) level of Spanish along the distribution of linguistic distance, while in the first group the trend is far from clear: French, English and German native speakers (in decile 4 of the linguistic distance distribution) report higher levels of Spanish than could be expected from their linguistic distance. We think this may be explained by higher ex-ante exposure to Spanish language and level of education, both of which considerably increase language learning efficiency. This is why the origin country, grouped in eight categories, is always controlled for in our multivariate analysis to proxy for the quality of the education system and cultural and institutional similarities between Spain and the origin countries. Moreover, in multivariate estimations errors are clustered around the exact country of origin reported by interviewees.

- *Figure 1 about here* -

Another set of instrumental variables refers to the combination of arrival age and being born in Spanish Speaking countries. Arrival age influences Spanish proficiency only for those who did not come from Spanish-speaking countries, with early arrivers hardly reporting lower levels of Spanish command than Spanish native speakers[5]. From the

[3] The regression is a probit model where reporting a native level of Spanish is explained from age at arrival, education attainment, a dummy that represents that at least part of the education took place in Spain, country of origin, spouses' self-reported level of Spanish, number of children in schooling age, linguistic distance and fixed effects for origin country groups. Results are available upon request.

[4] Critical age range of learning a foreign language stands between 5-15 years (Chiswick et al. 2008), but given the high level of Spanish proficiency we are imposing for our dependent variable to take value 1, only really early arrivers (under the age of 4) do report the highest proficiency regardless the distance between their mother tongue and Spanish. Given that pain became a migration receiving country in the last decades, the share of early arrivals about for about 5-6% only.

[5] Different pieces of research show that there is no significant distance in host language performance – ie., differences in means tests are not significant – between immigrants sharing the destination country mother tongue and the rest as long as they arrived at an early age (see Isphording (2014)). Given the way our

trends displayed in Figure 2 it is inferred that our instrument will be the interaction of arrival age and having a Spanish speaking origin. The latter has proved to be a good instrument in Bleakley and Chin (2004), and Budria *et al* (2017), as well as Yao and van Ours (2015).

- *Figure 2 about here* -

Part of the effect of arrival age 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 also means a higher likelihood of studying in the host country and, therefore, full recognition of the migrant's qualification. Moreover, the linguistic distance may reflect cultural similarities, as countries with the same language or languages derived from the common one may share more cultural similarities, which will make easier the integration of the immigrants upon arrival at the host country. This is why we include country clusters amongst the explanatory variables and, again, cluster errors across home countries.

Empirical strategy*: *IV models

In our prediction of both the probability of being in employment and job status (measured via ISEI) we are interested in testing whether Spanish proficiency (a) influences both labor market outcomes and (b) whether this influence is underestimated because of potential endogeneity problems. In order to satisfy (b) we use instrumental variables, namely, we adopt a GMM (generalized method of moments) – IV approach. The Stata routine ivreg2 here deployed allows for the estimation of a large array of test for the quality of the instrumental variables (Baum, Schaffer & Stillman, 2003).

Our empirical strategy (very much driven from Yao & van Ours (2015)) may be described as follows:
We first obtain OLS estimates for a linear probability model on the likelihood of being in paid work at the moment of the interview. This implies accepting the assumption of Spanish proficiency being exogenous to the employment status:

$$E_i = X_i\beta + \gamma L_i + \varepsilon_i \qquad (1)$$

where *Ei* denotes a dummy variable that takes value 1 if the individual is employed at the interview. X_i is a vector of exogenous variables that comprises age (and its square), gender, education attainment, reasons for migrate, region-specific unemployment quarterly rate and origin country groups. L_i is the dummy variable capturing a very high level of Spanish proficiency and β is a vector of parameters, with γ representing the effect of language proficiency on the likelihood of employment. Finally, ε_i is the error term. Standard errors are cluster at the level of the household.

Then, in order to correct potential bias in the parameters for language skills we use an instrumental variable approach, exploring two complementary strategies. The first one, inspired in Isphording (2014) and Isphording and Otten (2013), consists on the interaction between early arrival age and linguistic distance between migrants' home language and Spanish; the set of excluded instruments (vector I_i) comprises as well whether the

dependent variable is defined, we cannot prove that this is the case in our data – set.

migrant's highest level of education was obtained already in Spain. At the same time, the set of included instruments comprises for origin country groups fixed effects that point at economic and institutional differences with Spain (see Figure 1).
The vector of excluded instruments also comprises measures of exposure via household composition – number of co-resident children at school age and spouse's elapsed period of residence in Spain. A dummy indicating if any of her parents held a university degree and another one expressing whether the interviewee's last period in education had taken place in Spain [6] featuring efficiency in learning included in the set of excluded instruments.

$$L_i = X_i\beta_1 + \theta I_i + \varepsilon_{1i} \qquad (2)$$

The second strategy is inspired in Bleakley & Chin (2004) and Yao & van Ours (2015). In the baseline estimates to explain language skills L_i, we use one instrumental variable (I_i) that consist on the interaction between arrival age and a dummy variable indicating whether the migrant was born in a Spanish-speaking country. If immigrants born in Spanish speaking countries experience different assimilation trajectories from the rest, non-language age-at-arrival effects on labor market performance may not be the same for the two groups of immigrants (see Yao & van Ours (2015) for a detailed explanation). This is why we include age at arrival in the set of excluded instrumental variables. This is meant to allow for age-at-arrival effects on the language channel for immigrants who come from Spanish speaking countries. In addition, we add country of birth groups fixed effects to control for some non-language channels as well. Moreover, the same instruments regarding exposure (number of co-resident children at school age, spouse's elapsed period of residence in Spain) and efficiency (highly educated parents and last period in education in Spain) are also taken as further control variables in vector *Ii*.

Finally, in both scenarios β_1 is a vector of parameters for the included instruments, θ is a set of parameters for the set of excluded instruments, and ε_{1i} is an error term.

When ISEI is estimated via OLS, the empirical strategy takes into account that ISEI scores are only observed for employed individuals and the estimation of ISEI may therefore be subject to sample selection bias. We thus perform a two-step Heckman model, where the Inverse Mills Ratio is obtained from a probit model predicting the likelihood of employment from the very same set of explanatory variables as the linear probability model described above. In order to ease interpretation of results, the dependent variable, i.e., the ISEI score for the immigrant's occupation is standardized to the mean, so it ranks from 0 to 1.

$$ISEI_i = Z_i\beta + \gamma L_i + \delta IMR_i + \varepsilon_i \qquad (3)$$

The set of explanatory variables in (3) (vector Z_i) comprises personal characteristics and job features. The former entail gender, elapsed years of residence in Spain (and its square), age at which education was finished/interrupted and main reason to migrate[7]. The latter are the standardised occupation-specific "dealing with people" index, job status (self-

[6] This dummy has been built from the difference between age at which education was finished or interrupted and arrival age. If the event education is finished after the arriving to Spain we can assert that at least part of it took part in Spain.

[7] This time the variable allows to distinguish as well between those who migrated for employment reasons and had already a contract or job offer and those who did not have it.

employed versus both permanent and temporary employees), working part-time versus full-time and tenure (in months). As for the employers characteristics, they are broadly controlled for via sector of activity (in four categories) and firm size (below or beyond 10 employees). Finally, the origin country groups fixed-effects are also controlled for so that the coefficient linked to L_i expressing language proficiency reflects only the language skills, net from cultural or institutional peculiarities that influence the speed in the labor market insertion. Moreover, the sample is clustered across home country codes.

As for the IV-GMM approach, the two alternative strategies entail the set of excluded instruments (vector I_i) in equation (2).

4. Results

1) Spanish proficiency and immigrants employability

In Table 2 the coefficients for the linear model aimed at predicting the likelihood of being in employment at the interview are displayed. Columns (1) and (2) refer to the OLS before and after controlling for origin country group, while columns (3) and (4) report the coefficients for the two above described strategies in the IV-GMM approach. Migrants self-reporting a very high command in Spanish enjoy a 8.4 percentage points higher likelihood of being employed than the rest. Once the non-language effects of language proficiency are controlled for via the set of origin country dummies, the effect diminishes a bit, although the difference is not significant, as the confidence intervals of the relevant coefficients overlap. Under the IV framework, the size of the relevant coefficient considerably increases up to 26 percentage points. At the bottom of Table 2 a set of tests to verify the quality of the instruments is displayed (see Baum et al (2003) for a description of the different tests): our estimates do not suffer from neither under-identification nor over-identification, they are reasonably orthogonal to the errors in the first equation and overall language skills are indeed endogenous to the probability of being in employment.

This result points that the positive impact of Spanish proficiency on the likelihood of being employed is downward biased in the OLS estimates, as IV estimates show – in line with the international empirical evidence (Isphording, 2015). That bias is likely due to unobservable ability which affects the labor outcomes of immigrants as well that exceed the negative one due to measurement error.

***** Table 2 *****

The rest of variables follow the expected sign and significance levels in both the OLS and IV-GMM strategies: age and education attainment increase likelihood of being employed, while women's are a bit less likely to be in paid employment than men. Migrants whose main motivation to migrate was working are more likely to work as well than those with other motivations (except for those who migrated to study). The prevalent unemployment rate in the region of residence considerably reduces migrants' likelihood of being at work and, finally, non-European migrants tend to be less employable than European ones, with those from Latin American and North-African (Magreb) countries facing particular difficulties (these coefficients are not shown for space reasons but available on request).

1) Spanish proficiency and occupational status

Table 3 reports the estimates for the (standardized) ISEI scores. Again, columns (1) and (2) refer to the OLS before and after controlling for origin country group, while columns (3) and (4) report the coefficients for the two adopted strategies in the IV-GMM approach, described above. At the bottom of Table several tests for the quality of the instruments are displayed: overall, language skills are exogenous to ISEI. Consistently our sets of instruments suffer from over-identification. The preferred empirical strategy in this case would be OLS with sample selection. Proficiency in Spanish is not significantly correlated with ISEI ranking (the size of the negative coefficient is small and hardly significant).

**** Table 3 ****

The negative (if any) relation between proficiency in Spanish and ISEI score may have to do with the assignment of migrants from Spanish speaking countries, who are often mid and low-qualified, towards low and mid-skills occupations. At the same time, a non-negligible share of OECD migrants tends to hold the most qualified, supervisory or administrative positions. The origin country groups fixed effects – not shown for space reasons - confirm the occupational assignment for workers from different areas, with those from NMS (new member states) EU-countries, Latin American, North-African and Asian origin being assigned to lower status jobs than those from EU-15 countries, whose average status is only exceeded by North-American and Oceanians.

Interestingly enough, this result is confirmed when controlling for our "dealing with people" index; it seems paradoxical, but language use with people different from clients and co-workers and not language skills are positively related to ISEI. This may be related with supervisory positions, which are also controlled for.

As for the rest of the explanatory variables, their contribution to ISEI is as expected: women's occupational status is lower than men's; age when education was finished/interrupted is positively related with ISEI. The elapsed time of residence in Spain does not seem to influence ISEI, though. This is somehow worrying because it would possibly entail a rather low level of upward occupational mobility upon migration, for which further evidence is very much needed. Interestingly, those who migrated for employment reasons work in positions with lower status than those who migrated for family reasons and, overall, to undertake education. Job status is higher for the self – employed, while those in part-time positions work in lower ranked occupations, *ceteris paribus*. Firm size is not related to job status but jobs in the services sectors, both qualified and non-qualified[8], are featured by higher average status than those in primary and industrial activities.

c) Sensitivity checks: differences across migrant sub-groups

Given the high correlation between ISEI index and wages, our results are consistent with Swedberg (2010) and Ipshording (2013b), who obtain significant positive effect on access to employment, but not on income. The latter result is affected by the inclusion of Spanish

[8] Services sectors are classified in low-mid qualified (retail, hospitality and housekeeping, among others) and high qualified ones (mostly, financial and professional services, services related wo the public sector).

speakers in their sample, and it becomes more positive when they are excluded from it (Swedberg, 2010). Language skills are complementary to other types of human capital, with language skills endowments significantly differing across individuals with different education attainment. Moreover, good command of Spanish particularly contributes to wages for highly qualified workers and at the top of the wage distribution (see Budría & Swedberg (2015) and Budría, Ibarreta, & Swedberg (2017), respectively). In order to verify if our results may also confirm the complementarities between types of human capital we have estimated our models for non-Spanish speaking migrants and for sub-samples of migrants according to their level of education (results non-shown for space reasons but available from the authors upon request).

Our preliminary results point at returns to Spanish proficiency in terms of employment opportunities being larger in non-native Spanish speakers than in the overall sample of first generation immigrants. Moreover, the higher the education attainment, the higher return to language skills, confirming that they are complementary types of human capital. Still, results are so far exploratory as the quality of the instruments is not yet satisfactory in all sub-samples. Finally, and alike previous similar evidence on returns on wages, language skills are not clearly related to ISEI scores, as they are intrinsically connected with formal education and not so much with Spanish proficiency.

5. Conclusions

Determinants of the Spanish proficiency among immigrants and return to language skills as regards migrants' performance in the Spanish labor market has received increasing attention since the economic boom in the mid-2000s, when Spain became one of the most relevant destination countries in Europe. Moreover, Spanish language is widely spoken around the world. Still, occupational attainments of Spanish-native speaking migrants are featured by the level of development of their origin countries, and not by their command of Spanish (and, what is more, not even by their level of formal qualifications).

Most of the evidence about immigrants' labor market performance in Spain refer to the economic pre-2008 recession migration and economic boost, and one of the contributions of this paper is the analysis of a more recent period. In the new context, with lower levels of employment and stagnant incomes, good level of Spanish command is still very relevant for employability. Unfortunately, it is difficult to grasp if it is more or less relevant than before the crisis as our main explanatory variable is encoded in a completely different way than previous data-sets (NIS 2007).

The proficiency in Spanish language has also been proved to be endogenous to employability, particularly for non-native Spanish-speakers (confirming previous evidence), while exogenous to occupational status measured via ISEI index. We effectively instrument the language proficiency by a set of proxies for exposure to language and efficiency in learning, that turn not to contribute to improve the estimates for the occupational outcomes of migrants, not even when native Spanish speakers are excluded from the sample or when the estimates education attainment specific. The results about returns to Spanish command as regards occupational status differ from evidence in USA or UK or Germany. We think this is due to the fact that the allocation of migrants across occupations is more related to formal qualifications and institutional similarities than to language skills. Spanish native speaking migrants (usually, from Latin American countries) are not amongst the highest qualified ones, neither their qualifications are

always acknowledged in the Spanish labor market. Moreover, many of them are overeducated: in our sample, two thirds of immigrants from Latin America report their jobs are below their qualification, compared to half of those from EU-15, EFTA and North America or around one third of those coming from Magreb countries and Asia.

From our results we may indeed conclude that the acquisition of Spanish proficiency contributes to employability, and should therefore be an important part of labor market and social policies addressed to migrants. But, in order to improve occupational status of all migrants, educational policies contributing to migrants' general skills and formal educational attainment or updating and recognizing the skills they bring from their home countries would be imperative.

Longitudinal data-sets allowing to see if improvements in Spanish proficiency contribute to upward occupational mobility and/or if employment in itself contributes to improving Spanish proficiency would also help to better design migration policies as well as improve the quality of future estimations of monetary and non-monetary returns to language skills.

References

Adsera, A., & Pytlikova, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586), F49-F81.

Adsera, A. & Ferrer, A. M., (2015). *The effect of linguistic proximity on the occupational assimilation of immigrant men in Canada*, IZA Discussion Paper No.9499.

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.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata journal*, 3(1), 1-31.

Berman, E., Lang, K. & Siniver, E. (2003). Language Skill Complementarity: Returns to Im-migrant Language Acquisition, *Labor Economics*, 10(3):265-290.

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

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal*: *Applied Economics*, 2(1), 165-92.

Bloemen, H. (2013). *Language Proficiency of Migrants: The relation with Job satisfaction and Matching*, IZA Discussion Paper No. 7366.

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., 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.

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., & Miller, P. W. (2003). The complementarity of language and other

human capital: Immigrant earnings in Canada. *Economics of Education Review*, 22(5), 469-480.

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2004). Immigrants' language skills: The Australian experience in a longitudinal survey. *International Migration Review*, 38(2), 611-654.

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 Sook Lee, J., Rumberger, R.W., and T.G. Wiley (eds) *The education of language minority immigrants in the United States*, Bristol, UK: Multilingual Matters, pp. 72-91.

Chiswick, B.R., & Miller, P.W. (2007a). *The economics of language: International Analyses*. Routledge, New York.

Chiswick, B.R., & Miller, P.W. (2007b) Immigration: Trends, Consequences and Prospects for the United States, *Research in Labor Economics*, Vol. 27, 215–270.

Chiswick, B. R., & Miller, P. W. (2015). International migration and the economics of language. In Chiswick, B. R., & Miller, P. W. (eds) *Handbook of the Economics of International Migration* (Vol. 1) North-Holland. pp. 211-269

Chiswick, B. R, & Repetto, G. (2001). Immigrant Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and their Effects on Earnings, in S. Djajic (ed) *International Migration: Trends, Policies, and Economic Impact*, London: Routledge, pp. 204-228.

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants, *Journal of Population Economics*, 7(2): 133-156.

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. (2001). Lan-guage Fluency and Earnings: Estimation with Misclassified Language Indicators, *The Review of Economics and Statistics*, 83(4): 663-674.

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.

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.

Isphording, I. E., & Otten, S. (2013). The Costs of Babylon—Linguistic Distance in Applied Economics. *Review of International Economics*, *21*(2), 354-369.

Ishpording, I. (2013a). Disadvantage of Linguistic Origin: Evidence from Immigrant Literacy Scores, *Economics Letters*, 123(2): 236-239.

Ishpording, I. (2013b) *Returns to Local and Foreign Language Skills – Causal Evidence from Spain*, Ruhr Economic Papers No. 398.

Isphording, I. E. (2015). Language and labor market success. *International Encyclopedia of the Social & Behavioral Sciences*, 260-265.

Isphording, I. & Otten S. (2011). *Linguistic Distance and the Language Fluency of Immigrants,* Ruhr Economic Papers No. 274.

Isphording, I. & Otten, S. (2014). *Linguistic Barriers in the Destination Language Acquisition of Immigrants*, IZA Discussion Paper No. 8090.

Kulkarni, V. S. & Hu, X. (2014). English language proficiency among the foreign born in the United States, 1980–2007: Duration, age, cohort effects, *International Migration Review*, 48(3): 762-800.

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.

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

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 1 - Language skills, native Spanish – speakers and labor market outcomes

		Native level	Advanced	Inter-mediate	Beginner	Total
Sample composition	Overall sample	50.00	25.85	16.38	7.77	100.00
	Non-Spanish-speakers	20.72	39.95	26.66	12.68	100.00
	Spanish speakers	94.07	4.63	0.91	0.39	100.00
Employment rates	Overall sample	56.03	53.02	50.54	35.43	52.75
	Non-Spanish-speakers	55.12	52.32	49.9	35.79	50.16
	Spanish speakers	56.33	62.20	79.01	18.04	56.66
ISEI scores	Overall sample	33.85	34.86	27.22	29.66	32.85
	Non-Spanish-speakers	41.93	34.42	27.13	29.49	33.75
	Spanish speakers	31.23	39.65	29.82	45.87	31.66

Source: 2014 Spanish LFS ad hoc module.

Figure 1 - Proficiency in Spanish by linguistic distance and develpment in home country

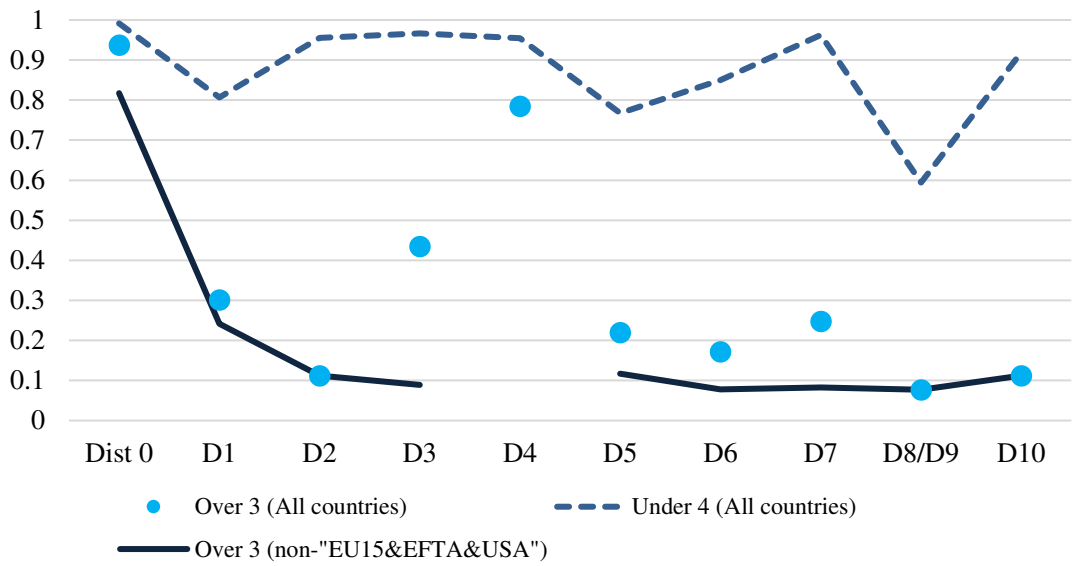

Figure 2 - Spanish Proficiency by arrival age and home country language

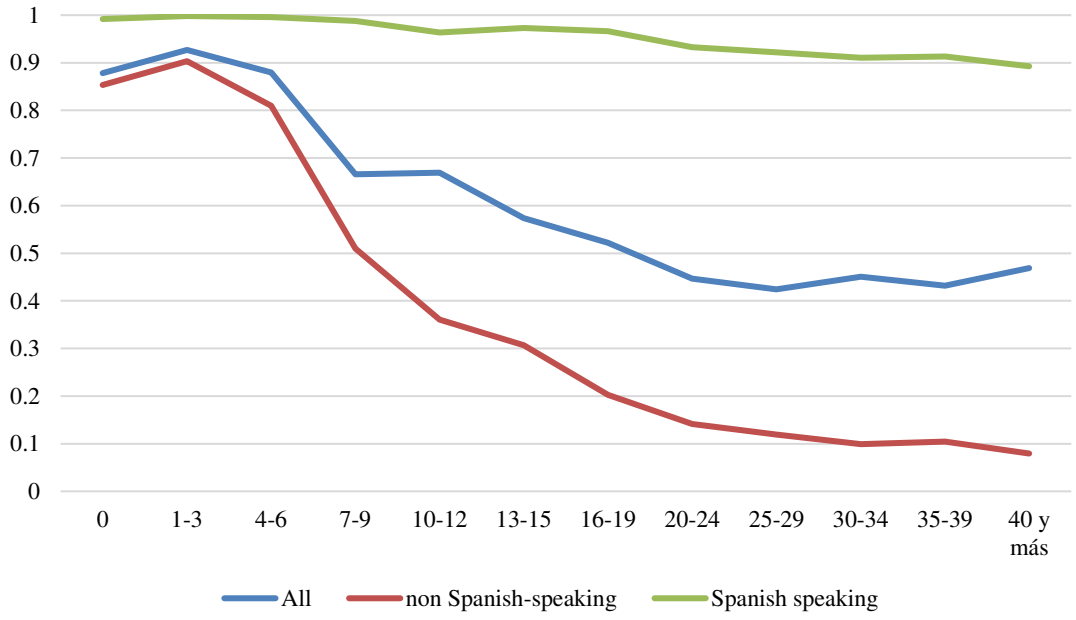

Table 2 OLS and IV estimates – being in employment – linear probability models

	OLS		IV-GMM	
	(I)	**(II)**	**Specif 1**	**Specif 2**
Proficiency in Spanish	0.084**	0.060**	0.266***	0.259***
	(0.035)	(0.025)	(0.057)	(0.062)
Age	0.061***	0.060***	0.061***	0.061***
	(0.005)	(0.005)	(0.003)	(0.003)
Age (squared)	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Gender: female	-0.062***	-0.066***	-0.063***	-0.065***
	(0.020)	(0.021)	(0.018)	(0.018)
Primary education	0.149***	0.083**	0.039	0.043
	(0.028)	(0.032)	(0.027)	(0.027)
Lower secondary	0.181***	0.096***	0.044*	0.038
	(0.031)	(0.030)	(0.024)	(0.025)
Upper secondary	0.215***	0.126***	0.076***	0.078***
	(0.032)	(0.029)	(0.024)	(0.025)
Higher education	0.300***	0.206***	0.178***	0.174***
	(0.032)	(0.030)	(0.024)	(0.024)
Family	-0.117***	-0.117***	-0.149***	-0.148***
	(0.020)	(0.018)	(0.017)	(0.016)
Education	-0.069	-0.068	-0.024	-0.035
	(0.054)	(0.050)	(0.039)	(0.039)
Refugee	-0.237***	-0.235***	-0.245***	-0.264***
	(0.086)	(0.083)	(0.086)	(0.077)
Other reasons	-0.082***	-0.090***	-0.080***	-0.072***
	(0.028)	(0.027)	(0.024)	(0.023)
Regional quarterly unemplomet rate	-0.512***	-0.462***	-0.481***	-0.477***
	(0.113)	(0.116)	(0.100)	(0.104)
Country of origin (in groups)	No	Yes	Yes	Yes
Constant	-0.695***	-0.540***	-0.599***	-0.603***
	(0.070)	(0.076)	(0.078)	(0.078)
Observations	5,729	5,729	5,644	5,729
Clusters	102	102	63	102
R^2	0.141	0.159	0.137	0.140
Quality of instruments (tests)				
F test of excluded instruments (F)	----	----	77.78***	7.15***
Underidentification: Kleibergen-Paap rk LM statistic	----	----	17.329*	16.989*
Weak identification: Cragg-Donald Wald F	----	----	94.485	57.750
Hansen J statistic (overidentification):	----	----	5.767	7.061
C statistic (exogeneity/orthogonality):	----	----	5.767	4.274
Endogeneity test of endogenous regressors:	----	----	3.894*	2.982*

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1; Source: 2014 Spanish LFS *ad hoc* module.

Table 3 OLS and IV estimates – ISEI index– with sample selection

	OLS		IV_GMM	
	(I)	**(II)**	**Specif 1**	**Specif 2**
Proficiency in Spanish (native level)	-0.034*	-0.031*	-0.101***	-0.045
	(0.019)	(0.016)	(0.029)	(0.029)
"Dealing with people different from co-workers"		0.314***	0.334***	0.329***
		(0.026)	(0.018)	(0.020)
Gender: female	-0.064***	-0.051***	-0.044***	-0.042***
	(0.009)	(0.008)	(0.006)	(0.006)
Age when education was finished / interrupted	0.007***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Supervising position	0.127***	0.125***	0.112***	0.116***
	(0.016)	(0.019)	(0.015)	(0.014)
Firm larger than 10 employees	0.017	0.006	0.003	0.007
	(0.012)	(0.011)	(0.007)	(0.008)
Part-time work	-0.037***	-0.019**	-0.015**	-0.011*
	(0.011)	(0.008)	(0.006)	(0.006)
Employee - permanent position	-0.073***	-0.031**	-0.036***	-0.035***
	(0.012)	(0.012)	(0.010)	(0.010)
Employee - temporary position	-0.061***	-0.030**	-0.035***	-0.034***
	(0.013)	(0.014)	(0.011)	(0.012)
Manufactures, energy, construction	0.268***	0.197***	0.196***	0.192***
	(0.013)	(0.014)	(0.011)	(0.012)
Low/mid qualified services	0.207***	0.064***	0.053***	0.049***
	(0.013)	(0.014)	(0.011)	(0.010)
High qualified services	0.503***	0.350***	0.344***	0.348***
	(0.015)	(0.022)	(0.014)	(0.016)
Country of origin (in groups)	Yes	Yes	Yes	Yes
Heckman's Mills Lambda	-0.097***	-0.111***	-0.141***	-0.131***
	(0.026)	(0.023)	(0.022)	(0.021)
Constant	0.105***	0.099***	0.167***	0.134***
	(0.038)	(0.032)	(0.033)	(0.037)
Observations	2,736	2,736	2,688	2,736
	90	90	90	90
R-squared	0.546	0.623	0.618	0.620
Quality of instruments (tests)				
F test of excluded instruments (F)			31,05***	229.01***
Underidentification: Kleibergen-Paap rk LM statistic			22.882*	20.129**
Weak identification: Cragg-Donald Wald F			38.770	85.590
Hansen J statistic (overidentification):			20.774*	22.670**
C statistic (exogeneity/orthogonality):			0.747	1.486
Endogeneity test of endogenous regressors:			0.000	0.061

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors in parentheses. Other control variables are reasons to migrate, Years of residence in Spain and its square; tenure – in months.

Source: 2014 Spanish LFS *ad hoc* module.

Appendix: Sample statistics

Table A.1. Means values of dependent and explanatory variables in multivariate models			
Dependent variables	**All sample**		**In employment**
In employment	0.527	ISEI (standardized)	0.348
		ISEI (st. dev)	(0.282)
Explanatory variables - personal features and migration history			
Spanish proficiency (native level)	0.505	Spanish proficiency (native level)	0.547
Age (mean)	38.672		
Age (st. dev)	(11.593)		
Male	0.469	Male	0.497
Female	0.531	Female	0.503
Education attainment			
Iliterate	0.058		
Primary education	0.120		
Lower secondary	0.250		
Upper secondary	0.319		
Higher education	0.253		
		Age when education ended (mean)	19.860
		Age when education ended (st. dev)	(5.971)
Arrived before 4 years old	0.046	Arrived before 4 years old	0.056
		Years of residence in Spain (mean)	14.053
		Years of residence in Spain (st. dev)	(10.181)
Reasons to migrate		Reasons to migrate	
Working	0.450	Working, no contract	0.397
		Workng, contract or job offer	0.144
Family	0.444	Family	0.343
Education	0.025	Education	0.029
Refugees	0.005	Refugees	0.004
Other reasons	0.077	Other reasons	0.083
Standardised language distance (mean)	0.514	Standardised language distance (mean)	0.475
Standardised language distance (st. dev)	(0.431)	Standardised language distance (st. dev)	(0.429)
Born in a Spanish speaking country	0.406	Born in a Spanish speaking country	0.444
At least one parent highly educated	0.172	At least one parent highly educated	0.171
At least partly educated in Spain	0.277	At least partly educated in Spain	0.220
Number of children in school age		Number of children in school age	
Does not live with children	0.441	Does not live with children	0.405
Lives with children, none of which is in school age	0.249	Lives with children, none of which is in school age	0.255
Lives with at least one child in school age	0.311	Lives with at least one child in school age	0.340
Spouse's time of residence in Spain		Spouse's time of residence in Spain	
Does not live with spouse	0.355	Does not live with spouse	0.305

Spanish-born spouse	0.177	Spanish-born spouse	0.205
Less than 10 years of residence	0.192	Less than 10 years of residence	0.220
10 or more years of residence in Spain	0.277	10 or more years of residence in Spain	0.269
Broad home country groups		Broad home country groups	
EU-15 / EFTA	0.148	EU-15 / EFTA	0.177
New EU-28 member States	0.167	New EU-28 member States	0.173
Other European countries	0.032	Other European countries	0.028
Latin American countries	0.420	Latin American countries	0.454
Magreb	0.161	Magreb	0.088
Rest of African countries	0.019	Rest of African countries	0.011
Asia	0.045	Asia	0.059
North America and Ocenia	0.007	North America and Ocenia	0.010
Other variables			
Regional quarterly unemploment rate (mean)	0.241	Heckman's Mills Lambda (mean)	0.683
Regional quarterly unemploment rate (st. dev)	(0.060)	Heckman's Mills Lambda (st. dev)	(0.276)
		Job features	
		Dealing with people different from co-workers continuously (mean)	0.447
		Dealing with people different from co-workers continuously (st. dev)	(0.296)
		Employment status	
		Self-employed	0.158
		Employee - permanent position	0.555
		Employee - temporary position	0.287
		Tenure (in months) (mean)	64.95
		Tenure (in months) (st. dev)	(69.54)
		Supervisory position	0.101
		Firm larger than 10 employees	0.365
		Part-time work	0.252
		Sectors of activity	
		Agricuture and primary activities	0.059
		Manufactures, energy, construction	0.155
		Low/mid qualified services	0.629
		High qualified services	0.157
Number of Observations	5,644	Number of Observations	2,688

Source: 2014 Spanish LFS *ad hoc* module. Numbers in brackets report standard deviations.