The Impact of Language Proficiency on Immigrants' Earnings in Spain

Santiago Budría
Pablo Swedberg

October 2012

The Impact of Language Proficiency on Immigrants' Earnings in Spain

Santiago Budría
CEEAplA and IZA

Pablo Swedberg
St. Louis University

ABSTRACT

The Impact of Language Proficiency on Immigrants' Earnings in Spain[*]

This article uses micro-data from the Spanish National Immigrant Survey to investigate the impact of Spanish language proficiency on immigrants' earnings. The results, based on Instrumental Variables (IV), point to a substantial earnings return to Spanish proficiency, of approximately 27%. This figure varies largely across educational groups, with high-qualified workers earning a premium of almost 50%. This conspicuous complementarity between formal qualifications and language skills poses a challenge for traditional language training policies, for these typically neglect the immigrants' heterogeneous educational background.

Corresponding author:

Santiago Budría
Department of Economics
University of Madeira
Rua Penteada 9000-390
Funchal
Portugal
E-mail: sbudria@uma.pt

---

1. Introduction and background

This is the first paper to explore the impact of Spanish language proficiency on the labour market earnings of immigrants in Spain. The motivation of the paper is twofold. First, there exists now a significant literature examining how language affects earnings and employment (Carliner, 1981, McManus et al., 1983, Grenier, 1987, Rivera-Batiz, 1992, Chiswick, 1998, Dustmann & Soest, 2001, Dustmann & Fabbri, 2003, Lui, 2007, Chiswick & Miller, 2010). Despite the fact that existing work typically shows a positive impact of language ability on earnings, the results are mostly based on English speaking countries, strongly case dependent and can be hardly transferred across labour markets. Existing research in Spain has focused on one of the Spanish economically leading regions, Catalonia, and its regional language, Catalan (Rendón, 2007, Di Paolo, 2011, Di Paolo & Raymond, 2012). This focus is partly due to the recent availability of the Survey of Living Conditions and Habits of the Catalan Population, carried out by the Statistical Institute of Catalonia. Despite pioneering, however, these efforts have not been seconded by studies focusing on Spanish, the nationwide spoken language, nor on the country as a whole. This paper is intended to fill this gap.

Secondly, the extent to which language skills affect an individual's labour market performance has implications about the income and poverty levels of immigrant families and ultimately affects their social and cultural integration to the host country society. It has been proven that immigrants have a negative wage gap respect to observationally equivalent natives, even though the degree of earnings assimilation is found to differ across studies (Friedberg, 2002, Hu, 2000, Adsera & Chiswick, 2007). At the national level, Amuedo-Dorantes & De la Rica (2007) find that immigrants reduce the wage gap in 15 percentage points during their first 5 years of residence in Spain. In Lacuesta et al. (2009) the initial wage differential with respect to natives with the same observable characteristics decreases with time spent in Spain. According to their estimates, assimilation of legal immigrants occurs rapidly, with a reduction of around half of the initial wage gap during the first 5-6 years after arrival. By contrast, Fernandez & Ortega (2008) fail to find an improvement in the immigrants' labour market conditions, especially in terms of job stability. All in all, the native immigrant

wage gap partly reflects the fact that migrants cannot fully use their human capital attributes. While the focus of the present paper is not on earnings assimilation, it addresses the impact on wages of one of the most important investments in human capital in the host country: language ability. Immigrants who possess language proficiency are less likely to be overqualified in the job (Blázquez and Rendón, 2012), most likely because they are better placed to obtain information about job opportunities and earnings, and to communicate information about their skills to employers. Because it fulfills a number of functions, language plays a relevant role in the process of integration. The case of Spain is particularly interesting, due to the massive migration flows experienced over the last decade. The number of foreign workers present in Spain increased by 2,259,330 people during the period 2001-2008, a rise that amounts to 53% of the labour force increase for that period. As a result, immigrants now represent a significant segment of the country's work force.

The paper makes use of the Spanish National Immigrant Survey, a large-scale immigration survey released recently by the Spanish National Statistics Institute. The empirical strategy is based on a set of earnings equations extended to include a control for the immigrant's level of Spanish language proficiency. As OLS estimates of the language earnings effect may be biased due to measurement error and unobserved heterogeneity affecting language skills and earnings, the paper adopts an Instrumental Variable (IV) approach. The excluded instruments pass well conventional quality and validity tests and the results under alternative combinations of excluded instruments are remarkably robust.

An additional question addressed by the paper is whether the relationship between Spanish language proficiency and earnings differs between groups with different educational attainments. There are reasons to believe that language proficiency and schooling are complementary inputs of the earning-generating process. First, schooling has a large impact on life chances, social mobility and labour market opportunities. To the extent that an imperfect knowledge of the host language is likely to hamper such opportunities, stronger effects from language proficiency among the highly educated may be expected. Indeed, the available evidence suggests that individuals with better language knowledge are more likely to end up in jobs commensurate with their

qualifications (Blázquez and Rendón, 2012). Secondly, language skills are more likely to represent a valuable asset only in occupations that require higher levels of formal education. Thirdly, we cannot preclude the possibility that host language proficiency acts as a signal to employers about the quality of the individual's post-compulsory education. Such heterogeneous effects may have salient implications for the design of effective integration policies. A common trend in OECD countries is the prioritization of labour market integration and the strengthening of educational aspects, including language training (OECD, 2012). In line with this view, the Spanish Strategic Plan for Citizenship and Integration acknowledged during the immigration boom period the fact that immigration poses specific problems that must be tackled, such as "the promotion of improvements in immigrants' knowledge of the official languages and social norms in Spain, prerequisites for a cohesive society and for the very social integration of immigrants" (Ministry of Labour and Social Affairs, 2007). Unfortunately, the scope attributed to such policies may be more modest than presumed if workers with low qualifications fail to reap relevant returns from language training. The evidence collected so far is however scarce and suggestive of diverging degrees of complementarity between schooling and language skills among immigrants (Chiswick & Miller, 2003, Casale & Posel, 2011). This paper sheds further light on this issue by assessing the interplay between Spanish proficiency and formal academic qualifications in the Spanish labour market. In doing so, the analysis takes a step towards the identification of the benefits that different population groups may obtain from immigrant-oriented policies of language training.

The next section describes the dataset, the estimating sample and the Spanish language proficiency central question. Section 3 describes the IV approach adopted in the paper and discusses the choice of excluded instruments. Section 4 presents estimates of the impact of Spanish proficiency on earnings and separates them by education level. The results are paired with a variety of validity and relevancy tests to assess the quality of the instruments and the robustness of the IV estimates. Section 5 contains the concluding remarks.

2. Data and definition of variables

The data is taken from the Spanish National Immigrant Survey (ENI, *Encuesta Nacional de Inmigrantes*), a large-scale immigration survey carried out by the Spanish National Statistics Institute. The data collection was conducted between November 2006 and February 2007 and was based on the Municipal Census (*Padrón Municipal*). The original survey sample comprises approximately 15,500 individuals. The ENI provides detailed information on the socio-demographic characteristics of immigrants and their previous and current employment status. Immigrants are defined as any individual born abroad (regardless of their nationality) who at the time of being interviewed had reached at least 16 years of age and had resided in a home for at least a year or longer or had the intention to remain in Spain for at least a year.

The estimating sample consists of private sector men who are between 18 and 65 years old and work regularly between 15 and 70 hours a week. Self-employed individuals, as well as those whose main activity status is paid apprenticeship, training, and unpaid family workers have been excluded from the sample. The case of women is disregarded on account of the extra complication of potential selectivity bias. Dropping observations including missing values leaves us with a final sample of 3,089 observations.

2.1 Spanish proficiency

The central question on the ENI is

- *Thinking of what you need for communicating at work, at the bank, with the public authorities/administration. How well do you speak Spanish?*

with possible answers ranging from 1 ('very well') to 4 ('need to improve'). These were used to define *SP,* a dummy variable that takes value one if the immigrant has Spanish

proficiency (1-very well), zero otherwise. This recoding is intended to maintain the cardinality of the variable and to ease the interpretation of its coefficient[1].

There are some shortcomings inherent to this type of questions that are worth mentioning. As in most empirical work, we are hampered by data availability. First, due to its direct implications for immigrant assimilation in the host country the ENI focuses on Spanish proficiency. In doing so, it disregards other languages that may be important in the Spanish labour market (mainly English). There is therefore a possibility that observationally identical individuals benefit from diverging earnings profiles due to unobserved proficiencies in other languages. Secondly, unlike other surveys used in the literature (Chiswick & Miller, 1999, Hayfron, 2001) the ENI does not contain information on different language skills, including reading, writing and communicating in different settings and life situations. Indicators related to these elements would prove fruitful in providing a more accurate assessment on the causal relationship between complementary language skills and earnings. Thirdly, the results of this paper will be based on a self-assessed question. Despite the use of subjective evaluations is standard in the field due to the high costs of test-based assessments of language ability, there is a possibility that respondents have different perceptions under identical circumstances of how well they speak Spanish. These concerns notwithstanding, the question included in the ENI provides a unique opportunity to investigate meaningful relationships contained in the data and fits well the conventions of the literature.

Table 1 provides summary statistics by proficiency level. Nearly 67% of the sample reports being proficient in Spanish, whereas the remaining 33% has limited language skills. There are some relevant differences between the two sub-samples. Proficient immigrants earn 22.7% higher hourly wages (6.80 against 5.54 €/hour) than non-proficient immigrant. They also exhibit higher education qualifications (11.27 against 8.18 years of schooling), slightly shorter professional experiences (19.98 against 21.67 years) and a higher probability of having a permanent contract (55.35% against

---

[1] The paper follows a stringent criterion by considering only individuals who claim to be able to speak Spanish 'very well'. Results under the alternative classification 1-2 against 3-4 displayed slightly lower returns and are available upon request.

38.31%). In terms of marital status and household size the two groups are similar, though. Proficient immigrants are mainly original from North and Latin-America (55.94%), Central and Western Europe (24.76%), and are more likely to work in the Technology & Sciences (19.53%) and the Manufacturing & Construction sectors (16.09%). By contrast, non-proficient immigrants are concentrated in the Maghrebian (33.14%) and Eastern European (36.94%) communities. Their most common occupations are Agriculture & Fishery (36.35%) and the residual category Other activities (39.93%).

3. Estimation strategy

The earnings equation is specified as follows,

$$ln\ w_i = X_i\beta + \gamma SP_i + \varepsilon_i \qquad (1)$$

where w is hourly earnings, X includes completed education level, potential labour market experience and its square, type of contract (temporary or permanent), marital status (single, divorced or widowed, reference: married), number of children at home, previous unemployment spells of 3 months or longer in Spain (yes or no), legal status (legal or illegal), 7 occupational dummies, the immigrant's region of origin (Maghreb, Sub-Saharan Africa, Eastern Europe, North and Latin-America, Asia or Oceania, reference: Central and Western Europe) and 19 dummies for region of residence in Spain. The choice of these variables is duly motivated by the immigration adjustment literature[2]. There are however two variables that deserve an additional remark. The first one is occupation. It has been argued that language proficiency effects on earnings operate mainly thorough occupation so that the inclusion of occupation dummies in the regression is expected to provide a conservative estimate of the effect of Spanish proficiency on earnings (Aldashev et al., 2009, Chiswick & Miller, 2010)[3]. The second

---

[2] Other conventional controls such as tenure and work experience in the home and host country were disregarded due to large item non-response.

[3] Returns to language proficiency were between 2 and 6 percentage points higher without occupation dummies and are available upon request. Removal of occupation variables from the regression equation

variable is years since migration, an information typically used in immigrant earnings equations that is omitted in the present analysis. This choice is due to its potential endogeneity with earnings. To the extent that low-earning immigrants are more likely to leave the country, a positive coefficient on this variable would be the result of a selection process instead of the consequence of true labour market assimilation. Still, and given the tradition, we computed complementary results including years since migration as an additional regressor. The results showed little variation and, thus, are not reported here.

Finally, the crux of the present analysis will be on *SP*, Spanish language proficiency. A typical problem is that *SP* may depend on unobservable individual characteristics that are potentially related to the unobservable earnings determinants. That would be the case if, for example, more productive and capable individuals are more likely to achieve Spanish proficiency. If that is the case, the estimated coefficient would not reflect benefits from language skills, but merely a spurious correlation. Addressing this issue involves specifying a first stage equation for Spanish proficiency,

$$SP_i = X_i\delta + Z_i\theta + v_i \qquad (2)$$

where Z contains the set of excluded instruments. The use of IV is also intended to reduce the extent of attenuation bias that may stem from errors in the measurement of the individual (self-assessed) Spanish proficiency level SP[4].

---

decreased the adjusted R-squared by about 15% and increased the coefficients on individual-level variables, including schooling, type of contract and legal status.

[4] Self-reported measures of speaking fluency typically suffer from misclassification/measurement errors, with the probabilities of over-reporting being higher than the probabilities of under-reporting (Dustmann and Soest, 2001). Dustmann and Fabbri (2003) find that measurement error and endogenous choice bias the language coefficients in opposite directions. However, it is difficult to correct for misclassification using cross-sectional data, as in the present case. There is a need for longitudinal immigration data that is currently not available for this survey in Spain. The use of the IV technique in the present context should be seen as a working compromise to mitigate the effect of measurement errors inherent in a self-assessed measure of literacy (Charette and Meng, 1993).

Instruments must be *valid* (i.e., uncorrelated with earnings, $E(\varepsilon_i|Z_i)=0$) and *relevant* (i.e., they must account for a significant variation in *SP*). Earlier studies addressing the endogeneity problem leave us with a variety of potential candidates. Thus, for example, Chiswick & Miller (1995) rely on whether married overseas, age, number of children and a birthplace concentration variable when instrumenting English language proficiency. Almost identical instruments are used in Chiswick (1998) to analyze the impact of Hebrew language usage among immigrants in Israel although, as in his previous work, the validity of the selected variables is not tested. Rendón (2007) shows that variables capturing the externality effect of the community of residence, origin variables, years since migration, age of arrival and a variable indicating whether the individual was affected by Catalonia's Linguistic Normalization law are significantly related to Catalan proficiency. However, he warns that some of these instruments, particularly the region of origin, might be also related to labour market performance. Gao and Smith (2011) rely on child information (the number of children living in the host city and having at least one child enrolled in primary school in the host city) to proxy for the individual's proficiency in Mandarin language. Arguably, parents' exposure to communication with their children in the host country language and access to the children's superior pronunciation skills acts as a transmission mechanism. This road is also explored by Di Paolo (2011) and Di Paolo & Raymond (2011), who assume that language use with their children (as well as other variables including arrival at the host country before age of 10) affects parents' language proficiency but does not directly affect their earnings.

For the purposes of the present paper a good set of potential instruments for Spanish language proficiency were identified in the ENI. These are three dummy variables that capture whether the respondent i) arrived in Spain before age 10, ii) has a child who is proficient in the Spanish language and iii) plans to stay in Spain for the next 5 years. These exclusion restrictions are confirmed by the encompassing tests of validity and relevancy reported in the next section.

4. Results

The OLS estimates of the earnings equation are reported in the first column of Table 2. According to the results, the impact of Spanish proficiency upon wages, γ, amounts to 4.8 percentage points (pp). However, before discussing the reliability and robustness of the coefficient to changes in the estimating strategy, it is illustrative to document the role of the remaining covariates included in the earnings equation.

The results are as follows. An additional year of schooling raises earnings by about 1.1 pp, a figure that is almost one order of magnitude below conventional estimates reported for the total Spanish population (see Budría, 2005, for a review). This result may well reflect the extent of labour market discrimination against migrant workers, according to which immigrants end up in low pay jobs that are not commensurate with their qualifications. As expected, professional experience is associated with higher earnings, though at a decreasing rate, whereas having a permanent contract and previous unemployment experience are associated with wages being about 5.5 pp higher and lower, respectively. There are conspicuous earnings differentials among immigrants from different regions of origin. Relative to the reference individual (an immigrant from Central-Western Europe), workers from Maghreb, Sub-Saharan Africa, Eastern Europe, America and Asia reap significantly lower earnings. The predicted pay penalty is between 9.6 (Asia) and 17.9 (North and South-America) pp. Finally, the results suggest roughly 40-50 pp higher earnings in the Management and Technology & Sciences sector, whereas Administration, Agriculture & Fishery and Manufacturing & Construction carry a lower (about 15 pp) despite statistically significant premium, relative to the reference category 'Other sectors'.

The other columns of Table 2 report the IV results emerging from alternative combinations of the instruments. As the role of the remaining covariates has been already documented and presents little variation, we move on to focus on the analysis of the impact of Spanish proficiency on wages, γ. Table 3 summarizes the point estimates and reports a variety of tests for the quality of the instruments. The IV results suggest that assuming exogenous *SP* yields a downward-biased prediction. Column (i) is based

on the full set of instruments and should be considered the benchmark specification, for it delivers the most favorable tests for the quality of the instruments (see below). In this case, Spanish proficiency is associated with a wage increase of 27.3 pp. This figure is quite robust to alternative specifications. In column (ii) the instrument with the lowest-despite significant explanatory power upon *SP* (whether the individual plans to stay in Spain for the next 5 years) has been dropped. This results in an almost negligible variation in the estimated return to language proficiency (27.5 pp). Column (iii) proceeds likewise and drops the instrument with the second lowest explanatory power (children's Spanish proficiency). This results in a modest decrease of the coefficient (23.0 pp) and an increase of its standard deviation. The loss of precision is however modest and the resulting effect is significant at conventional statistical levels. The just identified model that appears in column (iv) delivers an almost identical prediction (23.2 pp).

While these estimates seem high, they are towards the lower range of estimates reported by studies using a similar IV approach for language fluency in other countries. Chiswick and Miller (1995) encounter returns to English proficiency of between 40 and 57 pp among immigrants in Canada and the US, whereas Chiswick and Miller (2003) report estimates between 26 and 42 pp in Canada. Chiswick (1998) reports a figure above 35 pp for fluency in Hebrew among migrants in Israel. In Gao and Smyth (2011) the object of study is the return to standard Mandarin among internal migrants in China. In this case, the estimated impact amounts to some 40 pp. The results reported in Table 3 suggest that the earning gains from host language proficiency are large but relatively lower in Spain. Nevertheless they appear to be sensitively higher than the returns to proficiency in regional languages in Spain, for in Di Paolo & Raymond (2012) the return to Catalan proficiency is about 16 pp among immigrants in Catalonia.

An important concern with IV is the validity of the instruments. Validity is not assured if the excluded instruments have direct effects on earnings beyond those flowing indirectly through Spanish proficiency. This problem may yield biased estimates and will be exacerbated by a weak correlation between the endogenous variable and the instruments (Bound et al., 1995). To assess these issues, Table 3 reports various relevant

tests. In all cases the Hansen-Sargan test of orthogonality does not lead to rejection of the null hypothesis. This suggests that the selected instruments are uncorrelated to the earnings variable[5]. As for the weak IV problem, two diagnosis tests are reported: the F-test for the joint significance of the selected instruments and their relative contribution to R^2 in the *SP* equation. If the excluding restrictions are weak, i.e. have little explanatory power, then the bias in the estimated coefficients is expected to increase. However, the instruments are jointly significant in all columns and the F-statistics, above 45, are readily in the safe zone suggested by Stock et al. (2002). Therefore, the weak instruments hypothesis must be rejected. The R^2 of the first stage equation is remarkably high (> .50), with the contribution of the excluded instruments to this statistic decreasing from .0595 in the benchmark case to .0157 in column (v). The latter figure is only slightly above the lower range of values considered admissible in the literature. This inferior instrument quality arguably determines the artificially high return to Spanish proficiency found in the last column (39.9 pp).

Finally, the necessity of resorting to IV should be statistically assessed. Since the IV estimator always has larger asymptotic variance than the OLS estimator, there is loss of efficiency in the later approach. The Durbin-Wu-Hausman statistic reported in the table tests for the exogeneity of the instrumented variable. A rejection of the null hypothesis implies that the OLS estimator does not yield consistent estimates due to the endogeneity of *SP* and, therefore, the IV estimator should be used. The test rejects the null hypothesis and, thus, justifies the choice of the IV approach.

As a robustness check, Panel 2 reports complementary results where immigrants whose mother tongue is Spanish have been dropped ('Restricted sample'). Despite the inclusion of explicit controls for the immigrants' region of origin in the earnings equation, unobserved heterogeneity related to *SP* may persist due to immigrants' diverging language, cultural and social backgrounds. Interestingly, restricting the sample to individuals whose mother tongue is not Spanish results in only slight variations. The benchmark model yields a point estimate of about 23.0 pp and, due to

---

[5] Validity cannot be tested in the just-identified models of columns (iv) and (v).

smaller sample size, the estimates are now less stable. In two cases, (iii) and (iv), the coefficients on γ fail to be statistically significant. These two specifications should be the less preferred, though, according to the partial R^2 statistic ($< .03$). Nervertheless, combinations of instruments with a higher explanatory power (partial $R^2 > .03$) as the ones displayed in (ii) and (v) yield significant estimates.

4.1 The determinants of Spanish proficiency

To provide complementary evidence on the quality of the instruments and to shed some light on the determinants of *SP*, Table 4 reports the results of the first stage equation. Arrival before age 10 is a strong predictor of language ability, with immigrants in this group being some 27 pp more likely to be proficient in Spanish. The impact of having a child with good language skills is about one third as high (11 pp) but still statistically significant, whereas planning to stay in the country attracts a slightly positive coefficient (around 3 pp). These variables are significant at conventional levels across specifications.

Inspection of the table suggests that language skills are higher among more educated individuals. This observation is consistent with the bunch of the evidence reported in related research and probably stems from the close relationship between the individual's learning ability and the individual's level of schooling. A complementary explanation is that more educated individuals have strong economic incentives to language learning due to higher market wages and employment rates (Chiswick & Miller, 2003). Language skills depend negatively on total labor market experience, though at a decreasing rate. The direction of this effect is probably due to (ceteris paribus) lower language skills among older immigrants and the close relationship between professional experience and age. Having a permanent contract and previous unemployment experience in Spain are both positively associated with language proficiency. These effects match expectations. It is very unlikely that immigrants with poor language skills are offered a permanent position, whereas most immigrants without an unemployment history in the Spanish labour market are either recent immigrants or individuals entering

the labour force after periods of inactivity. Spells of inactivity are expected to speed up human capital depreciation, and to reduce the extent of social interactions leading to language learning. A similar reasoning, as well as a reduced bargaining power leading to a less favored labour market position, applies to illegal immigrants. Moreover, there are systematic differences in the degree of fluency by region of origin. Migrants from North and Latin-America and Oceania are more likely to be language proficient, due to the existence of large Spanish-speaking communities in these continents (Ecuador, Colombia, Argentina and Philippines, mainly). Reversely, fluency rates are lower ceteris paribus among immigrants from Maghreb, Sub-Saharan Africa, Eastern Europe and Asia. Finally, there are significant differences between occupation sectors, with workers in the Management, Technology & Sciences, and Manufacturing & Construction occupations being more likely to display proper language skills.

4.2 The complementarity between language proficiency and education

This section explores whether the relationship between Spanish language proficiency and earnings differs between education groups. The suspicion that certain deficits in academic qualifications may importantly hamper the returns to language skills motivates the analysis.

Table 5 reports the returns to Spanish proficiency by level of education. To avoid problems from small cell-size, only two education categories are considered. Panel 1 focuses on the sample of workers with completed upper secondary or higher education, whereas Panel 2 is devoted to immigrants with less than upper secondary education. Splitting the sample instead of estimating a common equation with a language proficiency-schooling interaction is intended to allow for a different endogeneity and earnings-determination process within the two groups. The results are strongly supportive of large returns to Spanish language proficiency among migrants with an upper secondary or higher education. According to the benchmark specification in Panel 1, highly educated individuals that possess Spanish language proficiency are expected to earn, ceteris paribus, 49.3 pp more than highly educated individuals with limited language skills. This figure almost doubles the 27.3 pp obtained previously from the

total sample, although it decreases slightly, to about 43 pp, in columns (iii) to (v). It is interesting to note that the reduction in sample size does not come at the cost of unstable results under alternative sets of instruments, as the variation across columns in the estimated effect is relatively modest. The tests for the excluded instruments suggest, again, inferior instrument quality as we move from left to right on the table, but they are broadly satisfactory. Panel 2 shows the results for the sample of workers that acquired less than upper secondary education. The resulting figures are almost 30 pp below the return obtained for highly educated individuals[6]. Moreover, in one case (column (iv)) the language effect fails to be statistically significant.

All in all, the results are partially supportive of earlier complementarity effects between education and language proficiency reported in the literature. Chiswick & Miller (2003) find higher returns to education among English proficient immigrants in Canada. In Di Paolo & Raymond (2012) highly educated workers earn a return from Catalan proficiency of about 25 pp, whereas workers with less than 9 years of education (presumably, less than secondary education) reap no return. Despite the fact that differences between studies must be interpreted cautiously, our results suggest that attaining Spanish language proficiency is a profitable investment for less educated immigrants as well. Casale & Posel (2011) examine the relationship between English language proficiency and earnings in South Africa. They find a very large additional wage premium for employed African men who are both proficient in English and have attained either a diploma or academic degree. Specifically, English proficiency together with the completion of some form of post-secondary qualification offers a return of 93 pp, against the 33 pp figure reaped by less educated, proficient workers. Despite the fact that sociological, economic and labour market differences between the two countries hamper any thorough comparisons, the results reported in the present study are suggestive of lower despite substantial complementarity effects in the Spanish labour market.

---

[6] The statistical significance of the cross-model difference was tested using the STATA postestimation command *suest* (p-value = 0.000).

5. Conclusions

This paper examined the impact of Spanish language proficiency upon earnings among immigrants in Spain. In doing so it adds to the scarce evidence collected among immigrants in a non-English speaking country. The recent and intense nature of immigration in Spain makes it a unique context for analysis: most immigrants arrived in the last decade, which makes them a more homogenous group and reduces the problems that arise from cohort effects.

The analysis was based on recent data from the Spanish National Immigrant Survey, and derives from an IV approach supported by good quality instruments. Being proficient in Spanish language raises immigrant earnings by about 27 pp. This estimate is reasonably robust to the combination of alternative instruments and falls to 23 pp when immigrants whose mother language is Spanish are excluded from the sample. These figures are similar to Bleakley & Chin's (2004) results for the return to English language proficiency among immigrants in the US, and considerably above the 13-18 pp range reported by Di Paolo (2011) and Di Paolo & Raymond (2012) for Catalan proficiency in the Spanish region of Catalonia. It is in our current research agenda to use a common specification and restrict our attention to Spain's bilingual regions in order to assess whether the returns to Spanish language proficiency are actually higher than the returns to regional languages in bilingual regions.

All in all, the results are important to enable policy makers to devise strategies and immigration policies that promote and guarantee economic and social stability. It would be advisable to provide language courses for immigrants upon arrival. Still, there are profound differences between education groups in terms of the language earnings premium (about 49 pp among the high-educated against 21 pp among the low-educated). This finding warns that indiscriminate language training policies may enhance economic inequalities among immigrants. In this respect, the provision of complementary schemes among the less qualified, including affordable education, training programs and instruction related to the Spanish labour market and legislation in order to ensure familiarity with work-related terms and usages might prove beneficial.

Because of the cross-sectional nature of our data, it is sensible to emphasize the need for some caution in the interpretation of the findings. The patterns outlined in the present paper deserve scrutiny in longitudinal data. Controlling for individual time-invariant effects will allow us to establish more accurate causal relationships by following people over time and investigating how the earnings of the *same* people change when they improve their language skills. There seem grounds here for the funding of an immigrant-based panel data survey in Spain to explore, among other things, the consequences for earnings of diverging levels of Spanish language proficiency.

References

Adsera, A. and Chiswick, B. (2007). 'Are there gender and country of origin differences in immigrant labor market outcomes across European destinations?', *Journal of Population Economics,* 20(3), pp.495-526.

Aldashev, A., Gernandt, J., and Thomsen, S. (2009). 'Language usage, participation, employment and earnings: evidence for foreigners in West Germany with multiple sources of selection', *Labour Economics,* 16(3), pp. 330-341.

Amuedo-Dorantes, C. and de la Rica, S. (2007). 'Labor market assimilation of recent immigrants in Spain', *British Journal of Industrial Relations,* 45(2), pp. 257-284.

Blázquez, M. and S. Rendón (2012), 'Over-education in multilingual economies: Evidence from Catalonia', forthcoming in *International Migration.*

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

Bound, J., Jaeger, D., and Baker, R. (1995). 'Problems with instrumental variable estimationwhen the correlation between instruments and the endogenous explanatory variables is weak', *Journal of the American Statistical Association,* 90(430), pp. 443-450.

Carliner, G. (1981). 'Wage differences by language group and the market for language skills in Canada', *The Journal of Human Resources,* 16(3), pp.384-399.

Charrete, M. and Meng, R. (1994). 'Explaining language proficiency. Objective versus self-assessed measures of literacy', *Economics Letters,* 44, pp. 313-321.

Chiswick, B. (1998). 'Hebrew language usage: Determinants and effects on earnings among immigrants in Israel', *Journal of Population Economics,* 11, pp. 253-271.

Chiswick, B. and Miller, P. (1995). 'The Endogeneity between language and earnings: international analyses', *Journal of Labor Economics*, 13(2), pp.246-288.

Chiswick, B. and Miller, P. (2003). 'The complementarity of language and other human capital: immigrant earnings in Canada', *Economics of Education Review*, 22(5), pp.469–480.

Chiswick, B. and Miller, P. (2010). 'Occupational language requirements and the value of English in the US labor market', *Journal of Population Economics,* 23, pp. 353-372.

Casale, D. and Posel, D. (2011). 'English language proficiency and earnings in a developing country: the case of South Africa', *Journal of Socio-Economics,*40, pp. 385-393.

Di Paolo, A. (2011). 'Knowledge of Catalan, public/private sector choice and earnings:evidence from a double sample selection model', *Hacienda Pública Española,* 197-(2), pp. 9-35.

Di Paolo, A. and Raymond, J.L. (2012). 'Language knowledge and earnings in Catalonia', *Journal of Applied Economics,* 15(1), pp. 89-118.

Dustmann, C, and van Soest, A. (2001). 'Language fluency and earnings: estimation with misclassified language indicators', *The Review of Economics and Statistics*, 83(4), pp. 663-674.

Dustmann, C. and Fabbri, F. (2003). 'Language proficiency and labor market performance of immigrants in the UK', *The Economic Journal,* 113, pp. 695-717.

Fernández, C. and Ortega, C. (2008). 'Labor market assimilation of immigrants in Spain: employment at the expense of bad job-matches?', *Spanish Economic Review,* 10, pp. 83-107

Friedberg, R. (2000). 'You can't take it with you? Immigrant assimilation and the portability of human capital', *Journal of Labor Economics,* 18(2), pp. 221-251.

Gao W. and Smyth, R. (2011). 'Economic returns to speaking "standard Mandarin" among migrants in China's urban labor market', E*conomics of Education Review,* 30(2), pp. 342-352

Grenier, G. (1984). 'The Effects of language characteristics on the wages of Hispanic-American males', *The Journal of Human Resources,* 19, pp. 33-52.

Hu, W-Y.(2000). 'Immigrant earnings assimilation: estimates from longitudinal data. *American Economic Review, Papers and Proc.* 90 (May): 368–72.

Lacuesta, A., Izquierdo, M., and Vegas, R. (2009). 'Assimilation of immigrants in Spain: A longitudinal analysis', *Labour Economics*, 6(6), pp. 669-678.

Lui, H.K. (2007). 'The returns to language ability in Hong Kong: before and after the handover', *Applied Economic Letters,* 14(2), pp. 121-125.

McManus, W., Gould, W., and Welch, F.(1983). 'Earnings of Hispanic men: the role of English language proficiency', *Journal of Labor Economics,* 1(2), pp.101-130.

Ministry of Labour and Social Affairs (*Ministerio de Trabajo y Asuntos Sociales*) (2007). 'Strategic plan for citizenship and integration', *Subdirección General de Información Administración y Publicaciones.* Available at:
http://extranjeros.empleo.gob.es/es/IntegracionRetorno/Plan_estrategico/pdf/PECIingles.pdf

OECD (2012). 'Migration policy developments, *International Migration Outlook 2012, OECD Publishing.* Available at: http://dx.doi.org/10.1787/migr_outlook-2012-6-en

Rendon, S. (2007). 'The Catalan premium: language and employment in Catalonia', *Journal of Population Economics,* 20, pp. 669-686.

Rivera-Batiz, F. (1992). 'English proficiency and the earnings of young immigrants in the U.S. Labor Markets', *Review of Policy Research,* 11(2), pp.165.175

Stock, J.H., Wright, J. and Yogo, M. (2002). 'A survey of weak instruments and weak identification in generalized method of moments'*, Journal of Business & Economic Statistics, American Statistical Association,* 20(4), pp. 518-529.

Tables

Table 1. Summary statistics by Spanish proficiency

	Proficient	Non-proficient
Share	.6679	.3321
Hourly wage	6.7974	5.5397
	(3.3343)	(2.5986)
Years of scholing	11.2672	8.1773
	(3.5800)	(4.9949)
Experience	19.981	21.666
	(9.6261)	(9.9128)
Permanent contract	.5535	.3831
	(.4972)	(.4864)
Single	.3601	.3363
	(.4801)	(.4726)
Divorced	.0641	.0799
	(.2447)	(.2713)
Married	.5759	.5838
	(.4943)	(.4931)
Children	.6403	.6335
	(.4800)	(.4820)
Previous unemployment experience	.3306	.3091
	(.4705)	(.4622)
Ilegal status	.3107	.3062
	(.4629)	(.4610)
Region of origin		
Maghreb	.0868	.3314
	(.2815)	(.4709)
Sub-saharan Africa	.0194	.0799
	(.1379)	(.2713)
Eastern Europe	.0751	.3694
	(.2637)	(.4828)
America	.5594	.0331
	(.4966)	(.1790)
Asia	.0107	.0556
	(.1027)	(.2291)
Oceania	.0011	.0001
	(.0311)	(.0041)
Central and western Europe	.2476	.1306
	(.4490)	(.3113)
Occupation sector		
Army	.0019	.0001
	(.0440)	(.0312)
Management	.0456	.0195
	(.2085)	(.1383)
Technology and Sciences	.1953	.0382
	(.3965)	(.1913)
Services	.1386	.1014
	(.3456)	(.3019)
Administration	.0456	.0127
	(.2086)	(.1119)
Agriculture and Fishery	.2123	.3635
	(.4090)	(.4812)
Manugfacturing, Construction	.1609	.0653
	(.3675)	(.2471)
Others	.1997	.3993
	(.2999)	(.4577)

Note to Table 1: i) Source: Spanish National Immigrant Survey; ii) Standard deviations are in parenthesis.

Table 2. Earnings equations: OLS and IV estimates

	OLS	(i) Benchmark	(ii)	(iii)	(iv)	(v)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Spanish language proficiency	.0478 ***	.2732 ***	.2746 ***	.2304 **	.2316 **	.3986 ***
	(.0165)	(.0799)	(.0799)	(.0951)	(.0951)	(.1529)
Years of scholing	.0114 ***	.0064 *	.0064 **	.0073 ***	.0073 ***	.0036
	(.0017)	(.0025)	(.0025)	(.0027)	(.0027)	(.0039)
Experience	.0083 ***	.0089 ***	.0089 ***	.0088 ***	.0088 ***	.0092 ***
	(.0022)	(.0026)	(.0026)	(.0025)	(.0025)	(.0027)
Experience $(x1000)^2$	-.1214 ***	-.1366 ***	-.1367 ***	-.1331 ***	-.1338 ***	-.1451 ***
	(.0448)	(.0534)	(.0531)	(.0529)	(.0529)	(.0564)
Permanent contract	.0553 ***	.0433 ***	.0433 ***	.0455 ***	.0456 ***	.0367 **
	(.0121)	(.0139)	(.0139)	(.0142)	(.0142)	(.0156)
Single	-.0181	-.0166	-.0166	-.0169	-.0169	-.0157
	(.0152)	(.0160)	(.0160)	(.0159)	(.0159)	(.0167)
Divorced	-.0533 **	-.0287	-.0286	-.0333	-.0334	-.0150
	(.0239)	(.0267)	(.0268)	(.0269)	(.0269)	(.0319)
Children	.0149	.0111	.0111	.0118	.0118	.0090
	(.0152)	(.0166)	(.0166)	(.0164)	(.0164)	(.0174)
Previous unemployment experience	-.0547 ***	-.0670 ***	-.0671 ***	-.0647 ***	-.0647 ***	-.0738 ***
	(.0123)	(.0137)	(.0137)	(.0137)	(.0137)	(.0162)
Ilegal status	.0080	.0096	.0096	.0093	.0093	.0104
	(.0136)	(.0139)	(.0139)	(.0138)	(.0138)	(.0144)
Region of origin						
Maghreb	-.1526 ***	-.0789 **	-.0785 **	-.0925 **	-.0929 **	-.0380
	(.0225)	(.0352)	(.0352)	(.0389)	(.0389)	(.0557)
Sub-saharan Africa	-.1484 ***	-.0666	-.0661	-.0817 *	-.0822 *	-.0211
	(.0326)	(.0439)	(.0439)	(.0470)	(.0470)	(-.0668)
Eastern Europe	-.1408 ***	-.0447	-.0441	-.0624	-.0629	-.0087
	(.0225)	(.0411)	(.0411)	(.0460)	(.0460)	(.0703)
America	-.1792 ***	-.2256 ***	-.2259 ***	-.2170 ***	-.2168 ***	-.2514 ***
	(.0180)	(.0260)	(.0260)	(.0280)	(.0280)	(.0378)
Asia	-.0955 **	.0103	.0110	-.0091	-.0097	.0692
	(.0387)	(.0695)	(.0695)	(.0708)	(.0708)	(.0993)
Oceania	.2953	.2230 ***	.2225 ***	.2363 ***	.2367 ***	.1827 **
	(.2187)	(.0762)	(.0762)	(.0773)	(.0773)	(.0890)
Occupation sector						
Army	.2441 *	.2433	.2433	.2435	.2435	.2428
	(.1379)	(.1657)	(.1658)	(.1646)	(.1646)	(.1697)
Management	.3912 ***	.3663 ***	.3662 ***	.3709 ***	.3711 ***	.3525 ***
	(.0324)	(.0552)	(.0552)	(.0553)	(.0553)	(.0578)
Technology and Sciences	.4902 ***	.4755 ***	.4754 ***	.4782 ***	.4783 ***	.4674 ***
	(.0222)	(.0270)	(.0270)	(.0268)	(.0268)	(.0298)
Services	-.0151	-.0187	-.0188	-.0181	-.0180	-.0208
	(.0199)	(.0199)	(.0200)	(.0198)	(.02198)	(.0208)
Administration	.1323 ***	.1188 ***	.1287 ***	.1213 ***	.1213 ***	.1112 ***
	(.0331)	(.0312)	(.0312)	(.0311)	(.0311)	(.0333)
Agriculture and Fishery	.1544 ***	.1555 ***	.1555 ***	.1553 ***	.1553 ***	.1561 ***
	(.0159)	(.0153)	(.0153)	(.0152)	(.0152)	(.0160)
Manugfacturing, Construction	.1600 ***	.1262 ***	.1260 ***	.1325 ***	.1326 ***	.1074 ***
	(.0208)	(.0227)	(.0227)	(.0235)	(.0235)	(.0312)
Constant	1.4770 ***	1.3688 ***	1.3682 ***	1.3888 ***	1.3894 ***	1.3086 ***
	(.0455)	(.0626)	(.0626)	(.0670)	(.0670)	(.0898)
Controls for the 19 Spanish Communities	Yes	Yes	Yes	Yes	Yes	Yes
R^2	.3944	.3575	.3571	.3699	.3702	.3050
No. of observations	3,089	3,089	3,089	3,089	3,089	3,089

Table 3. The impact of Spanish language ability on hourly earnings

	OLS	IV				
		(i) Benchmark	(ii)	(iii)	(iv)	(v)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Panel 1. Full sample						
Spanish language proficiency	.0478 ***	.2732 ***	.2746 ***	.2304 **	.2316 **	.3986 **
	(.0165)	(.0799)	(.0799)	(.0951)	(.0951)	(.1529)
Instruments						
Arrived before age 10	--	Yes	Yes	Yes	Yes	no
Child proficient in Spanish	--	Yes	Yes	No	No	Yes
plans to stay in Spain	--	Yes	No	Yes	No	No
R^2	.3944	.3575	.3571	.3699	.3702	.3051
Sargan-Hansen χ^2 test	--	1.15	.88	.29	--	--
F-test	--	63.14 ***	94.80 ***	77.37 ***	154.89 ***	48.71 ***
R^2 in first stage equation	--	.5327	.5327	.5253	.5253	.5111
Partial R^2		.0595	.0594	.0446	.0446	.0157
Durbin-Wu-Hausman test	--	8.77 ***	8.87 ***	3.99 **	4.04 **	5.82 ***
No. of observations	3,089	3,089	3,089	3,089	3,089	3,089
Panel 2. Restricted sample						
Spanish language proficiency	.0639 ***	.2297 **	.2286 **	.1668	.1644	.2802 **
	(.0178)	(.0862)	(.0867)	(.1328)	(.1337)	(.1220)
Instruments						
Arrived before age 10	--	Yes	Yes	Yes	Yes	no
Child proficient in Spanish	--	Yes	Yes	No	No	Yes
plans to stay in Spain	--	Yes	No	Yes	No	No
R^2	.4749	.4481	.4484	.4645	.4651	.4292
Sargan-Hansen χ^2 test	--	.42	.38	.07	--	--
F-test	--	31.90 ***	47.54 ***	28.21 ***	55.85 ***	46.85 ***
R^2 in first stage equation		.4481	.4016	.3843	.3841	.3878
Partial R^2		.0526	.0523	.0248	.0246	.0304
Durbin-Wu-Hausman test	--	3.90 **	3.81 **	0.63	0.59	3.34 *
No. of observations	1,739	1,739	1,739	1,739	1,739	1,739

Table 4. The determinants of Spanish language ability

	(i) Benchmark	(ii)	(iii)	(iv)	(v)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Arrived before age 10	.2686***	.2674***	.2709***	.2701***	
	(.0218)	(.0216)	(.0218)	(.0217)	
Child proficient in Spanish	.1117***	.1115***			.1148***
	(.0160)	(.0160)			(.0164)
Plans to stay in Spain	.0278*		.0340**		
	(.0173)		(.0173)		
Years of scholing	.0199***	.0199***	.0213***	.0213***	.0206***
	(.0019)	(.0019)	(.0019)	(.0019)	(.0019)
Experience	-.0060**	-.0060**	-.0045*	-.0046*	-.0039
	(.0026)	(.0026)	(.0026)	(.0026)	(.0026)
Experience $(x1000)^2$	.0530**	.1241**	.1103*	.1105*	-.0820
	(.1237)	(.0530)	(.0528)	(.0528)	(.0535)
Permanent contract	.0363***	.0361***	.0398***	.0397***	.0491***
	(.0136)	(.0135)	(.0137)	(.0137)	(.0138)
Single	-.0035	-.0030	-.0131	-.0128	.0032
	(.0179)	(.0178)	(.0178)	(.0177)	(.0180)
Divorced	-.0644**	-.0646**	-.1017***	-.1017***	-.0710**
	(.0272)	(.0272)	(.0267)	(.0267)	(.0284)
Children	-.0199	-.0202	.0183	.0179	-.0227
	(.0193)	(.0193)	(.0184)	(.0184)	(.0197)
Previous unemployment experience	.0381***	.0380***	.0459***	.0458***	.0461***
	(.0132)	(.0132)	(.0133)	(.0133)	(.0134)
Ilegal status	-.0329**	-.0331**	-.0296**	-.0297**	-.0106
	(.0146)	(.0146)	(.0147)	(.0147)	(.0147)
Region of origin					
Maghreb	-.2321***	-.2325***	-.2399***	-.2402***	-.3178**
	(.0302)	(.0302)	(.0304)	(.0304)	(.0289)
Sub-saharan Africa	-.2556***	-.2557***	-.2742***	-.2743***	-.3427**
	(.0431)	(.0431)	(.0437)	(.0437)	(.0427)
Eastern Europe	-.3066***	-.3066***	-.3204***	-.3204***	-.4112**
	(.0312)	(.0312)	(.0312)	(.0312)	(.0286)
America	.3055***	.3055***	.3008***	.3009***	.2116***
	(.0231)	(.0231)	(.0232)	(.0232)	(.0201)
Asia	-.3661***	-.3664***	-.3874***	-.3876***	-.4470**
	(.0519)	(.0519)	(.0516)	(.0516)	(.0514)
Oceania	.2029***	.2037***	.1831***	.1837***	.3401***
	(.0725)	(.0725)	(.0421)	(.0422)	(.0979)
Occupation sector					
Army	-.0094	-.0110	-.0038	-.0049	-.0026
	(.1218)	(.1218)	(.1266)	(.1266)	(.1029)
Management	.0942**	.0936**	.1059***	.1056***	.0981**
	(.0389)	(.0388)	(.0393)	(.0392)	(.0397)
Technology and Sciences	.0760***	.0758***	.0828***	.0827***	.0581**
	(.0248)	(.0248)	(.0249)	(.0249)	(.0250)
Services	.0101	.0101	.0145	.0144	.0118
	(.0219)	(.0219)	(.0219)	(.0219)	(.0224)
Administration	.0465	.0466	.0478	.0478	.0585*
	(.0324)	(.0325)	(.0322)	(.0322)	(.0327)
Agriculture and Fishery	-.0008	-.0012	-.0024	-.0026	-.0032
	(.0177)	(.0177)	(.0179)	(.0179)	(.0181)
Manugfacturing, Construction	.1121***	.1115***	.1288***	.1285***	.1324***
	(.0225)	(.0225)	(.0226)	(.0225)	(.0231)
Constant	.4507***	.4443***	.4223***	.4179***	.5064***
	(.0543)	(.0528)	(.0540)	(.0526)	(.0532)
R^2	.5327	.5327	.5253	.5253	.5111
No. of observations	3,089	3,089	3,089	3,089	3,089

Table 5. The impact of Spanish language ability on hourly earnings, by education groups

| | OLS | IV | | | | |
| | | (i) Benchmark | (ii) | (iii) | (iv) | (v) |
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Panel 1. Workers with upper sec. educ. or more						
Spanish language proficiency	.0622 ***	.4927 ***	.4951 ***	.4241 **	.4264 **	.4314 **
	(.0230)	(.1746)	(.1747)	(.2026)	(.2036)	(.2145)
Instruments						
Arrived before age 10	--	Yes	Yes	Yes	Yes	no
Child proficient in Spanish	--	Yes	Yes	No	No	Yes
plans to stay in Spain	--	Yes	No	Yes	No	No
R^2	.4592	.2931	.2915	.2728	.2713	.3292
Sargan-Hansen χ^2 test	--	.41	.07	.33	--	--
F-test	--	23.11 ***	34.69 ***	29.14 ***	58.34 ***	16.72 ***
R^2 in first stage equation	--	.4422	.4422	.4367	.4367	.4309
Partial R^2		.0295	.0295	.0199	.0199	.0099
Durbin-Wu-Hausman test	--	10.94 ***	11.04 ***	7.32 **	7.38 **	3.48 ***
No. of observations	1,760	1,760	1,760	1,760	1,760	1,760
Panel 2. Workers wih less than upper sec. educ.						
Spanish language proficiency	.0264	.2143 **	.2125 **	.1609 *	.1581	.6199 **
	(.0238)	(.0816)	(.0819)	(.0867)	(.0871)	(.2889)
Instruments						
Arrived before age 10	--	Yes	Yes	Yes	Yes	no
Child proficient in Spanish	--	Yes	Yes	No	No	Yes
plans to stay in Spain	--	Yes	No	Yes	No	No
R^2	.2532	.2173	.2181	.2348	.2356	.4292
Sargan-Hansen χ^2 test	--	3.13	2.90 *	.38	--	--
F-test	--	35.26 ***	52.39 ***	45.40 ***	90.00 ***	15.24 ***
R^2 in first stage equation		.6359	.6358	.6312	.6311	.5987
Partial R^2	--	.1044	.1042	.0931	.0926	.0131
Durbin-Wu-Hausman test	--	6.37 **	6.23 **	2.85 *	2.71 *	5.64 *
No. of observations	1,329	1,329	1,329	1,329	1,329	1,329