

WORKING PAPER SERIES

CEEApIA WP No. 17/2010

**The Determinants of Spanish Language
Proficiency among Immigrants in Spain**

**Santiago Budría
Pablo Swedberg**

December 2010

The Determinants of Spanish Language Proficiency among Immigrants in Spain

Santiago Budría

Universidade da Madeira (DGE e CEEApIA)

Pablo Swedberg

Department of Business and Social Sciences, St. Louis University

Working Paper n.º 17/2010
Dezembro de 2010

RESUMO/ABSTRACT

The Determinants of Spanish Language Proficiency among Immigrants in Spain

This article uses micro-data from the Spanish National Immigrant Survey (Encuesta Nacional de Inmigrantes-ENI in Spanish) carried out in 2007 among immigrants in Spain. In recent years, Spain has received unprecedented immigration flows. A substantial number of immigrants cannot communicate adequately in the language of the country to which they immigrate. Among the multiple reasons for the lack of host language proficiency one can distinguish factors such as a low level of educational attainment, not having been provided with adequate opportunities to learn the host language, living in ethnic enclaves or having arrived at an older age. Language skills (including oral and written ability) play a crucial role in the determination of the immigrants' social and economic integration in the host country. As a consequence, analyzing the source of foreign language acquisition is crucial for understanding the immigrants' economic, social and political involvement. The results show that an increase in educational attainment is associated with a higher level of Spanish spoken proficiency. Language ability is also associated with the country or region of origin. The results show that immigrant men and women from the Maghreb and Asia, as well as men from Eastern Europe and Sub Saharan Africa show a significantly weaker command over spoken Spanish than Western Europeans.

JEL classification: F22, J15, J24, J40

Keywords: Immigration, Language, Education

Santiago Budría
Departamento de Gestão e Economia
Universidade da Madeira
Edifício da Penteada
Caminho da Penteada
9000 - 390 Funchal

Pablo Swedberg
Department of Business and Social Sciences
St. Louis University
Madrid Campus
Avenida del Valle, 34
28003 Madrid
Spain

The determinants of Spanish language proficiency among immigrants in Spain

Santiago Budría*, Pablo Swedberg**

*CEEApIA and Department of Economics, University of Madeira

**Department of Business and Social Sciences, St. Louis University (Madrid Campus)

December 2010

Abstract

This article uses micro-data from the Spanish National Immigrant Survey (Encuesta Nacional de Inmigrantes-ENI in Spanish) carried out in 2007 among immigrants in Spain. In recent years, Spain has received unprecedented immigration flows. A substantial number of immigrants cannot communicate adequately in the language of the country to which they immigrate. Among the multiple reasons for the lack of host language proficiency one can distinguish factors such as a low level of educational attainment, not having been provided with adequate opportunities to learn the host language, living in ethnic enclaves or having arrived at an older age. Language skills (including oral and written ability) play a crucial role in the determination of the immigrants' social and economic integration in the host country. As a consequence, analyzing the source of foreign language acquisition is crucial for understanding the immigrants' economic, social and political involvement. The results show that an increase in educational attainment is associated with a higher level of Spanish spoken proficiency. Language ability is also associated with the country or region of origin. The results show that immigrant men and women from the Maghreb and Asia, as well as men from Eastern Europe and Sub Saharan Africa show a significantly weaker command over spoken Spanish than Western Europeans.

JEL classification: F22, J15, J24, J40

Keywords: Immigration, Language, Education

I. INTRODUCTION

A substantial number of immigrants cannot communicate adequately in the language of the country to which they immigrate. Among the multiple reasons for the

lack of host language proficiency one can distinguish factors such as a low level of educational attainment, not having been provided with adequate opportunities to learn the host language, living in ethnic enclaves or having arrived at an older age. Language skills (including oral and written ability) play a crucial role in the determination of the immigrants' social and economic integration in the host country. As a consequence, analyzing the source of foreign language acquisition is crucial for understanding the immigrants' economic, social and political involvement. A variety of studies have analyzed the process of language acquisition among immigrants in recipient countries such as Israel, Canada, the U.S., Germany and Australia.

Studies carried out by Chiswick (1991, 1997) Chiswick and Miller (1992, 1994, 1998, 2001 and 2007), Espenshade and Fu (1997), Dustmann (1997), Shield and Price (2000), Chiswick, Miller and Lee (2002, 2005) have analyzed this subject in different geographical contexts and under different circumstances. Most of these studies conclude that language proficiency is positively correlated with educational attainment and negatively related with age at arrival. Language skills improve substantially with the time spent in the host country. Moreover, empirical studies have almost unanimously found that language proficiency has a positive impact on earnings.

This paper carries out the first extensive research on Spanish language acquisition among immigrants in Spain. Spain is a case of particular interest since immigration has been very intense over the last decade. Annual inflows of immigrants, on average around 500,000 since year 2000, have raised the percentage of foreign population in Spain from 2% to 11% over an 8-9 years period. In 2007 Spain ranked second among OECD countries in the aggregate number of annual immigrant inflows just behind the United States (OECD, 2008). Interestingly, over this period of strong immigration growth, foreigners accounted for more than 80% of the total growth in the working age population and for almost 50% of the growth in employment (Izquierdo, Lacuesta, Vegas, 2009). Moreover, the increase in immigration took place in a period of robust economic growth, with a 3.6% average GDP growth and a considerable fall in the level of unemployment.

Given the importance and dimension of the phenomenon, extensive research is needed to devise strategies and immigration policies that guarantee the economic and social well-being of immigrants in Spain. However, accurate statistical data has been

scarce due to the recent nature of immigration in Spain and the subsequent lack of large-scale surveys. This in turn has limited and weakened the quantity and quality of research on immigration. This chapter uses micro-data from the Spanish National Immigrant Survey (NIS in English or Encuesta Nacional de Inmigrantes-ENI in Spanish) carried out among immigrants in Spain between November 2006 and February 2007. The survey was published by the National Statistics Institute (NSI-INE in Spanish) in 2008. This is the first extensive immigration survey carried out in Spain. It examines 15,465 respondents and their families. Immigrants are defined as individuals born abroad, irrespective of their nationality. The NIS comprises a single cross-section, which presently does not allow for a longitudinal analysis.

The NIS collects a wide range of statistical data about immigrants, including multiple questions on their socio-economic status, their mother language, as well as their Spanish and regional language speaking ability. In particular, the survey includes questions on Spanish language proficiency that I use as the dependent variable in the model that explores the determinants of Spanish language proficiency. The relevant survey questions for this chapter are stated as follows:

- *Besides your mother language, what other languages do you know?*

For the respondents that list Spanish as one of the foreign spoken languages, an additional question is posed:

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

1. Very Well 2. Well 3. Sufficient 4. Need to improve

Destination language acquisition will be analyzed using the human capital framework. It was not only until the 1980s that economists viewed immigrant language skills as a form of human capital and analyzed it in this context (Carliner, 1981; McManus, Gould, & Welch, 1983; Tainer, 1988). The model used in this article to measure the determination of language learning was presented by Chiswick (1992, 1995) based upon the human capital perspective. Language skills satisfy the necessary requirements to be considered human capital: they are embodied in the individual, useful in production and consumption activities, and are costly for the individual willing carry out the investment.

Following Chiswick and Miller (1992, 1998, and 2007), I categorize the elements that affect the acquisition of foreign language skills as exposure, efficiency, and economic factors. Exposure factors include the immigrant's mother language, the duration of residence and the number of children. Efficiency factors include age at migration, educational attainment, region of origin and motivation for migration (economic vs. refugee). Finally, economic incentives include expected duration of stay (temporary vs. permanent).

The statistical analysis includes males and females aged 16 and older. Only respondents from non-Spanish speaking countries (6,989 individuals) are included in the analysis since immigrants from Spanish-speaking countries already speak the destination language as their mother language when they arrive in Spain. Because of the likely gender differences in the determinants of language acquisition, we will compute separate equations for men and women. We analyze the impact of the determinants of Spanish-language proficiency using Ordinary Least Squares and Ordered Logit.

The main results that emerge from this study are as follows. An increase in educational attainment is associated with a higher level of Spanish spoken proficiency. The impact of schooling on language acquisition seems stronger for men than for women. Studying in Spain further increases the odds being in a higher language proficiency category. Moreover, the presence of children in the family shows a strong negative impact on women's language skills, as in previous research from other countries. Speaking a Romance language (French, Italian, Romanian or Portuguese) as a mother tongue increases the chances of being in a higher speaking proficiency category. Belonging to an association is positively associated with Spanish speaking ability. Finally, Spanish language ability is also associated with the country or region of origin. There are several broad groupings in my analysis: Western Europe (the reference category), Eastern Europe, Maghreb, Sub-Saharan Africa, America (including South, Central and North-America), and Asia. The results show that immigrant men and women from the Maghreb and Asia, as well as men from Eastern Europe and Sub Saharan Africa show a significantly weaker command over spoken Spanish than Western Europeans.

The remainder of the paper is organized as follows. Section II outlines a model of destination language proficiency for immigrants and reviews the literature on the

economic determinants of migration and destination language acquisition. Section III presents and explains the data and variables used for the statistical analysis. Section IV presents the methodology and Section V discusses the results, summarizes and provides the conclusions.

II. THE DETERMINANTS OF HOST COUNTRY LANGUAGE PROFICIENCY

II.1 The Economics of Language: a model of language acquisition for immigrants

Immigrant destination language can be analyzed using the human capital framework that will be described in Section V. In this context, acquiring a language has benefits in consumption and in labor market participation, but there is also a financial cost of acquiring these skills as well as an opportunity cost derived from foregone earnings while learning the language. These investments are made in anticipation of future benefits that may be in the form of higher earnings in the labor market, lower costs of consumption, greater political involvement or larger social networks, to name just a few.

It was not until the 1980s that economist began to study immigrant language skills as a form of human capital and analyzed it in this context (Carliner, 1981; McManus, Gould, & Welch, 1983; Tainer, 1988). Immigrants who speak the destination language will find it easier to obtain a job and are more productive in their job leading to a higher level of income.

Language skills have different dimensions including listening, speaking, writing and reading skills. Census and survey data nearly always rely on self-reported answers and most of the data are drawn for speaking ability only. However, some authors find that speaking and writing skills are highly correlated (Chiswick 1991, Chiswick and Repetto 2001) and follow show the same patterns in language acquisition.

Following Chiswick and Miller (1992, 1998, and 2007), we categorize the factors that affect the acquisition of foreign language skills as *exposure*, *efficiency*, and *economic incentives*:

1) *Exposure factors*: Exposure can happen either before or after immigration. Pre-immigration exposure depends on the degree to which the destination language is

used in the origin country and the study of the language in the home country. In the case of ex-colonies, a dichotomous variable could be used to capture these effects. Post-immigration exposure will mainly depend on the duration of residence, assuming that the immigrant has lived in the country continuously. A quadratic form will also be added to the model to capture the effect of an additional year in the destination country since the effect of duration is likely to be non-linear and decreasing. In other words, proficiency increases rapidly in the first years, but at a decreasing rate.

Moreover, ethnic enclaves play a major role in the immigrant's language acquisition. An immigrant can avoid speaking the local language living or working in neighborhoods in which others speak his or her home language. Settling in areas with others from the same origin provides for economies in communication, consumption, information and the labor market. This can be measured using the proportion of the population that speaks the immigrant's origin language in the area. The greater the extent to which an immigrant can avoid communicating in the destination language, the slower he is likely to acquire destination language skills. Moreover, because of economies of scale in the production of ethnic goods consumed by an immigrant group, the cost of these goods is lower the larger the size of the particular immigrant community.

Chiswick and Miller (1996) explore the effects of ethnic networks on language proficiency using a unique survey data set, Issues and Multicultural Australia 1988 that provides measures for various dimensions of ethnic linguistic environment. They present different models for speaking, reading and writing proficiency and find that a high concentration of minority-language speakers has a statistically significant negative effect on all three language skills. Furthermore, Chiswick and Miller (2003) also analyze the impact of immigrant concentrations on proficiency in destination language skills using the 1990 U.S. Census of Population. Again, their results show that a smaller minority language concentration ratio and living in a rural area are both associated with greater English proficiency among immigrants. Particularly among Mexican immigrants, greater access to Spanish language radio stations leads to poorer English language proficiency.

Dustmann (1997) warns that ethnic enclaves are potentially endogenous in a language equation since those immigrants that tend to move into an ethnic

neighborhood might do so because of their lack of native language proficiency. According to River-Batiz (1996), immigrants with lower proficiency in the destination language may tend to seek employment opportunities in ethnic enclaves. Unfortunately the survey does not provide information on the number of immigrants living in the same municipality nor the municipality of residence itself. The only information provided refers to the province of residence.

The language spoken in the household may also impact the individual's foreign language proficiency. Chiswick, Lee, and Miller (2005) show strong correlations between the different family members' dominant language fluency skills. In particular, they show that correlations in destination language proficiency are stronger between spouses than between parents and children. Moreover, when marriage takes place after immigration, it is more likely that the spouse is not proficient in his mother language and this may facilitate the use and learning of the destination language. Chiswick, Lee and Miller (2002) also find a substantial positive correlation between the language skills of the spouses due to both measured and unmeasured variables.

The presence of children on destination language acquisition is significant but complex. On the one hand, the rapid learning of the destination language among children may enable them to teach their parents the dominant language skills. On the other hand, children may serve as translators, reducing the incentive for their parents to learn the destination language. Since this impact is more likely to happen in consumption activities rather than work activities, mothers are more likely to be negatively affected. Children may also affect labor supply, especially among females, reducing the chance and incentive to acquire foreign language skills. Finally, parents who seek to transmit the culture of their home country may encourage the learning of the home language to enable them to communicate with their grandparents and cousins or facilitate return migration in the future, limiting the chance to develop destination language skills among all family members. Children thus have positive (teaching) and negative (the rest) effects on their parents' dominant language acquisition. As a consequence, the sign of the overall effect is ambiguous and is expected to differ between genders. The impact is more likely to be negative for the mother. Chiswick, Lee and Miller (2002) find that having children has a negative impact on the mother's destination language proficiency, but not on the father's.

2) *Efficiency*: Some individuals find it easier to learn a language than others and this may be an important factor in language acquisition. An older age at migration is expected to have a negative impact on foreign language acquisition since younger individuals seem to have a greater capacity to learn languages for biological reasons (Service 1993). There also seems to be a critical age threshold beyond which an immigrant's learning of a second language may become very difficult. Moreover, educational attainment is expected to have a positive impact on foreign language acquisition. More educated individuals may have a greater learning ability (innate or acquired in school) or other unobserved variables (motivation) that enhance both education and language skills.

Dustmann (1997) argues that the father's educational background may determine his and his children's intellectual broadness. As a consequence, individuals from families with higher educational background may develop a stronger interest for acquiring foreign language skills. In fact his analysis shows that parental educational background has a strong positive impact on German speaking and writing abilities among immigrants in Germany.

Additionally, an individual's mother tongue and its distance to the destination language (in this analysis, Spanish) will also affect foreign language acquisition. The linguistic distance between Italian and Spanish is considerably smaller than between Arabic and Spanish. The lower the linguistic distance between the mother tongue and the destination language, the greater the level efficiency to learn a destination language. More investment in destination language training would be required to achieve the same level of proficiency if the origin language is more distant.

In an interesting paper Chiswick, Beenstock and Repetto (2001) separate country of origin effects from language of origin effects in the acquisition of Hebrew language skills in Israel. They use data from the 1972 Census of Israel and the Immigrant Absorption Surveys carried out during the 1970s. Their results show that immigrants from most language origins are less proficient in Hebrew than those from Arabic-speaking countries. Arabic is the closest language to Hebrew (both are Semitic languages), which is consistent with their Semitic origin. English speakers show the least proficiency in speaking, reading and writing Hebrew as they make slower progress than others in acquiring Hebrew language skills. These findings point to the existence of

linguistic distance effects in destination language acquisition. They seem to be important and usually take time to emerge. The disadvantage that they find for English speakers may derive from the fact that English is an international language and therefore reduces the incentive for immigrants in acquiring Hebrew proficiency. Their analysis shows the importance of both country of origin and linguistic distance in the acquisition of a foreign language.

Chiswick and Miller (1998, 2005) develop a quantitative measure of the distance between English and a set of other languages. In particular, they create an index for English, based on the ability of Americans to learn a variety of foreign languages in a given period of time. That is, the lower the scores on standardized foreign language tests, the greater the distance between these languages and English. They use their indexes as an explanatory variable and find that linguistic distance has a negative impact on immigrant's language proficiency. Linguistic distance may be related to the self-selection of immigrants. An example is the large amount of Romanians moving to Spain in the last decade. Romanians are by far the largest group of Eastern European immigrants in Spain. Romanian is a Romance language, closer to Italian, French, Spanish and Portuguese. The proximity of the languages makes Spanish language acquisition easier for Romanian immigrants.

Finally, admission criteria may also have an impact. Refugees tend to be less favorably selected based on their socio-economic background and often have less time to prepare for the move and less incentive to invest in human capital due to the uncertainty derived from their status. As a consequence, they are expected to learn less quickly than other immigrants. On the contrary, economic migrants are the most favorably selected for labor market and are therefore expected to learn the destination language faster. Typically, economic immigrants have the highest level of destination language proficiency, followed by family immigrants and refugees tend to have the lowest levels (Chiswick, 2007, 2008)

3) *Economic Incentives*: the longer the expected residency, the stronger the incentive for immigrants to acquire foreign language skills. A greater geographical distance may imply a higher cost of return migration and should be associated with greater dominant language skills. However, the further the country of origin from the destination country, the more intense the self-selection in immigration as well.

Dustmann (1999) uses data from the German Socio-Economic Panel to analyze whether temporary immigrants' acquisition of language capital is sensitive to the intended duration of residence in the host country. The analysis distinguishes between temporary migration where the return date is exogenous and temporary migration where migrants choose their date of return. His results show clearly that those who intend to remain permanently have higher probability of being proficient in the host language. Language fluency is thus negatively affected by the migrant's return propensity. This suggests that immigrants who plan to remain a longer period of time in the host country invest more in human capital.

III. THE DATA

III.1 The National Immigrant Survey (La Encuesta Nacional de Inmigrantes)

The final results from the National Immigrant Survey (NIS in English or Encuesta Nacional de Inmigrantes-ENI in Spanish) were released by the National Statistics Institute of Spain (www.ine.es) in July 2008. The survey was carried out in collaboration with the Universidad Complutense de Madrid. The project was initiated by the Working Group for the Study of Population and Society (GEPS- in Spanish), that was set up in 1997 at the School of Political Science and Sociology of the Universidad Complutense de Madrid.

This is the first large-scale immigration survey carried out in Spain. It examines 15,465 respondents and their families, defined as individuals born abroad, irrespective of their nationality. It does not include those immigrants with foreign citizenship that were born in Spain, but does include foreign-born individuals with Spanish citizenship. The participants are 16 years or older that either have been living in Spain for more than one year or have the intention to do so. The survey was implemented between November 2006 and February 2007 and covers the entire Spanish territory. According to the National Institute of Statistics (NIS) there were 2.16 million households in Spain in which at least one of their members was aged 16 or over and was born outside Spain. There were 1.02 million households solely comprising immigrants and 1.14 million households comprised of a combination of immigrants and non-immigrants.

The survey is divided into a series of sections:

1) *Identification* of the respondent and his family members, including information on gender, age, nationality, country of birth, and year of arrival in Spain of all the members in the household.

2) *Socio-demographics*: this section includes the main socio-demographic characteristics of the respondent and his family members including civil status, educational attainment, working status and language proficiency. It also contains information about the characteristics of the residence in which they currently live.

3) *Migration experience*: countries in which the respondent and his closest relatives have lived since birth and before arriving in Spain.

4) *Home country conditions prior to departure*: information related to the family, work and economic environment (push factors) in their home country.

5) *Arrival in Spain*: conditions under which the immigrant arrived in Spain, including his journey and the expenses derived from the trip.

6) *Work experience in Spain*: focuses on the current and first job in Spain, including occupational status, occupation and earnings among other job characteristics.

7) *Residential experience in Spain*: main characteristics of the houses in which they lived and their regional mobility in Spain.

8) *Home country relationship*: reasons for moving to Spain, contact with their relatives and friends at home, remittances and properties in their home country.

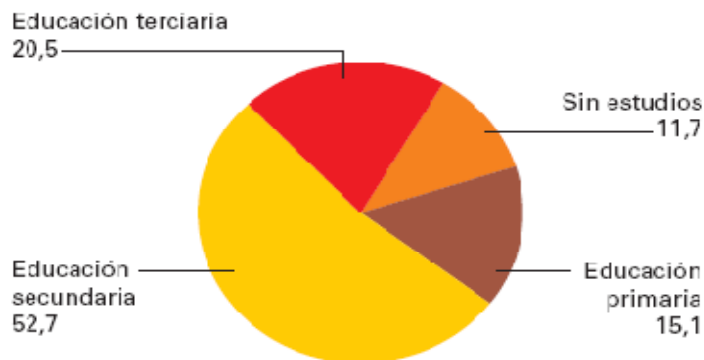
9) *Social involvement in Spain*: information including participation in local elections, membership in associations, and future plans to remain in Spain.

As mentioned earlier, individuals are considered immigrants if they were born abroad, regardless of their current citizenship. As a consequence, the survey includes a considerable proportion of immigrants with Spanish citizenship-22% of which 7% were Spanish by birth and the rest-15% acquired their citizenship after living in Spain.

The survey shows that the majority of the immigrants included arrived in Spain in the last 10 years from a relatively limited number of countries, compared to other European Union recipient countries, Canada or the US. The average number of years that an immigrant has been living in Spain is 11 years and the standard deviation is 12

years. In contrast, the median immigrant arrived in 2001 and has been living in Spain for 6 years. The most important immigrant origin regions or countries are Latin-America, Morocco, Romania and Western Europe. On the one hand, Western European immigrants are older on average and show a higher level of educational attainment. They tend to work in more qualified jobs or are already retired. On the other hand, Latin-American, African, Asian and Eastern European immigrants arrived more recently and hold low-skilled positions in the labor market. We are mainly interested in their level of educational attainment, their mother language and their Spanish speaking proficiency. Figure III.1 shows the distribution of educational attainment among immigrants:

Figure III.1: Educational attainment of Immigrants in Survey

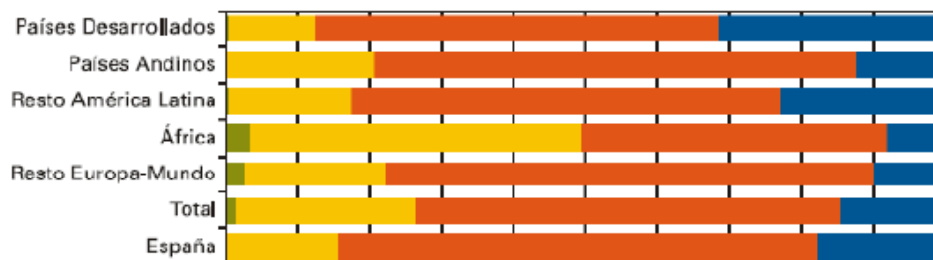


Source: ENI Report

No Education, Primary education, Secondary: yellow, Tertiary

Figures III.2a and III.2b show a comparison of educational attainment by gender for native Spaniards with educational attainment for immigrants from different geographical world regions.

Figure III.2a: Educational attainment for men by region (age 20-34)

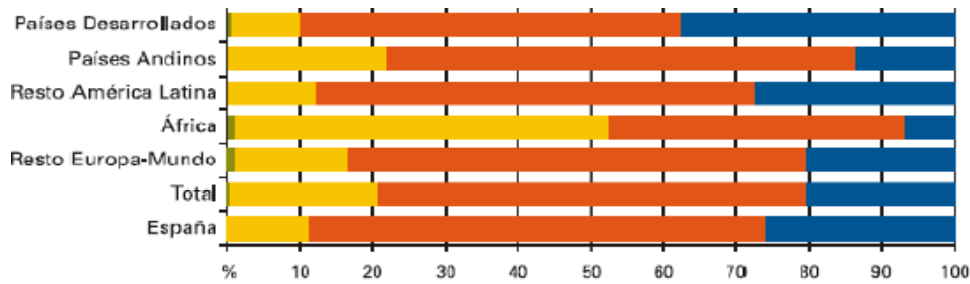


No Education, Primary education, Secondary, Tertiary

Source: ENI & Population Census

Categories: Western Countries, Andean Countries, Rest of Latin-America, Africa, Rest of Europe-World, Total, Spain

Figure III.2b: Educational attainment for women by region (age 20-34)



No Education, Primary education, Secondary, Tertiary

Source: ENI & Population Census

Categories: Western Countries, Andean Countries, Rest of Latin-America, Africa, Rest of Europe-World, Total, Spain

We observe from the above figures that educational attainment among Spaniards is not substantially different from educational attainment among immigrants. Immigrants from Western countries have the highest level of educational attainment and Africans the lowest. Immigrant women are more educated than men on average.

As regards to mother language, Figure III.2 shows the majority of the immigrants in the survey speaks Spanish as a mother language (44.9%) and will not be considered for our Spanish language proficiency analysis. Non-Romance Indo-European languages represent 19.4% of the sample and 18.2% of the immigrants speak Romance languages including Romanian, French, Italian and Portuguese as shown in Figure 7.

Figure III.3: Mother languages spoken as a percentage of the total immigrants



Source: ENI report

Language categories: Spanish, Indo-European languages, Latin languages, Afro-Asian languages, Oriental, African, Non-classified

The survey includes questions on Spanish language proficiency that will be used as the dependent variable in the models that explore *the determinants of Spanish language proficiency*. In particular, the survey questions are stated as follows:

- **Besides your mother language, what other languages do you know?**

For the respondents that list *Spanish* as one of their spoken foreign languages, an additional question is posed:

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

1. Very Well 2. Well 3. Sufficient 4. Need to improve

We will create a four-level categorical variable named SPANISH PROFICIENCY, which shows the immigrants' Spanish language proficiency level. The number of immigrants speaking Spanish as a foreign language is 6,989:

Table 2.1: Spanish language proficiency

Spanish language proficiency	Frequency	Percent
Needs to improve	1,027	14.69
Sufficient	960	13.74
Good	2,224	31.82
Very good	2,778	39.75
Total	6,989	100.00

According to the survey data in Table 1, the majority of immigrants (71.5%) show a good or very good level of spoken Spanish proficiency.

IV. METHODOLOGY

IV.1 The determinants of Spanish language proficiency

We generate a conceptual equation with the hypothesized signs for the *Spanish language proficiency model* following:

SPANISH PROFICIENCY = f(Age at migration (-), Years since migration (+), YSM² (-), Educational attainment (+), Education in Spain (+), Children (?), Romance language background (+), Plans to stay in Spain (+), Association membership (+), Country of Origin, Region of Residence)

Following the literature, we expect immigrants that arrived at an older age to be less fluent in Spanish. The length of their period of residence in Spain should have a positive impact at a decreasing rate. Moreover, the individual's educational attainment and having acquired some formal education in Spain is expected to be positively correlated with his Spanish speaking ability. Having children may affect the parents' language skills differently depending on the parent's gender. We expect a negative impact on women and a non-significant effect on men's destination language proficiency. Romance language background is positively related to linguistic distance and is therefore expected to impact Spanish language acquisition positively. Furthermore, belonging to an association and planning to stay in Spain more permanently should have a positive impact. We finally add dummy variables for the geographical area of origin (Eastern Europe, Western Europe, Sub-Saharan Africa, America, Oceania and the Maghreb) and the region of residence (Autonomous Community).

The regression analysis includes males and females aged 16 and older -we carried out the same analysis for 24 and older and obtained similar results. Only respondents from non-Spanish speaking countries (6,989) are included in the analysis since immigrants from Spanish-speaking countries speak the destination language as their mother language when they arrive in Spain. Because of the likely gender differences observed in the determinants of host language acquisition, we compute separate equations for men and women. I will note the cases in which the direction or the strength of the estimated coefficients varies significantly between men and women.

The impact of the determinants of Spanish-language proficiency is analyzed using an Ordinary Least Squares (OLS) and an Ordered Logit analysis. The Spanish language-proficiency measure SPANPROF is the dependent variable. The advantage of OLS over logit models is that the coefficients are easier to interpret. The disadvantage is that OLS estimates may be biased due to the presence of censored dependent variables. The predictions derived from the model may be outside the permissible range of 0 to 4,

and as a consequence the regression disturbance term may be heteroskedastic. If sample selection is entirely random, then OLS estimates are unbiased. If sample selection depends on the explanatory variables and additional random terms that are independent of x and u , then OLS is also consistent. If sample selection is correlated with error term, then OLS is inconsistent.

If we changed intervals and decided that the distances are not all equal, that would change the slope. To avoid this problem, we can use ordered logit. It is based on the idea of a latent dependent variable, which we can only observe as a set of categories – but in fact, it is a continuous variable. However, the major conclusions do not seem to be sensitive to the choice of the estimation technique according to previous analyses (Chiswick and Miller 1998b).

V: REGRESSION ANALYSIS RESULTS

The choice of the statistical method is guided by the ordinal nature of the dependent variable (Spanish language proficiency). Our findings will be based on two regression techniques: Ordinary Least Squares and Ordered Logit. The dependent variable is based on qualitative responses to the question “Considering your needs in the workplace, at the bank, with the authorities, how well do you speak Spanish?” Responses are coded 1 (needs to improve), 2 (sufficient), 3 (well), and 4 (very well). A positive (negative) regression coefficient means that an increase in the value of the corresponding explanatory variable is expected to raise (lower) the respondent's Spanish proficiency level.

The coefficients obtained from the ordered logit analysis have meaning in terms of relative odds. Suppose that ordinal responses are classified into J categories. We can model the logit of y_{ij} , the cumulative probability that an individual with characteristics vector x_i is in response category j or lower ($j = 1, 2, \dots, J-1$), as a linear function of the predictors. If P is the vector of ordered-logit regression coefficients, then $\exp\{-pk\}$ is the odds ratio or the proportionate change in the odds of being in category j or less that is produced by a one-unit increase in X_{ik} , the k th explanatory variable.

We run separate regression analyses for men and for women in order to compare the direction and intensity of the explanatory variables between genders. We expect

differences between genders regarding the strength of the coefficients for some variables. We carry out the analysis for immigrants 16 years and older. The results for the immigrant group 24 years and older are very similar. As a consequence, we only present the results for the sample that includes individuals that are 16 years and older in Table V.1.

Table V.1: The determinant of Spanish language proficiency for male immigrants

Variable	OLS1		OLS 2		ORD.LOGIT1		ORD.LOGIT2	
	Coefficient	t-stat	Coefficient	t-stat	Odds Ratio	z-stat	Odds Ratio	z-stat
Constant	2.152***	16.67	2.356***	17.00				
Years Migr.	.0486***	13.06	.046***	11.79	1.115***	13.37	1.109***	12.09
YSMSQUARE	-.0005***	-8.52	-.0005***	-8.23	.998***	-8.53	.998***	8.26
Education	.0607***	17.03	.056***	14.92	1.134***	16.70	1.124***	14.67
Study Spain	.1049	1.77	.113	1.93	1.623***	3.67	1.661***	3.82
Children	-.0815*	-2.25	-.066	-1.83	.872*	-1.84	.910	-1.24
No plans	.0992	1.12	.100	1.14	1.291	1.46	1.296	1.48
Stay Spain	.1774*	2.27	.171*	2.20	1.465*	2.49	1.455*	2.45
Move third	.1183	0.72	.109	0.67	1.212	0.57	1.162	0.44
Age Arrival	-.0184**	-3.15	-.015**	-2.67	.944***	-4.34	.947***	-4.07
AASQUARE	.00001	0.25	-.00005	-0.59	1.0002	1.67	1.0001	0.86
Association	.1482**	3.17	.129**	2.77	1.448***	3.66	1.376**	3.12
Romance L.	.2974***	8.77	.222***	5.45	1.884***	8.88	1.575***	5.28
Maghreb			-.180***	-3.35			.623***	-4.13
Subs-Africa			-.336***	-4.71			.476***	-4.97
Eastern			-.146**	-3.00			.667***	-3.99
America			.075	0.95			1.255	1.25
Asia			-.286***	-3.74			.480***	-4.61
Oceania			.178	0.43			2.740	0.76
Adjusted R ²	.2359		.2428					
Likelihood								
Ratio Chi ²					1015.04		1064.64	
Sample	3349		3349		3349		3349	

* $p < .05$ ** $p < .01$ *** $p < .001$

Table V.I presents the results for the analysis of the determinants of Spanish speaking proficiency using OLS and Ordered Logit. The results provided for the Ordered Logit regression are shown in Odds Ratios. For example, for model 1 we can say that the odds of speaking Castilian Spanish very well versus speaking the language well, sufficient or not sufficiently well are 1.8 times higher for those immigrants that have a Romance Language background. The main patterns that emerge do not vary significantly with the statistical methodology used. We only find minor differences for the impact of having acquired some education in Spain. Accordingly, we will interpret the results focusing on the results obtained using Ordered Logit.

Both models show that years since migration have a statistically significant positive effect on Spanish speaking proficiency. The negative coefficient on the squared term indicates decreasing returns to years since migration, possibly because most of the language learning takes place in the first years after arrival. More importantly, educational attainment has a strong and significant positive impact on the ability to speak Spanish. The reason is that more educated individuals may have a greater learning ability (innate or acquired in school) or other unobserved variables (motivation), which enhance both educational attainment and language skills. Moreover, attending school in Spain increases their odds of being in a higher language proficiency category.

Other findings show that age at arrival has a significant negative effect on Spanish speaking ability, since younger individuals seem to have a greater capacity to learn languages for biological and environmental reasons. The negative coefficient of having children is statistically significant at a 95% level, only when geographical dummies are not included. The impact of children is weaker for men since they are less involved in child upbringing. Speaking a Romance language as their mother tongue increases their odds of being in a higher proficiency category. The similarity of the languages as regards to grammar, vocabulary and pronunciation allows for a faster and better learning of Spanish.

Moreover, belonging to an association and planning to stay in Spain more permanently is positively associated with Spanish language proficiency. Temporary migration leads to lower incentives to invest in human capital since the benefits will be collected for a shorter period of time. Previous research shows that those who intend to remain more permanently have a higher probability of being proficient in the destination language. Language fluency is thus negatively affected by the migrant's return propensity since immigrants who plan to remain for a longer period of time in the host country invest more in human capital.

Finally, Spanish language usage is also linked to the country or region of origin. There are several broad groupings in our analysis: Western Europe (our reference category), Eastern Europe, Maghreb, Sub-Saharan Africa, America (including South, Central and North-America-14.6% of the non-Spanish speaking immigrants from the Americas), Asia and Oceania. The results show that immigrant men from Maghreb, Sub-Saharan Africa, Eastern Europe and Asia show a significantly weaker command

over spoken Spanish than Western Europeans. Asians and Sub-Saharan Africans are among the least proficient immigrants.

These results may reflect the linguistic distance between Indo-European and African/Asian languages or the curriculum and quality of their educational systems. Another reason may be related to cultural differences and ethnic concentrations in the neighborhood and/or at the workplace. Ethnic enclaves make the host country language more difficult to access for particular immigrant groups. If the propensity to create ethnic enclaves is different for individuals from different geographical and cultural backgrounds, this may be reflected in the regional dummies. Unfortunately, there is no data on the proportion of the population that speaks the immigrant's origin language in the area of residence. As regards to residency, only information on the province (a broad category) is available in the survey. This limitation does not allow for a distinction between immigrants living in urban and rural areas either. A final set of explanation for the differences between immigrant groups includes self-selection, different institutional settings, geographical distance and reward systems that may lead to differences in the types of immigrants that enter the country. These are sources of endogeneity. Immigrants that are more able to learn languages in general or Spanish in particular move to Spain and that makes it more difficult to find the direction of causality.

We add a third model including dummy variables for Spanish region of residence in the appendix. Madrid is the benchmark region and there are several regions that are associated with a lower level of Spanish speaking proficiency. Surprisingly, most of the regions where immigrants seem to be at a disadvantage in Castilian Spanish proficiency are not bilingual Autonomous Communities (Andalusia, Aragon, Asturias, Extremadura, Galicia, Murcia and La Rioja). These are regions where rural jobs are more common and this may affect the acquisition of language proficiency. Adding the regional dummies does not seem to affect the strength of the coefficients for the other determinants of Spanish language proficiency. Table 2.9 shows the results for immigrant women.

Table V.2: The determinant of Spanish language proficiency for female immigrants

	OLS1		OLS 2		ORD.LOGIT1		ORD.LOGIT2	
Variables	Coefficient	t-stat	Coefficient	t-stat	Coefficient	z-stat	Coefficient	z-stat
Constant	2.538***	20.63	2.683***	20.22				
Years Migr.	.0486***	14.52	.047***	12.82	1.121***	14.61	1.118***	12.91
YSMSQUARE	-.0005***	-9.93	-.0005***	-9.43	.998***	-10.03	.998***	-9.50
Education	.0517***	13.97	.047***	12.01	1.118***	14.02	1.109***	12.07
Study Spain	.0514	0.71	.066	1.20	1.490**	3.05	1.473**	2.94
Children	-.1648***	-4.60	-.156***	-4.34	.734***	-4.05	.755***	-3.65
No plans	-.0860	-0.98	.082	-0.95	.918	-0.49	.910	-0.54
Stay Spain	.0514	0.71	.0491	0.67	1.209	1.33	1.190	1.22
Move third	.2662	1.57	.2473	1.46	1.863	1.69	1.719	1.47
Age Arrival	-.0178**	-3.09	-.017**	-2.99	.934***	-4.95	.933***	-4.95
AASQUARE	-.00002	-0.34	-.00005	-0.67	1.0003	1.93	1.0003*	2.18
Association	.0975*	2.09	.0896	1.91	1.414***	3.23	1.380**	2.98
Romance L.	.2578***	8.03	.2004***	5.26	1.803***	8.55	1.590***	5.92
Maghreb			-.175**	-3.11			.649***	-3.54
Subs-Africa			-.160	-1.62			.657*	-2.08
Eastern			-.044	-0.96			.836	-1.83
America			-.069	-0.58			.865	-1.02
Asia			-.325***	-3.58			.458***	-4.08
Oceania			.317	0.91			1.673	0.60
Adjusted R ²	.2317		.2373					
Likelihood								
Ratio Chi ²					1036.64		1061.47	
Sample	3419		3419		3419		3419	

* $p < .05$ ** $p < .01$ *** $p < .001$

Table V.2 presents the female results for the analysis of the determinants of Spanish speaking proficiency using OLS and Ordered Logit. The results provided for the Ordered Logit regression are shown in Odds Ratios. For model 1 we can say that the odds of speaking Castilian Spanish very well versus speaking the language well, sufficient or not sufficiently well are 1.41 times higher for those immigrant women that participate in associations.

The results for immigrant women show that years since migration have a significant and positive effect on Spanish speaking proficiency. Educational attainment shows a strong and positive effect on the ability to speak Spanish, but the effect is slightly weaker for women than for men. Studying in Spain increases the odds of being in a higher proficiency. Again the impact is slightly weaker than for immigrant men. Age at arrival is negatively associated with Spanish proficiency, as expected.

Moreover, having children shows a strong statistically significant negative impact on women's destination language skills. The negative impact is stronger for the mothers since children have a greater impact on female labor supply and mothers are more involved in the domestic environment upbringing of their children and the transmission of their culture including their home language. Speaking a Romance language increases the odds of being in a higher proficiency category. Belonging to an association has a positive impact on Spanish language proficiency. Interestingly, planning to stay in Spain has no significant positive impact on women's language proficiency as opposed to men's. Women's labor force participation rate is typically inferior to men's and their contributions to the household's income tend to be lower. As a consequence, their incentive to invest in language acquisition may be lower.

Finally, Spanish language usage is also associated with country or region of birth, although the differences are weaker than for men. For women, differences remain statistically significant just for Maghrebians and Asians. Ethnic concentration in the workplace might be stronger among men. Moreover, cultural differences among Maghrebians and Asians may lead to differences in the language learning process between genders. We include dummy variables for region of residence in the appendix.

Differences between men and women

Amongst the most important differences between genders, we find that the impact of education and studying in Spain on Spanish language ability is significantly stronger for men. A higher level of educational attainment is likely to allow men to achieve a stronger upgrade in their professional status as compared to women due to a certain level of gender discrimination in the labor market. This professional upgrade may lead them to jobs that are less physical and allow for more linguistic interactions.

Furthermore, children barely have an impact on men's language skills, but do have a strong statistically significant negative for women. The result confirms that children are more likely to affect language learning for women than for men. Children may also affect labor supply, especially among females, reducing the chance and incentive to acquire foreign language skills. Planning to stay in Spain is positively associated with men's language skills, but does not seem to be related with women's language acquisition. This may be related to the fact that men are more willing to invest

in human capital-Spanish language acquisition, since they are more likely to work and obtain higher salaries during their stay in the destination country.

We finally add interaction terms between educational attainment and country of origin in the appendix of the chapter. I intend to capture differences in the impact of schooling on Spanish language proficiency between different immigrant groups. I find a statistically significant positive interaction term for immigrant men from Africa (including Maghreb and Sub-Saharan) and for immigrant women from Maghreb. For these immigrants educational attainment has a stronger impact on the acquisition of destination language proficiency. It is therefore important to encourage and promote schooling programs especially among African migrants.

V.2 Self-selection, endogenous choice and measurement error

There are several methodological issues that we need to address regarding the use of OLS and Ordered Logit in the determination of language proficiency. On the one hand, the estimated coefficients for the language proficiency variable may be biased due to identification problems that arise from endogeneity or unobservable variables. This may lead to an upward bias as a consequence of the positive correlation between unobserved heterogeneity in language proficiency and some explanatory variables. For example, those who do better in language acquisition may be induced to get more schooling and to use the acquired language more in social interactions or associations. This problem may be corrected using I.V., but it is important to find adequate instruments. In other words, the practical difficulty with IV estimation is finding an instrument that is significantly correlates with educational attainment but also orthogonal to the residuals of the main equation. A number of recent studies have questioned the interpretation given to IV estimates and their usefulness for policy evaluation (Shield and Price, 2000).

Moreover, the coefficient for language ability may also be biased as a result of measurement error. Measurement error in the self-reported language variable is expected to lead to a substantial downward bias in the estimated coefficients. This occurs when respondents have the tendency to over- or under-report consistently in surveys. Charette and Meng (1993) focus on the nature and potential impact of

measurement error resulting from the use of self-assessed and discrete measures of language fluency. Their findings show that while the measurement error may distort the estimates of the determinants of language ability, they do not seem to affect the estimates in an earnings determination model. According to the authors, the use of an I.V. technique may mitigate the effect of measurement errors inherent in a self-assessed measure of literacy.

Finally, Dustmann and Soest (2001) show that self-reported measures of speaking fluency suffer from misclassification errors. Their estimation results for the language equation indicate that the probabilities of over-reporting are higher than the probabilities of under-reporting. However, the way misclassification is modeled does not have much effect on the coefficients in the language equation. The authors conclude that the estimates of the determinants of speaking fluency appear to be robust. 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.

VI SUMMARY, CONCLUSIONS AND DISCUSSION

This paper has analyzed the determinants of Spanish language proficiency for male and female immigrants in Spain, applying the theoretical model presented by Chiswick for the acquisition of foreign language skills. The model is based on the human capital theory. The data was obtained from the National Immigrant Survey, the first large-scale survey among immigrants carried out in Spain. The survey was implemented between November 2006 and February 2007 and covers the entire Spanish national territory. The results and patterns found for Spain are similar to those obtained in previous analyses for other major immigrant recipient countries like Germany, the United States, Canada, Australia and Israel. However, this is one of the few analyses carried out in a non-English speaking country. This is a relevant fact since English is the most widely used second language and many immigrants speak the English language already when they arrive in the destination country. Moreover, the recent and intense nature of immigration into Spain makes the country a unique context for this type of analysis. Most of the immigrants in Spain arrived in the last decade, which makes the composition of the group more homogenous. The present research is limited to

individuals aged 16 years and older. Both male and female immigrants are considered, in contrast to most of the recent literature that focuses on male adults aged 25-64 years.

The results show that the number of years since migration has a positive impact on Spanish speaking proficiency, as expected. The strength of the impact decreases as time passes. Language proficiency tends to be lower among those that arrive at an older age, since younger immigrants learn languages easier and more quickly, for biological and environmental reasons. Moreover, higher levels of educational attainment are associated with a higher Spanish speaking proficiency category. The impact of education on language acquisition is stronger for men than for women.

Having children has a stronger negative impact on women's language acquisition, since women are more involved in their upbringing. Speaking a Romance language as mother tongue increases the odds being in a higher language proficiency category. Furthermore, belonging to an association has a positive impact on language acquisition and planning to stay in Spain more permanently is also positively associated with language learning, although just for men.

Spanish language usage is also associated with country or region of origin. There are several broad groupings in my analysis: Western Europe (the reference category), Eastern Europe, Maghreb, Sub-Saharan Africa, America (including South, Central and North-America), Asia and Oceania. The results show that immigrant men and women from Maghreb and Asia, as well as just men from Eastern Europe and Sub-Saharan Africa show a significantly weaker command over spoken Spanish than Western Europeans, adjusting for other determinants of Spanish language proficiency.

Finally, living in the Autonomous Communities of Andalusia, Aragon, Galicia, Murcia, Extremadura and La Rioja is associated with a weaker Spanish speaking proficiency. Interestingly, among these regions Galicia is the only bilingual region, where Galician and Spanish coexist as the official languages. I will discuss the acquisition of regional language proficiency in Spanish bilingual regions in the following chapter of my dissertation.

The direction of the impact for the determinants of Spanish language proficiency was remarkably similar between men and women. Among the few variables giving rise to statistically significant gender differences in the strength of the impact are

educational attainment and studying in Spain. The positive impact is significantly stronger for men in both cases. In contrast, the negative impact of having children is significantly stronger for women and barely significant for men. Finally, planning to stay in Spain more permanently is positively associated with Spanish language acquisition just for men.

Cultural, social and economic assimilation in the destination country are crucial for keeping a cohesive and stable society, especially under the particular circumstances of massive newly arrived immigration as in Spain. In 2005 a series of riots and civil unrest developed in the *banlieus* of Paris and quickly spread to other areas in France. Both in 1981 and 1991, extensive rioting in France occurred along lines similar to those that happened in 2005.

Moreover, the fact that the rapid increase in immigration occurred in Spain in the last decade makes the learning of culture, traditions, values and language by immigrants even more relevant. From an analytical point of view the problem arises from the concept of assimilation and the difficulty of measuring it. Earnings, employment and language skills are easier to estimate than culture, traditions and values. We analyze the impact of the determinants of Spanish language proficiency since destination language acquisition contributes to cultural and social assimilation. Economic assimilation through higher earnings in the labor market and more chances of full-time employment are other substantial benefits derived from host language acquisition.

The results from our research are important to enable policy makers to devise strategies and immigration policies that promote and guarantee economic and social stability. The analysis shows that immigrants with higher educational attainment, receiving some formal education in Spain and planning to stay in Spain for a longer period of time, show better spoken Spanish language ability. As a consequence, it would be advisable to set up an extensive introduction program for immigrants. The program should partly consist of Spanish language instruction, combined with vocational training, in order to ensure familiarity with work-related terms and usages. Classes should be organized by education level of participants to ensure adequate learning. There is also a need for professional development in teaching Spanish as a second language. Moreover, policies that favor permanent migration over temporary

migration would also allow for a better social and economic integration. They may provide immigrants with more stability and time to acquire the necessary host language skills.

This research has important implications for understanding and predicting Spanish language experiences for future immigrants from various geographical areas in different Autonomous Communities. The results show that African and Asian immigrants seem to be at disadvantage and may require more intensive Spanish language training programs at arrival. Moreover, immigrants seem to encounter more difficulties in acquiring language skills in several Autonomous Communities.

APPENDIX

Appendix 1a: The determinant of Spanish language proficiency for male immigrants including region of residence (OLS)

<u>Variable</u>	<u>Coefficient</u>	<u>Std. Err.</u>	<u>t-statistic</u>
Years s/Migration	.0451532***	.0039191	11.52
YSMSQ	-.0005242***	.000067	-7.82
Education	.0555483***	.0038107	14.58
Study Spain	.1113796	.0590044	1.89
Children	-.0694285	.0361781	-1.92
Stay Spain	.1635459*	.077508	2.11
No plans	.0830425	.088055	0.94
Move third	.1085647	.1627359	0.67
Age Arrival	-.0172637**	.0058484	-2.95
AASQ	-.0000206	.0000809	-0.26
Association	.129451**	.0468371	2.76
Romance Lang	.2404668***	.0412362	5.83
Maghreb	-.1720275**	.0561229	-3.07
Subs-Africa	-.3369639***	.0713688	-4.72
Eastern Europe	-.1756637***	.050429	-3.48
America	.0557661	.0790021	0.71
Asia	-.2795134***	.077236	-3.62
Oceania	.1024162	.4155549	0.25
Andalusia	-.3583504***	.0783357	-4.57
Aragon	-.1878597*	.088301	-2.13
Asturias	-.3408967**	.1303925	-2.61
Balearic Islands	.0103593	.080317	0.13
Canary Islands	-.1630573	.1007608	-1.62
Cantabria	-.039519	.1216113	-0.32
Castile-Leon	-.0172157	.0945778	-0.18
Castile-La Mancha	-.0895811	.0880241	-1.02
Catalonia	-.0981641	.0682702	-1.44
Valencian Comm.	-.0961568	.0719949	-1.34
Extremadura	-.2849016**	.1079275	-2.64
Galicia	-.3763098**	.1124614	-3.35
Murcia	-.2382021**	.0810905	-2.94
Navarre	.1315203	.0790384	1.66
Basque Country	-.1264132	.0996121	-1.27
La Rioja	-.2287046*	.0980786	-2.33
Adjusted R ² Likelihood		.2546	
Sample		3349	
* $p < .05$	** $p < .01$	*** $p < .001$	

Appendix 1b: The determinant of Spanish language proficiency for female immigrants including region of residence (OLS)

<u>Variable</u>	<u>Coefficient</u>	<u>Std. Err.</u>	<u>t-statistic</u>
Years s/Migration	.0476167***	.003706	12.85
YSMSQ	-.0005689***	.000062	-9.18
Education	.0420026***	.004010	10.47
Study Spain	.0337038	.055669	0.61
Children	-.1474521***	.035551	-4.15
Stay Spain	.0751605	.071814	1.05
No plans	-.0967344	.086014	-1.12
Move third	.2119107	.1672781	1.27
Age at arrival	-.0222219***	.0057571	-3.86
AASQ	.0000248	.0000821	0.30
Association	.1005461*	.0462869	2.17
Romance Language	.1965677***	.0369266	5.32
Maghreb	-.0755927	.0592628	-1.28
Sub-Saharan Africa	-.1833613	.0983098	-1.87
Eastern Europe	-.0364729	.0474026	-0.77
America	-.0485933	.063912	-0.76
Asia	-.3448879***	.090552	-3.81
Oceania	.3526159	.3448797	1.02
Andalusia	-.3637603***	.0694982	-5.23
Aragon	-.3428458***	.0871274	-3.93
Asturias	-.1957487	.1132833	-1.73
Balearic Islands	-.0542341	.069078	-0.79
Canary Islands	-.1806132	.0988471	-1.83
Cantabria	.0787709	.1103036	0.71
Castile-Leon	.0636323	.0882614	0.72
Castile-La Mancha	-.050073	.0878244	-0.57
Catalonia	-.0698668	.068124	-1.03
Valencian Community	-.1213282	.0672932	-1.80
Extremadura	-.5267494***	.1001439	-5.26
Galicia	-.4219609***	.1017048	-4.15
Murcia	-.4028271***	.0864417	-4.66
Navarre	-.0148627	.0779256	-0.19
Basque Country	-.0982313	.0994921	-0.99
La Rioja	-.3132154**	.0950379	-3.30
Adjusted R ² Likelihood		.2591	
Sample		3419	

* $p < .05$ ** $p < .01$ *** $p < .001$

Appendix 2a: The determinant of Spanish language proficiency for male immigrants including interaction terms between origin and education (OLS)

<u>Variable</u>	<u>Coefficient</u>	<u>Std. Err.</u>	<u>t-statistic</u>
Years s/Migration	.0461***	.00393	11.72
YSMSQ	-.0005***	.00006	-8.30
Education	.0336***	.00737	4.57
Study Spain	.1168*	.05899	1.98
Children	-.0686	.03632	-1.89
No plans	.1031	.08837	1.17
Stay in Spain	.1725*	.07769	2.22
Move third	.1284	.16345	0.79
Age at arrival	-.0139*	.00585	-2.38
Age arrival SQ	-.00008	.00008	-1.03
Association	.1456**	.04702	3.10
Romance	.1949***	.04119	4.73
Maghreb	-.5359***	.10663	-5.03
Sub-Saharan Africa	-.7653***	.14613	-5.24
Eastern	-.3912**	.13665	-2.86
America	-.0073	.23616	-0.03
Asia	-.5077**	.18230	-2.78

Oceania	1.7518	1.41316	1.24
Maghreb*Edu	.0353***	.00932	3.79
Subsaharan*Edu	.0426**	.01357	3.14
Eastern*Edu	.0220	.01157	1.90
America*Edu	.0082	.01929	0.43
Asia*Edu	.0179	.01541	1.17
Oceania*Edu	-.1419	.12079	-1.18
Constant	2.6023	.15566	16.72
Adjusted R ² Likelihood		0.2460	
Sample		3349	

* p<.05 **p<.01 ***p<.001

Appendix 2b: The determinant of Spanish language proficiency for female immigrants including interaction terms between origin and education

Variable	Coefficient	Std. Err.	t-statistic
Years s/Migration	.0476***	.00375	12.70
YSMSQ	-.0005***	.00006	-9.34
Education	.0315***	.00659	4.79
Study in Spain	.0631	.05600	1.13
Children	-.1549***	.03603	-4.30
No plans	-.0887	.08723	-1.02
Stay Spain	.0449	.07284	0.62
Move third	.2467	.16918	1.46
Age at arrival	-.0174**	.00576	-3.03
Age arrival SQ	-.00005	.00008	-0.66
Association	.0988*	.04692	2.11
Romance	.1805***	.03653	4.94
Maghreb	-.4956***	.10202	-4.86
Sub-Saharan Africa	-.3802	.20173	-1.89
Eastern	-.2568*	.12589	-2.04
America	-.0749	.16494	-0.45
Asia	-.5172*	.25914	-2.00
Oceania	.0354	2.45795	0.01
Maghreb*Edu	.0371***	.00970	3.83
Subsaharan*Edu	.0208	.02010	1.04
Eastern*Edu	.0196	.01031	1.90
America*Edu	.0029	.01412	0.21
Asia*Edu	.0168	.02166	0.78
Oceania*Edu	.0233	.18117	0.13
Constant	2.8783	.14778	19.48
Adjusted R ² Likelihood		0.2373	
Sample		3419	

* p<.05 **p<.01 ***p<.001

APPENDIX 3

Dependent variables

Spanish language proficiency (SPANLANG)

SPANLANG is an ordinal variable on a four level scale: needs to improve, sufficient, good and very good.

Explanatory variables

Age at arrival

Age is available in single years. Age at arrival was obtained by subtracting years of residence in Spain from their age at the time of the interview.

Years since migration (YSM)

The survey contains the question: *In what year did you arrive in Spain?* We subtract that number from the year of the survey-2007 to obtain the number of years lived in Spain. We form a continuous measure and add a quadratic form to capture decreasing returns to years of residence in Spain.

Education (EDU)

This variable records the highest education level that the respondent has achieved. It is constructed with seven different levels: no formal education, incomplete primary education, primary education, first cycle of secondary education, second cycle of secondary education, first cycle of tertiary education, second cycle of tertiary education. Following common practice, years of schooling have been measured as follows for the regression analysis: 0 years for individuals without any formal education; 3 years for incomplete primary education; 6 years for primary education; 10 years for completing lower secondary education; 12 years for completing higher secondary education; 15 years for the first cycle of university education; and 17 years for the second cycle of university education.

Education in Spain (EDUSPAIN)

This variable is derived from the following questions: *Did you finish your education in Spain?* We will construct a dichotomous variable and I expect a positive impact on Spanish language proficiency.

Children (CHILD)

This is also a dichotomous variable set to one if the respondent has any children.

Association

This is a dichotomous variable that is derived from the following question: *Do you participate in any of the following groups or associations that are not specifically dedicated to foreigners?*

Romance language background (ROM)

This variable is defined as being born in a non-Spanish Romance-language country: Brazil, Portugal, Italy, Romania and France.

Stay in Spain, No plans, Move third

These are dummy variables set to one for those planning stay in Spain, move to third country or have no specific plans. The reference category is that for those planning to return to their home country in the following five years and zero for the rest.

Region in Spain (CCAA)

The following autonomous communities are included for the regional-language analysis: Navarre, Catalonia, Valencia, Balearic Islands, the Basque country and Galicia. Catalonia is used as a benchmark group since it is the main language used by the regional government and local administration.

Region of Birth

Dichotomous variables are included for region of birth with Western Europeans as a benchmark. Maghreb, Eastern Europe, Asia, Oceania, Sub-Saharan Africa and America (including North, Central and South) are the other categories.

BIBLIOGRAPHY

Aja E., Carbonell, F., Funes J. and Vila I. (2000). La inmigración extranjera en España. Los retos educativos. *Colección Estudios Sociales, No.1, La Caixa.*

Aja E. and Arango J. (2006). Veinte años de inmigración en España. Perspectivas Jurídica y Sociológica (1985-2004). *Fundació CIDOB (eds.)*

Borjas, G. (1989). Economic Theory and International Migration. *International Migration Review, Vol.23, No.3.*

Borjas, G. (1994). The Economics of Immigration. *Journal of Economic Literature, Vol.32, No.4, pp. 1667-1717.*

Borjas G (1999) The economic analysis of immigration. In: Ashenfelter O, Card D (eds) Handbook of labour economics, vol. 3A. North Holland

- Charrete, M. and Meng, R. (1994). Explaining language proficiency. Objective versus self-assessed measures of literacy. *Economics Letters* 44, pp. 313-321.
- Chiswick, B. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal of Labor Economics*, Vol.9, No.2, pp.149-170.
- Chiswick, B. (1998a). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics*, Vol.11, pp. 253-271.
- Chiswick, B. (1998b). English Language Fluency among Immigrants in the United States. *Research in Labor Economics*, Vol.17, pp. 151-200.
- Chiswick, B. (1999). Are Immigrants Favorably Self-Selected? *The American Economic Review*, Vol.89, No.2, pp.181-185.
- Chiswick, B. (2008). The Economics of Language: an introduction and overview. *IZA Discussion paper No. 3568*.
- Chiswick, B. and Miller, P. (1994). Language Choice Among Immigrants in a Multi-Lingual Destination. *Journal of Population Economics*, 7(2), pp. 119-131.
- Chiswick, B. and Miller, P. (1994). Language and Labor Supply: The Role of Gender among Immigrants in Australia. *Research on Economic Inequality*, Vol.5, pp.153-189.
- Chiswick, B. and Miller, P. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, Vol.13, No.2, pp.246-288.
- Chiswick, B. and Miller, P. (1996). Ethnic Networks and Language Proficiency Among Immigrants. *Journal of Population Economics*, 9(1), pp.19-35.
- Chiswick, B. and Miller, P. (1998). Language Skill Definition: a Study of Legalized Aliens. *International Migration Review*, 32(4), pp. 877-900.
- Chiswick, B. and Miller, P. (1999). Language Skills and earnings among Legalized Aliens. *Journal of Population Economics*, Vol.12, pp.63-89.
- Chiswick, B. and Miller, P. (2001). A model of destination-language acquisition: application to male immigrants in Canada. *Demography*, Vol.38, No.3, pp.391-409.
- Chiswick, B. and Miller, P. (2002a). Do enclaves matter in immigrant adjustment?. *IZA discussion paper No. 449*.
- Chiswick, B. and Miller, P. (2005). Linguistic Distance: a Quantitative Measure of the Distance between English and other Languages. *Journal of Multilingual and Multicultural Development*, 26, pp. 1-11.
- Chiswick, B. and Miller, P. (2007a). Modeling Immigrants' language skills. *IZA discussion paper No. 2974*
- Chiswick, B., Miller, P., and Lee, Y.L. (2002) Family Matters: The role of the family in Immigrants' destination language acquisition. *IZA Discussion paper No. 460*.

Chiswick, B., Miller, P., and Lee, Y.L. (2005). Parents and Children Talk: English Language Proficiency within Immigrant Families. *Review of Economics of the Household*, 3, pp.243-268.

Chiswick, B., Beenstock, M., and Repetto, G. (2001). The Effect of Linguistic and Country of origin on Immigrant Language Skills: Application to Israel. *International Migration*, 39(3), pp.33-60.

Chiswick, B. and Repetto, G. (2001). Immigrants Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and the Effects on Earnings. In Slobodan Djajic (ed.), *International Migration: Trends, Policy and Economic Impact*, New York: Routledge, 2001, pp. 204-228.

Chiswick, B. and Wenz, M. (2006). The Linguistic and Economic Adjustment of Soviet Jewish Immigrants in the United States, 1980-2000." *Research in Labor Economics*, Vol.24, pp. 179-216.

De la Rica, S. and Ortega, F. (2009). Economic and Cultural Gaps among Foreign-born Minorities in Spain. *IZA Discussion Paper No. 4115*.

De Palo, D., Faini, R., and Venturini, A. (2007). The social assimilation of immigrants. *Social Protection Discussion Paper No 0701*.

Dustmann, C. (1997). The Effects of Education, Parental Background and Ethnic Concentration on Language. *The Quarterly Review of Economics and Finance*, Vol.37, pp.245-262.

Dustmann, C. (1999). Temporary Migration, Human Capital, and Language Fluency of Migrants. *Scandinavian Journal of Economics*, 101, pp.297-314.

Epenshade, T., and Fu, H. (1997). An Analysis of English Language Proficiency among U.S. Immigrants. *American Sociological Review*, Vol. 62, No.2, pp. 288-305.

Herm A. (2008) Population and social conditions. *Eurostat 98/2008*.

Hernández M.T. and Villalba F. (2008). Immigrants in Spain: sociolinguistic issues. *International Journal of Sociology and Language* 193/194, pp. 177-190.

Izquierdo, M., Lacuesta, A. and Vegas R. (2009). Assimilation of Immigrants in Spain: A Longitudinal Analysis. *Banco de España, Documento de Trabajo 0904*.

Kossoudji, S. (1988). English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men. *Journal of Labor Economics*, Vol.6, No.2, pp.205-228.

Mayda, A.M. and Patel, K. (2006). International Migration Flows: The Role of Destination Countries' Migration Policies. *Working Paper 2006-3 Center for Retirement Research*.

Münz, R. (2007) Aging and Demographic Change in European Societies: Main Trends and Alternative Policy Options. *Social Protection Discussion Paper No 0703*.

OECD (2008), International Migration Outlook, Annual Report 2008, *OECD, Paris*.

- Ortega Masagué, C. (2005). La situación de los inmigrantes en España: Un análisis descriptivo. *Documento de trabajo, 2005-08*.
- Psacharopoulos, G. (1995). The profitability of Investment in Education: Concepts and Methods. *Human Capital Working Paper 15280, World Bank*.
- Reher, D.S. et al. (2008). Informe Encuesta Nacional de Inmigrantes (ENI-2007). *Documento de Trabajo, Instituto Nacional de Estadística 2/08*
- Rivera-Batiz, F. (1990). English Language Proficiency and the Economic Progress of Immigrants. *Economic Letter, Vol.34, pp.295-300*.
- Rivera-Batiz, F. (1992). English Proficiency and the Earnings of Young Immigrants in the U.S. Labor Markets. *Review of Policy Research, Vol.11. (2) pp.165.175*
- Rivera-Batiz, F.(1996). "English Language Proficiency, Quantitative Skills, and the Economic Progress of Immigrants." pp. 60-76 in *Immigrants and Immigration Policy: Individual Skills, Family Ties, and Group Identities*, edited by H. Duleep and P. Wunnava. Greenwich, CT: JAI Press.
- Rivera-Batiz, F. (2007). Migration: The Globalization of International Labor Flows, its Causes and Consequences. *Working Paper prepared for the World Bank Trade Policy Executive Program*
- Sandell, R. (2008) A social Network Approach to Spanish Immigration: An Analysis of Immigration into Spain 1998-2006. *Fedea working paper 2008-33*.
- Service, E. (1993). Differences between Young and Older Adults in Learning a Foreign Vocabulary. *Journal of Memory and Language, Vol.32, pp.608-623*
- Shields, M. and Price, S. (2002). The English language fluency and occupational success of ethnic minority immigrant men in English metropolitan areas. *Journal of Population Economics, Vol.15, pp.137-160*.
- Tainer, E. (1988). English Language Proficiency and the Determinants of Earnings among Foreign-Born Men. *The Journal of Human Resources, Vol.23, No.1, pp.108-122*.
- Zimmermann, K. (1995). Tackling the European Migration Problem. *The Journal of Economic Perspectives, Vol.9, No.2, pp.45-62*.
- Zimmermann, K. (2004). European Labour Mobility: Challenges and Potentials. *IZA Discussion Paper No. 1410*.