

# Foreign languages for the labor market: an analysis of the role of compulsory education in Europe

*Daniele Mazzacani*

Faculty of Economics and Management, Free University of Bozen-Bolzano  
daniele.mazzacani@economics.unibz.it

*Abstract: It is well known that proficiency in languages is important for the labor outcomes of natives, and the economic literature generally shows positive effects for those able to use multiple languages in the domestic labor market. In this context, compulsory education is likely to play an important role in identifying to what extent additional languages matter for the native workforce. Indeed, institutional education systems are often the main providers of individual skills in a country, including language skills, and compulsory education is reasonably unrelated to individual characteristics affecting choices of language acquisition. However, while some studies on co-official languages and labor in multilingual countries focus on compulsory schooling, it seems that no studies have yet been made on foreign languages and compulsory schooling. As a first step for future analyses on foreign languages and labor in Europe, in this paper I try to analyze whether compulsory education affects the foreign language proficiency of European adult natives. I find that being taught foreign languages during compulsory schooling has a positive effect of between 3 and 5 percent on the probability of knowing them.*

*Keywords: education; compulsory education; labor; labor market; foreign language; language learning.*

This study is based on data from the Eurostat Adult Education Survey (AES) 2007 and 2011. The responsibility for all conclusions drawn from the data lies entirely with the author.

# 1. Introduction

There is a growing belief that multiple language knowledge is an increasingly important asset in domestic labor markets. Existing research on economics and languages generally confirms this perception, and finds positive returns for native workers who can speak more languages. A number of studies dealing with co-official languages in multilingual countries highlight the crucial role of formal education, and particularly compulsory education, in providing skills in those languages to native workers. By contrast, however, to our knowledge no studies have assessed the impact of the education system on native workers' proficiency in "proper" foreign languages even though the "foreign" nature of these languages seems to make them particularly suitable for learning through formal education, especially in Europe. Preparing the ground for future analyses on language and labor outcomes, the present study concentrates on evaluating the effectiveness of European education systems at building the foreign language skills of their native adults. The basic idea is that, if the education of individuals is mostly supplied through formal education, the same should hold true for their language learning. In Europe, this assumption is underpinned by the fact that foreign languages have been part of the school curricula of most countries for several decades, even though the languages taught and the length of time they are taught may differ. Against this background, I try to make use of the role of compulsory education to investigate whether European education systems contribute to the foreign language knowledge of their citizens. Results seem to suggest that compulsory education does play a role, albeit limited.

The remainder of the paper proceeds as follows: subsection 1.1 frames the research question in the related literature; section 2 introduces the model and the identification strategy; section 3 presents the full database and the restricted sample, derived in accordance with the empirical approach, as well as some descriptive statistics; section 4 discusses the main results, and section 5 concludes.

## 1.1 Related Literature

As mentioned, the importance of competences in additional languages for native workers is already recognized by the economic literature. Various empirical studies stress that knowing languages other than mother tongue is a valuable asset on national labor markets and can influence individual labor outcomes. However, the concept of "additional language" is quite general, and a better specification enables two partially different research lines to be identified in the literature. In fact, the second languages of interest can be either *co-official*

languages in multilingual countries or regions, where they are non-native for a part of the population, or *foreign languages*, and therefore neither native nor official in a country or any of its regions. Much of the existing research focuses on the former, how “other” official languages contribute to the economic outcomes of native workers in several national and regional labor markets.<sup>1</sup> The languages that have been studied range from the long-standing, autochthonous ones, like French in Canada or Catalan in Spain, to the ones more recently inherited by countries that were former colonies of Europe or the United States.<sup>2</sup> Most of the papers specify individual labor outcomes as monthly or hourly wage differentials (Albouy, 2008; Angrist & Lavy, 1997; Azam, Chin, & Prakash, 2013; Cattaneo & Winkelmann, 2005; Christofides & Swidinsky, 2010; Grin & Sfreddo, 1998; Shapiro & Stelcner, 1997). Some research also looks at additional outcomes, such as the probability of employment, working hours and occupational status (Borooah, Dineen, & Lynch, 2009; Cappellari & Di Paolo, 2018; Chakraborty & Bakshi, 2016; Rendon, 2007). On the other hand, relatively few papers deal with indigenous workforce and foreign languages, and almost all of those that do analyze the effect of language learning on wages, within or between European countries (Di Paolo & Tansel, 2015; Ginsburgh & Prieto-rodriguez, 2011; Ginsburgh & Prieto-Rodriguez, 2007; Stöhr, 2015; Williams, 2011). Only one institutional report (Araújo, Costa, Flisi, & Calvo, 2015) and one article (Donado, 2017) explore the relationship between foreign languages and the employability of workers who speak them in Europe.

Importantly, the existing studies also emphasize the major role of *formal* education – namely, education provided by schools, colleges, universities and other institutional systems – in the process of language learning. A number of the studies mentioned above include the education system in their analyses. In their seminal paper, Angrist and Lavy (1997) use the “Arabization” of middle and secondary schooling in Morocco, which changed the language of instruction from French to Arabic, to measure the effects of loss of fluency in French on wages. Along similar lines, Chakraborty and Bakshi (2016) estimate the premium associated to English in India after a language policy intervention at the state level that abolished its teaching as a subject in public primary schools. Thanks to

---

1 Among the most studied regions and countries there are Quebec and the whole of Canada (e.g. Albouy, 2008; Christofides & Swidinsky, 2010; Shapiro & Stelcner, 1997), Switzerland (Cattaneo & Winkelmann, 2005; Grin & Sfreddo, 1998) and Catalonia (Cappellari & Di Paolo, 2018; Rendon, 2007).

2 The extent to which colonial languages are part of the linguistic environment of former colonies or not is an issue that goes beyond the scope of this paper. Here I include the studies on former colonies in the “co-official languages” group because of the sociolinguistic differences between languages with an official past in a country, and which still play today an important role in its administration and business (officially or *de facto*), and languages that are less related (i.e., foreign) to that country, which have never had such status or prominence.

the federal structure of the country and the detailed data at the district level, the authors also built a control group from states not subject to this policy change in the same period. In Europe, two studies on Catalonia and Catalan build on the crucial role of formal education. Although to varying degrees, both focus on the Language Normalization Act of 1983, which promoted the use of Catalan through the education system of the autonomous community, to investigate the effects of knowing the language on employment (Rendon, 2007) or earnings, working hours and occupational status (Cappellari & Di Paolo, 2018). Regardless of the geographic areas and languages they analyse, all these studies describe the key role of formal education in building effective identification strategies, and generally find that skills in co-official languages have positive effects for native workers. The two basic ideas underpinning all these studies are that: (i) the education of individuals is largely supplied by national or regional education systems, and the same should hold true for their language instruction, and (ii) institutional changes occurring in the system at the *compulsory* education level are mandatory, and thus reasonably unrelated to unobservable individual characteristics that could potentially affect comparisons. Moreover, natives of a country or a region share common institutional and cultural backgrounds that further reduce the potential heterogeneity of learning channels (e.g., compared to individuals with migratory backgrounds), strengthening the presumption that formal education systems may be one of the main providers of individual skills, including language competences.

However, all of the above studies deal with co-official languages that are part of the cultural heritage of multilingual countries, in most of which there has been a change in the language medium of instruction of the national or regional education systems. To the best of my knowledge, none of the studies on native workers analyses the role of education in producing skills in foreign languages, although they lend themselves particularly well to such an approach. Indeed, foreign languages are extraneous to the linguistic environment of a country and its related identities, they enjoy neither official status nor privileged positions, and in business or society they do not dominate over domestic languages. All these features make foreign languages a form of human capital that is likely to be largely acquired through formal education, even during the compulsory period, and so can benefit the strong points mentioned above on which an identification strategy can be built. Like any investment in human capital, language learning comes at the cost of time, resources, and alternative skills. Thus, a better understanding of how education systems provide language skills and the impact

that these have on labor outcomes could usefully inform individual and policy decisions on this issue.

As a first step to subsequent analyses on foreign language knowledge and labor outcomes, this paper addresses the first part of the problem and seeks to evaluate whether formal education influences the language competences of European adult natives and, if it does, to what extent. Assuming that the education system is the major supplier of this kind of skill is plausible in Europe, where foreign languages have been part of the school curricula in most countries for several years, and where over two-thirds of individuals report having learned foreign languages through lessons at school (European Commission, 2012).<sup>3</sup> Previous studies on the economic value of foreign languages in Europe or individual European countries do not try to estimate and use the impact of the national education system on foreign language proficiency because of limitations in data or reliable sources of exogenous variation. Here I try to overcome this problem by using the data in the study by Aparicio Fenoll and Kuehn (2017) on compulsory schooling and migration, which provides information on the introduction of foreign languages in a number of European education systems over the last decades. I use these data to compare the language proficiency of natives in a country exposed or not exposed to foreign languages in compulsory schooling with the language proficiency of natives of different countries, exposed or not exposed to language teaching because of the different education policies in different countries. The identification strategy and the model used are better detailed in the following section.

## 2. Empirical Approach

Foreign languages were already being taught in some European states in the 19<sup>th</sup> century, although they were circumscribed to secondary education and the small fraction of the population who had access to it. As from the middle of the 20<sup>th</sup> century, they started to be taught by the mass education systems throughout the continent, although with considerable differences between countries (Eurydice, 2001). This analysis focuses on the second half of the 20<sup>th</sup> century when foreign language teaching became increasingly common in the primary education of European natives. For this reason, it would seem that investigating the relationship between educational level and language proficiency would be sufficient to shed some light on the role of the former in acquiring the latter. However, regardless of the aforementioned differences in implementation, building a multiple regression

---

<sup>3</sup> A total of 68% of Europeans say they have learned foreign languages at school, making formal education the most common language-learning channel in Europe (European Commission, 2012: 8). Although respondents could give multiple answers, other channels appear to be used much less.

of foreign language knowledge on the level of education may be misleading, despite the controls added. Unobservable individual characteristics (such as motivation and talent) may influence the propensity to acquire additional levels of education and foreign language skills, which are themselves a product of the education system, hence generating an omitted variable bias that would alter the resulting estimates.

Linking my work to what has been done in the literature on multilingual countries and co-official languages, in this paper I propose to address the problem from the perspective of foreign language courses in compulsory schooling systems in Europe. Indeed, while individual choices of language acquisition through education are affected by heterogeneity, its teaching in *compulsory* education is determined at the government level and requires all pupils to study the language. Therefore, this policy produces an exogenous variation in language proficiency that is reasonably unaffected by selection problems related to education. To build my identification strategy, I draw on detailed data from the work of Aparicio Fenoll and Kuehn (2017) on compulsory schooling laws and migration in Europe. In their effort to investigate how schooling laws affect the propensity to migrate, authors see being taught foreign languages during compulsory education as a critical factor, since individuals exposed to a foreign language during compulsory schooling are more likely to move to countries where that language is native.<sup>4</sup> To determine which subjects may have been exposed to compulsory language teaching, they collect information on national reforms introducing foreign languages in a number of European education systems over the last fifty years or so of the twentieth century, and the languages offered. In the same spirit as Aparicio Fenoll and Kuehn (2016, 2017), I consider compulsory language teaching as exogenous to individual heterogeneity and apply a model to reliably identify its effects.

Data on language reforms are for individual countries and include the languages taught and the year they were introduced. Reforms are the result of specific national contexts and different education policies, so they differ between countries in terms of the time of introduction and the languages chosen. Importantly, in each country the reforms are implemented at a specific stage of the schooling system and from a specific year onwards, so the first cohort (i.e., the birth year) affected can be determined from the structure of the national system. Because these data are available in the database on European natives, I can know whether individuals born in a particular year in a particular country were exposed to compulsory language teaching. The differences in the years and

<sup>4</sup> As Aparicio Fenoll and Kuehn state in their paper, the relevance of compulsory language teaching for migration choices already emerged in one of their previous studies on foreign language proficiency and migration (2016).

stage of schooling that languages were introduced lead to different birth cohorts being treated, which generates the variations in identification. In fact, in each European state, the reform determines a *within-country variation* between the cohorts of individuals who were or were not exposed, identified as the treated and non-treated group, respectively. However, a comparison based only on these two groups would not pick up differences among generations and could lead to biased results. To mitigate these effects, I incorporate in the analysis the *cross-country variation* generated by the different years in which language learning was introduced in different European countries, which makes it possible to compare treated and non-treated individuals for the same birth cohort. The latter are used as a control group for the former. Unfortunately, due to limitations of the data available, in this study I cannot isolate the effect of compulsory teaching on proficiency at the single-language level. As said, in the data on reforms, the sets of foreign languages introduced differ between countries. Moreover, while some states impose a specific language, others allow their pupils to choose one language from a given group. Lastly, the database on natives to which the identification strategy applies records the languages spoken by respondents at the time of the survey, but it does not provide information on where they acquired them. Therefore, in this setting, it is not possible to ascribe language skills to compulsory teaching for the single foreign languages of interest. To overcome this issue, I pool all the foreign languages introduced in compulsory schooling by the different countries: so, what I evaluate in this paper is the *overall effect* of teaching a set of foreign languages during compulsory education on proficiency in the same set of languages. My empirical analysis focuses on the foreign languages introduced by the countries that form the restricted sample discussed in section 3: namely, English, French, German and Spanish. These are four of the five most widely taught and spoken foreign languages in the European Union (European Commission, 2012), which confirms the relevance of this analysis in the European context.<sup>5</sup> In order to find the effects of interest, I estimate the following linear probability model:

$$l_{sict} = \beta' x_{ict} + \theta T_{bc} + \gamma E_{ict} + \delta_c + \mu_b + \eta_t + \varepsilon_{ct} \quad (1)$$

On the left-hand side of equation 1,  $l_{sict}$  is a dummy variable for the foreign language skills of individual  $i$  of country  $c$  in survey year  $t$ , which takes value 1 if the respondent reports to know one of the four languages considered, and 0 otherwise. On the right-hand side, dummy variable  $T_{bc}$  indicates exposure to

<sup>5</sup> The fifth most widespread foreign language in the EU is Russian, which was not introduced by any of the countries examined in this study. Moreover, in these countries, only 0.2 % of respondents report knowing Russian as a foreign language.

language classes during compulsory schooling, which varies depending on country  $c$  and cohort (birth year)  $b$ ,  $E$  represents the highest level of formal education attained by the individual and  $x$  is a vector of additional control covariates (gender, age and its square). The model also includes fixed effects for country ( $\delta_c$ ), birth cohort ( $\mu_b$ ) and survey year ( $\eta_t$ ), respectively. Errors are clustered at the level of country-survey year ( $\varepsilon_{ct}$ ).

The data on native Europeans used in this analysis are from the survey years 2007 and 2011, thus encompassing the global financial crisis of 2008 that affected European economies and their labor markets, shrinking economic activity and considerably increasing unemployment rates. The crisis may have impacted on the acquisition of individual language skills through complex dynamics that operate outside formal education at two different levels. At the generational level, it may have had different impacts on cohorts of people who share common variations over time because they were born in a given year. The crisis may have influenced their job expectations in different ways and changed their propensity to invest in foreign languages. For instance, younger people in the most affected countries may have considered emigrating, so increasing their commitment to learning foreign languages, while the older ones may have had fewer incentives to learn other languages. Alternatively, older and less educated individuals could have chosen to acquire language skills, to get out of unemployment or to find a better job. Similar dynamics related to the crisis may also have operated at the country level, again influencing individual incentives to invest in languages. For example, by raising (or reducing) firm orientation towards foreign trade, therefore increasing (decreasing) demand for foreign language skills in the domestic labor market. The linguistic information in the European database is on adults who reported their competences at the time of the survey. For reasons such as those discussed above, the economic crisis that occurred between the two survey years could have affected the acquisition of individual language skills by adult individuals outside formal schooling. Hence, if the effects of the crisis were not taken into account, their influence on the language skills in the data could bias the estimate of the treatment coefficient. In an attempt to address these confounding factors, I expand the model including two fixed-effect interactions, reported in equation 2:

$$S_{ict} = \beta' x_{ict} + \theta T_{bc} + \gamma E_{ict} + \delta_c + \mu_b + \eta_t + \eta_t \cdot \delta_c + \eta_t \cdot \mu_b + \varepsilon_{ct} \quad (2)$$

where the interaction term  $\eta_t \cdot \delta_c$  accounts for the economic cycles in different countries, and  $\eta_t \cdot \mu_b$  allows the model to control for the different status of the cohorts in each survey year.



Finally, there are a couple of additional points to bear in mind when considering model estimates, in both specifications. First, the data do not contain information on whether natives have *actually* studied foreign languages during compulsory schooling. For this reason, the regressor of interest  $\tilde{T}$  represents the *potential* exposure to the treatment, and the estimates presented in section 4 below should be interpreted as the Intention to Treat (ITT) effects of the policy. Second, in a cross-country analysis such as the one proposed here, clustering error terms reflect the idea that observations from the same country and year of survey may share unobserved components.

The empirical approach outlined above is applied to a subset of countries in the dataset providing the information on European natives. The section below introduces the Eurostat dataset, shows how it reduces to the restricted sample and discusses its characteristics.

### 3. Dataset and Descriptive Statistics

The analysis on European countries and adult natives is based on the Eurostat Adult Education Survey (AES) and pools two cross-section waves, from 2007 and 2011. The pooled data originally consisted of about 400,000 adult individuals, although several countries were not suitable for the analysis because of their intrinsic characteristics or the lack of necessary variables. In fact, after selection, the restricted sample amounts to around a quarter of the original observations. Besides standard sociodemographic variables, the AES offers information on the highest level of completed education (ISCED-97 codes) and a quite detailed linguistic profile of the respondents, including their mother tongue and other non-native languages spoken. In particular, it records the two best-known international languages *other than mother tongue* and their level. All the linguistic variables are self-reported, and knowledge of international languages is ranked on a three-level evaluation grid (basic, sufficient and good). The AES covers almost 40 international languages but, as indicated in section 2, this study focuses on four of the five most taught international languages in Europe (English, French, German and Spanish – hereinafter also “most taught” languages). A first look at the full dataset seems to confirm it is worth concentrating on them: six out of ten individuals claim to speak one international language other than their mother tongue and for seventy-four percent of these it is one of the four most taught ones. On the other hand, four out of ten Europeans in the pooled data speaks no international language at all. And fewer than three in ten respondents know a second international language. Again, more than seventy percent of them speak a “most taught” language.

However, it is crucial to note that the concept of *international* language used in the database does not necessarily coincide with the one of *foreign* language to which this analysis refers. According to Eurostat, in the AES data the term “international language” applies to every language spoken in more than one country. In European bi- or multilingual states, like Belgium or Finland, all the official languages have long been included in compulsory education, but not all of them are necessarily viewed as mother tongues by all speakers. In these countries, formally co-official (i.e., non-foreign) languages could then be perceived and reported as international languages, making it difficult to allocate the treatment only to actual foreign languages, and confounding different effects in the estimates. To avoid this, I exclude from the analysis all the countries that are officially multilingual in more than one international language.

Given the aim of the present study, the restricted sample is limited to native individuals of the European countries discussed. To reduce the confounding effects on foreign language skills of recent immigration, natives are regarded only as those individuals born in a country and citizens of the same country. In other words, considering a generic country *C*, a respondent in the dataset must be both born in *C* and a citizen of *C* to be identified as a native. Focusing on natives in this way excludes those states that do not record country of birth, citizenship or both. I also limit the study to individuals aged between 25 and 64 at the time of the survey. I do this to exclude younger respondents, who could still be students and hence more likely to be subject to formal education in foreign languages, and the older ones, who presumably had less need to know foreign languages. Besides discarding the observations that fall outside that age range, the few countries that partially anonymize the age variable are also removed. Furthermore, in the 2007 data a number of countries lacked crucial variables for this analysis, and were therefore excluded from the sample.

Finally, the history of foreign language teaching in Central and Eastern European states is particularly complex, which affects their eligibility for this analysis. In almost all of these countries, Western European languages were systematically introduced into the education systems around the mid-1940s. However, with the end of the war and their entry into the Soviet sphere of influence, Russian became the first foreign language until the fall of the Berlin Wall. Nonetheless, in the same decades, many countries allowed the teaching of Western foreign languages in some geographical areas or school, and private, informal teaching spread widely to compensate for the lack of supply in the public system (Eurydice, 2001). The early introduction of Russian precedes the cohorts available in the dataset and, together with the lack of data on other foreign languages in the socialist period, excludes Central and Eastern Europe from the analysis.

**Figure 1. Countries included in the analysis**



This leaves seven countries available: Austria, Cyprus, France, Norway, Portugal, Spain and Sweden (see figure 1). For three of them – Austria, France and Spain – the official language of the country coincides with one of the most taught international languages. Despite the restrictive definition of native applied above, a few individuals from these countries report the official language not as a mother tongue but as their first or second international language, thus creating “false positives” in the analyses of foreign language competences. Possible reasons for this behavior may be a remote immigration background or political reasons.<sup>6</sup> To account for this issue, I identify and exclude from the sample the limited number of observations from Austria, France and Spain reporting German, French or Spanish to be an international language. As a result, the final restricted sample includes 95,083 observations, of which 50,094 are from 2007 and 44,989 from 2011, distributed among the seven countries. As previously stated, the set of languages introduced into the compulsory education systems of these countries is composed of English, French, German and Spanish.<sup>7</sup>

<sup>6</sup> For instance, in Spain, for reasons of strong regional identities (Catalan, Basque and Galician) and the related political issues.

<sup>7</sup> In 1993, Austria added Italian to the set of available languages, along with English and French. However, this addition is recent and applies only to those born from 1985 onwards, who represent about 3 % of the Austrian sample. Of these, respondents who report Italian as their first or second international language represent only 0.2 % of the national sample. Unsurprisingly, repeating the analyses presented in section 4 with or without Italian in the set of foreign languages leaves the results unchanged.

Table 1 shows the percentages of speakers of international languages in the restricted sample, at any level of skills, splitting international languages into the ones most taught and others. About 57% of those interviewed report they can speak at least one international language, which for the 93% of them is English, French, German or Spanish. Moreover, even among those who speak a second language (about 23%), in almost nine cases out of ten it is one of the most taught ones.

**Table 1. Speakers of international languages in the restricted sample**

	No language	Most taught language	Other language	Total (%)
First international language	42.9	53.3	3.8	100.0
Second international language	77.3	19.4	3.3	100.0

Results in percentages. Total sample: 95,083 observations. The set of most taught international languages includes English, French, German and Spanish.

Table 2 goes more into detail and reports knowledge of international languages at any level of skill distributed by age, decades, and formal education completed. Panel A highlights that younger cohorts are more likely to report knowing one international language other than their mother tongue, almost always one of the most taught set. Although the second international language shows a similar distribution, its incidence in the sample is rather small, about one third of that of the first. These figures seem to suggest a positive generational trend in language learning, which may in turn depend on formal education. Along these lines, panel B depicts the relationship between knowledge of international languages and level of formal education achieved, aggregating the ISCED codes into “low”, “intermediate” and “high”. What emerges is that the more educated individuals are, the more they report speaking at least one international language. Also in panel B, the positive trend holds for both first and second international languages, but again with a far lower magnitude for the latter. The fact that younger

**Table 2. Speakers of international languages in the restricted sample by age group and by educational level****Panel A. International language knowledge among age groups**

Age group	First international language				Second international language			
	No language	Most taught language	Other language	Total (%)	No language	Most taught language	Other language	Total (%)
25-34	27.5	68.3	4.2	100.0	69.6	26.0	4.4	100.0
35-44	37.4	58.5	4.1	100.0	75.8	20.6	3.6	100.0
45-54	47.5	48.8	3.7	100.0	79.6	17.6	2.8	100.0
55-64	56.7	39.8	3.5	100.0	83.0	14.4	2.6	100.0

**Panel B. International language knowledge by completed level of education**

Level of education completed	First international language				Second international language			
	No language	Most taught language	Other language	Total (%)	No language	Most taught language	Other language	Total (%)
High	14.0	81.8	4.2	100.0	56.5	38.0	5.5	100.0
Intermediate	33.6	62.5	3.9	100.0	77.0	19.0	4.0	100.0
Low	69.3	27.1	3.6	100.0	91.6	7.1	1.3	100.0

Results in percentages. Total sample: 95,083 observations. The set of most taught international languages includes English, French, German and Spanish.

individuals also tend to be more educated (as reported in table A1 in the appendix) reinforces the intuition that formal education can be an important driver of language learning. However, relying on simple descriptive correlations in the data could be misleading. As mentioned in the previous section, both trends shown in table 2 may depend on unobserved individual heterogeneity, or be the result of other generational trends. Moreover, the table presented above does not account for differences between countries.

It should be pointed out that, in the restricted sample, the concepts of international language and foreign language coincide when I focus on the four most taught international languages. Indeed, none of them is an official language in four of the seven countries investigated, and in the three countries where they are, false positive observations are discarded. Moreover, both table 1 and table 2 confirm that in the restricted sample the four most taught languages represent the bulk of international languages spoken by natives, to an even greater extent than in the whole dataset, whereas the group of other languages shows a rather marginal weight. It is therefore unlikely that excluding the other international languages significantly affects the results. Thus, when implementing the model, the speakers of other international languages are aggregated to those who report no language, and only English, French, German and Spanish are considered to be foreign languages. Consequently, the descriptive statistics for the restricted

sample simplify and convey three main messages. First, in the seven countries analyzed, the set of foreign languages taught during compulsory schooling covers almost all non-native language skills. Second, while many individuals know one foreign language, only a small percentage of them report knowing a second one. Third, Table 2 shows a positive relationship between education and proficiency in foreign languages, which is consistent with the assumptions of the literature and confirms the importance of controlling for the level of education in the model.

The results presented in the section below focus only on the first foreign language best known by natives in the restricted sample, and exclude the second one. This is because, when languages were first taught in compulsory education, not all countries required their students to learn more than one. Therefore, any estimation of the effects of language treatment on the second foreign language would require only countries teaching two foreign languages to be examined, otherwise different policies would be compared. Regrettably, the limited number of country and survey years available prevents further investigation of a second language. Focusing the analysis on the first foreign language reported should still capture the most important effect, given the reduced weight of the second language shown in tables 1 and 2. Furthermore, the estimates evaluate the effects of compulsory language teaching only for sufficient and good levels of language proficiency. This is because, in AES data, the basic level corresponds to rather low skills, which may be more sensitive to self-assessment errors (for instance, being reported by individuals who actually have no such skills). Since this could in turn confound estimates of treatment effects, in this preliminary approach, basic skills are assimilated to the lack of language knowledge. Hence, in the restricted sample, I set the dependent variable equal to 1 for the individuals who know the first foreign language at a sufficient or good level, and 0 otherwise.

Table 3 closes the third section by providing descriptive statistics for the variables used in empirical analysis, differentiating between individuals exposed and not exposed to language classes during compulsory education. The comparison shows a considerable difference in language knowledge: 23.6 % of non-exposed subjects report knowing one foreign language at a sufficient or good level, while this percentage increases to 44.1 % for exposed subjects. As for gender, the number of women and men is almost identical in the two groups. Due to the identification strategy, individuals exposed to the reforms are on average ten years younger than those who did not receive compulsory language instruction. Finally, the treated individuals are on average better educated than their counterparts, as highlighted by a higher incidence of upper, post-secondary and tertiary levels of education in their group (ISECD levels 3, 4 and 5, respectively).

**Table 3. Descriptive statistics by exposure to compulsory language teaching**

Variables	Non-exposed	Exposed
Language skills	23.6	44.1
Female	53.1	51.9
Age	49.1 (10.2)	38.7 (9.5)
<i>Education</i>		
ISCED 1	28.6	7.2
ISCED 2	24.2	16.2
ISCED 3	23.8	40.7
ISCED 4	1.4	2.5
ISCED 5	22.0	33.5
Total (%)	100.0	100.0
Observations	58,532	36,551

Results in percentages. Total observations: 95,083. Standard deviations, reported only for continuous variables, are in parentheses. Language skills refers to the knowledge of the first best-known foreign language between English, French, German or Spanish, at a sufficient or good level. Education variables are according to the ISCED-97 codes, though here ISCED 1 also includes pre-primary education (ISCED 0), and ISCED

5 aggregates first and second stage of tertiary education (ISCED 6). Post-secondary and non-tertiary education (ISCED 4) is a marginal category in the countries analyzed.

## 4. Results

Table 4 presents the main coefficients of the regression model described in section 2. The baseline specification of the model in column 1 refers to the first equation. The rest of the table expands the model gradually: columns 2 and 3 introduce interaction terms separately, and column 4 considers them together, as in the second equation.

When not considering potential changes between the two years of survey, estimates suggest that native individuals who received language teaching during their compulsory education have around a 3.6 % higher probability of speaking a foreign language at a sufficient or good level. Accounting for the two dynamics potentially intervening between 2007 and 2011, which in section 2 I hypothesize may come from the 2008 global financial crisis, has different effects on the estimates. Introducing the interaction between country and year of survey fixed effect incorporates potential changes at the national level between the two years, but this only marginally affects the coefficient of language

reforms, and does not change its statistical significance. On the other hand, both magnitude and significance increase appreciably when the model controls for changes in cohort effects. Including the different trajectories of birth cohorts over time makes exposed natives between 5.1 and 5.2 % more likely to have sufficient or good language skills than their non-exposed counterpart do. The differences between columns 2, 3 and 4 clarifies that, in this setting, the increase in coefficient estimates is almost entirely driven by cohort effects. Regardless of the specification adopted, the empirical analysis of the restricted sample shows that being taught either English, French, German or Spanish during compulsory schooling has an impact on the language skills of adult individuals in the same set of languages. It is also worth noting that formal education other than compulsory language teaching plays a major role in acquiring foreign language skills, displaying the highest coefficient in table 4. Indeed, each additional level of education completed increases the probability of respondents reporting sufficient or good language skills in one of the foreign languages analyzed by 13 %.<sup>8</sup> Moreover, education is highly significant and not influenced by any of the interactions added to the model. Looking at the other covariates, women appear to be 1.5 % less inclined to report foreign language knowledge at least at a sufficient level, with a significant and constant estimate across the different specifications. However, in this setting, it is not entirely clear whether this difference signals a gender gap or is the result of other factors. After controlling for the birth cohort effects and their interaction with the survey years, age and its square have no relevance in explaining language competences. Lastly, a closer look at the specific foreign languages reported by individuals allows for a few qualitative considerations on which of the languages best explains the effects of language reforms. As illustrated by figure A1 in the appendix, in all the countries of the restricted sample, for all the age groups, English is by far the first foreign language best known by natives.<sup>9</sup> Percentages vary across countries, and show that English (at a sufficient or good level) is widely known as the first foreign language in the Scandinavian countries, followed by Austria and Cyprus.<sup>10</sup> In the Romance-speaking countries of southwestern Europe, English remains the most widely spoken language in a relative sense, but is not so widespread. Other

8 My study focuses on the effects of compulsory language teaching on the language skills of adult individuals, not on the effects of further teaching at other educational levels. However, it is reasonable to assume that different levels of education influence language acquisition in different ways. These relationships could be the object of future analyses.

9 The graphs also confirm, for each country, the positive generational trend in language learning emerging in table 2.

10 The spread of English as a foreign language in Cyprus is presumably linked to the British administration (between the late nineteenth and mid-twentieth century), as well as to the continuing British military presence on the island.



foreign languages play a minimal role, with percentages that are appreciable only in Portugal, Spain and France. Furthermore, all the countries in the restricted sample include English in their language set: therefore, although the data do not enable language skills to be attributed entirely to language reforms, it is reasonable to assume that English drives most of the effect of the compulsory language teaching discussed in this study.

**Table 4. Compulsory language teaching and language skills**

Variables	(1) language skills	(2) language skills	(3) language skills	(4) language skills
Compulsory language teaching	0.036* (0.018)	0.037* (0.018)	0.051*** (0.016)	0.052*** (0.015)
Female	-0.015** (0.006)	-0.015** (0.006)	-0.015** (0.006)	-0.015** (0.006)
Formal education	0.130*** (0.011)	0.130*** (0.011)	0.130*** (0.011)	0.130*** (0.011)
Age	-0.010 (0.013)	-0.009 (0.012)	0.029 (0.036)	0.029 (0.035)
Age <sup>2</sup>	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Country fixed effects	YES	YES	YES	YES
Birth year fixed effects	YES	YES	YES	YES
Year of survey fixed effects	YES	YES	YES	YES
Country <input checked="" type="checkbox"/> Year of survey	NO	YES	NO	YES
Birth year <input checked="" type="checkbox"/> Year of survey	NO	NO	YES	YES
Observations	95,083	95,083	95,083	95,083
Adjusted R <sup>2</sup>	0.343	0.344	0.343	0.344

Dependent variable: dummy for the first foreign language between English, French, German or Spanish, at a sufficient or good level. Female is a dummy variable for gender. Formal education is expressed in ISCED-97 levels. Clustered robust standard errors at the country-survey year level are reported in parentheses: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 5. Conclusions

The analysis presented in this paper suggests that, in the European countries analyzed, teaching foreign languages during compulsory schooling has a positive and significant effect on the proficiency of adult natives in the same languages. My estimates range from 3 to 5 %, depending on whether the model controls for dynamics related to the global financial crisis of 2008. This is a limited effect

compared to that of formal education completed – from a third to half of its magnitude. However, the language reforms used in the identification strategy are determined at the institutional level and, as such, they produce a variation that is reasonably exogenous to individual unobserved characteristics. The results presented here indicate that the assumptions underlying the studies on co-official languages found in the literature may be extended to the case of natives and foreign languages, making compulsory language teaching a valuable instrument to study the impact of foreign language skills on a variety of outcomes, both related and unrelated to the labor market.

## 6. References

- ALBOUY, D. (2008). “The wage gap between Francophones and Anglophones: A Canadian perspective, 1970-2000”. *Canadian Journal of Economics/Revue Canadienne d'économique*, 41(4), 1211–1238.
- ANGRIST, J. D., & LAVY, V. (1997). “The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco”. *Journal of Labor Economics*, 15(1), 48–76.
- APARICIO FENOLL, A., & KUEHN, Z. (2016). “Does Foreign Language Proficiency Foster Migration of Young Individuals within the European Union?” In M. GAZZOLA & B.-A. WICKSTRÖM (Eds.), *The Economics of Language Policy* (pp. 331–355). Cambridge, Massachusetts: The MIT Press.
- APARICIO FENOLL, A., & KUEHN, Z. (2017). “Compulsory Schooling Laws and Migration Across European Countries”. *Demography*, 54(6), 2181–2200.
- ARAÚJO, L.; COSTA, P. D.; FLISI, S., & CALVO, E. S. (2015). *Languages and Employability*. Brussel: Joint Research Centre (European Commission).
- AZAM, M.; CHIN, A., & PRAKASH, N. (2013). “The Returns to English-Language Skills in India”. *Economic Development and Cultural Change*, 61(2), 335–367.
- BOROAH, V. K.; DINEEN, D. A., & LYNCH, N. (2009). “Language and occupational status: Linguistic elitism in the Irish labour market”. *Economic and Social Review*, 40(4), 435–460.
- CAPPELLARI, L., & DI PAOLO, A. (2018). “Bilingual schooling and earnings: Evidence from a language-in-education reform”. *Economics of Education Review*, 64(February), 90–101.
- CATTANEO, A., & WINKELMANN, R. (2005). “Earnings differentials between German and French speakers in Switzerland”. *Schweizerische Zeitschrift Für Volkswirtschaft Und Statistik*, 141(2), 191–212.

- CHAKRABORTY, T., & BAKSHI, S. K. (2016). „English language premium: Evidence from a policy experiment in India”. *Economics of Education Review*, 50, 1–16.
- CHRISTOFIDES, L. N., & SWIDINSKY, R. (2010). “The economic returns to the knowledge and use of a second official language: English in Quebec and French in the rest-of-Canada”. *Canadian Public Policy / Analyse de Politiques*, 36(2), 137–158.
- DI PAOLO, A., & TANSEL, A. (2015). “Returns to Foreign Language Skills in a Developing Country: The Case of Turkey”. *Journal of Development Studies*, 51(4), 407–421.
- DONADO, A. (2017). “Foreign Languages and their Impact on Unemployment”. *Labour*, 31(3), 265–287.
- European Commission. (2012). *Europeans and their Languages. Special Eurobarometer 386*, (June).
- EURYDICE (2001). *Foreign Language Teaching in Schools in Europe*. Brussels: Eurydice.
- GINSBURGH, V., & PRIETO-RODRIGUEZ, J. (2007). “La prima salarial de las lenguas extranjeras en el mercado de trabajo español” [Foreign language wage premium in the Spanish labor market]. *Cuadernos Económicos de ICE*, (74), 129–146.
- GINSBURGH, V., & PRIETO-RODRIGUEZ, J. (2011). “Returns to Foreign Languages of Native Workers in the European Union”. *Industrial and Labor Relations Review*, 64(3), 599–618.
- GRIN, F., & SFREDDO, C. (1998). “Language-based earnings differentials on the Swiss labour market: Is Italian a liability?” *International Journal of Manpower*, 19(7), 520–532.
- RENDON, S. (2007). “The Catalan premium: Language and employment in Catalonia”. *Journal of Population Economics*, 20(3), 669–686.
- SHAPIRO, D. M., & STELCNER, M. (1997). “Language and earnings in Quebec: Trends over twenty years, 1970-1990”. *Canadian Public Policy / Analyse de Politiques*, 23(2), 115–131.
- STÖHR, T. (2015). “The returns to occupational foreign language use: Evidence from Germany”. *Labour Economics*, 32, 86–98.
- WILLIAMS, D. R. (2011). “Multiple language usage and earnings in Western Europe”. *International Journal of Manpower*, 32(4), 372–393.

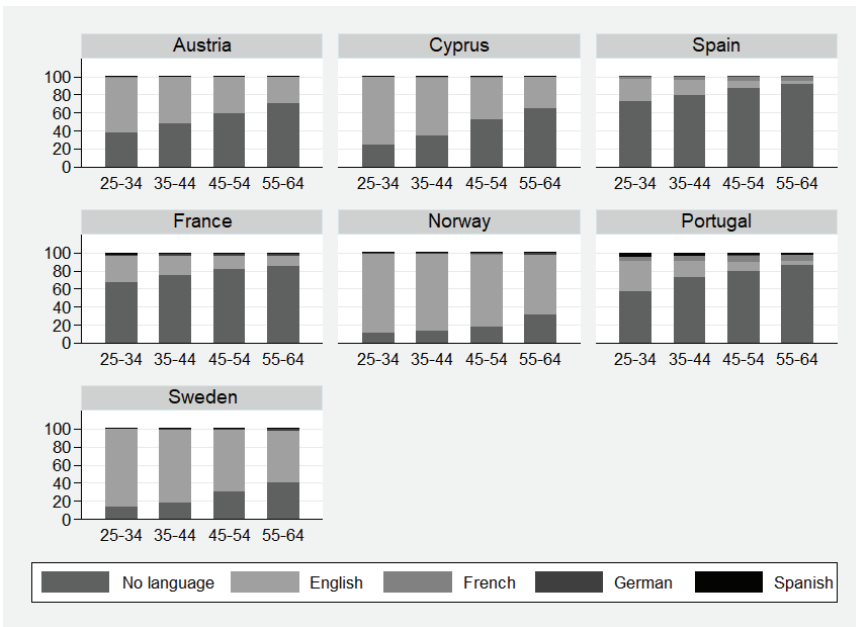
# Appendix

**Table A1. Completed level of education by age group**

Age group	Completed level of education			Total (%)
	Low	Medium	High	
25-34	26.3	32.9	40.8	100.0
35-44	35.4	32.1	32.5	100.0
45-54	45.4	31.1	23.5	100.0
55-64	56.6	25.5	17.9	100.0

Results in percentage. Total sample: 95,083 observations.

**Figure A1. First foreign language at sufficient or good level by country and age group**



For each country, the x-axis reports the population divided into four age groups. Percentages on the y-axis refer to the specific age groups.