Bilingualism, Linguistic Distance and Adult's Skills Endowment: Evidence from PIAAC

Guida Ayza Estopà
Universitat de Barcelona – UB School of Economics

Advisor: Antonio Di Paolo
Universitat de Barcelona – UB School of Economics

Master Thesis – MSc in Economics
June, 2020

Abstract. Do bilinguals outperform monolinguals on skills development? If so, is there heterogeneity in the effect of bilingualism across countries? I combine data of the PIAAC with the Levenshtein measure of linguistic distance and I estimate the effects of bilingualism in the scores gained in the literacy and numeracy tests of the PIAAC by Ordinary Least Squares. I perform the analysis using data on 7 European and post-Soviet countries with a significant percentage of a bilingual population. I find a bilingual advantage in the countries where most of the bilinguals speak the same two languages as well as a detrimental effect of a higher linguistic distance. Moreover, I exploit the Spanish case to identify if the general intuition holds for the within-country case. Results indicate that a linguistic immersion in two well-established languages can help to develop a bilingual advantage on literacy and numeracy skills if both languages do not differ sharply between them. These findings are relevant for the design of language-in-education policies since they suggest the importance of fostering bilingualism in modern societies.

Key words: Bilingualism, Numeracy, Literacy, Linguistic distance, Language-in-education policies, PIAAC.
JEL codes: I20, I21, I25, I28, J24, Z13.

UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

1. Introduction

Multilingualism is becoming a common feature in an increasingly globalized world. About 60% of the world's population is estimated to know and use two or more languages (Costa et al., 2008; Grosjean, 2010). Increasing globalization and migrant flows lead to multilingual realities. Nowadays, it is common for many people to acquire skills in languages different from their mother tongue during adulthood, but many others become bilingual during infancy because they are exposed to more than one language during the first years of life. Speaking several languages is highly valued in the academic and working world and the evidence points out that learning more than one language during childhood is easier. However, being raised in a bilingual environment not only implies being fluent in two or more languages but also developing cognitive advantages. Earlier exposure to bilingualism strengthens attention and cultural sensibility (knowing different languages often means to be aware of the existence of diverse cultural habits), it also fosters multitasking, increases communication skills, and has several mental health advantages (Bialystok et al., 2009). The rise of information and communication technology, as well as the associated increase in the demand for skills in literacy and numeracy, display a particular challenge to identify all the possible factors that can enhance them. Given that these cognitive advantages play a key role in skill acquisition, one should expect a bilingual advantage on literacy and numeracy skills. These disparities in the level of competences might affect labor market outcomes, which is ultimately translated into economic and social inequalities.

The effect of bilingualism is an intricate phenomenon. Although in cognitive science literature there is still a discussion about whether bilingualism is good, in economic literature there is increasing evidence that bilingualism has positive effects on education and labor market outcomes. Basically, what cognitive science literature finds is that it could be positive for some aspects (as cognition or monitoring) and negative for some others (as vocabulary richness or speed of response), albeit the overall effect seems to be positive. Moreover, when studying the impact of bilingualism on the individual's skills, there is no right answer or approach as all countries have different realities.

It is likely that the controversy arises from this fact. Several studies find that bilingual individuals not only outperform monolinguals on skill acquisition (Kovelman, 2008; Bialystok et al., 2009) but also prove that bilingual education has a positive effect in the labor market (Angrist & Lavy, 1977; Capellari & Di Paolo, 2018) and educational outcomes (Pearson et al., 1993; Aparicio, 2018). Nevertheless, some researchers argue that bilingualism may harm children's learning potential and educational outcomes (Hakuta, 1986; Thordardottir et al., 2006) – just the opposite approach.

Therefore, in this paper, I study whether: (i) There is an impact of early bilingualism on numeracy and literacy skills among adults, which might explain the positive effects on labor market outcomes – otherwise these beneficial outcomes would be just due to signaling. (ii) The effect of bilingualism varies by country, probably due to the diverse historical reasons for which a country has a significant proportion of bilingual

population. (iii) The effect of bilingualism depends on the dissimilarity between the two languages, which might also reconcile the controversial findings in the cognitive science field. To this effect, I use data from the *Programme for the International Assessment of Adult Competencies (PIAAC)* of the *OCDE*, which contains information about competences in literacy and numeracy of the adult population[1], as well as specific information about bilingualism during childhood and the different languages spoken. Thanks to having specific information about the languages spoken by bilingual individuals in the sample, I can combine the PIAAC data with the Levenshtein distance; a measure for the linguistic distance between different languages developed by the *German Max Planck Institute for Evolutionary Anthropology*. For more robustness, I only retain data of countries in which the percentage of a bilingual population is relatively high[2] and is not due to the presence of indigenous languages. I estimate the effect of having two native languages in the scores gained in the numeracy and literacy tests by Ordinary Least Squares. I consider a set of characteristics that can differ between individuals and could be correlated with the fact of being bilingual as well as with the competencies. Thus, apart from the basic demographic aspects (i.e. gender, age, and education), I also take into account other variables such as the parent's country of origin. Furthermore, in the regressions, I allow for the possibility of heterogeneous effects of bilingualism by country including interaction effects. In the last regression, I also include the measured linguistic distance between both languages of bilinguals in order to assess if the degree of linguistic distance influence the impact of bilingualism on skills.

It is important to take into account this heterogeneity between countries because, as argued above, bilingualism is a rather complex topic. Some factors such as the historical reasons for which in a country there is a significant bilingual population, the variety of language spoken and the number of speakers of each, as well as the linguistic distance between the natives' languages of the same individual, may affect the scope of bilingualism's impact. To the best of my knowledge, this paper is the first one to analyze the impact of bilingualism using data from PIAAC and considering this possible heterogeneous effect by country. Moreover, apart from studying the country's historical characteristics, my research considers other important confounders such as whether an individual's native languages are different from the test's language and the country's official languages. Another interesting contribution is the use of the Levenshtein distance, including this measure I allow for the possibility that bilingualism may affect skills differently when both languages are closer or further apart, which can help us figure out what originates these differences.

In the context of an increasing bilingual population and the popularization of bilingual education, it is important to understand the impact of bilingualism in order to design proper policies, to promote the positive aspects and, if any, mitigate the negatives. The countries that I use for the study present different realities and cultures. I keep seven western European and post-soviet countries[3] that have a significant proportion of a bilingual population. In them, one can find a wide range of languages, which can be more

[1] In order to avoid mixing up bilingualism with migration experiences, I only keep data of native speakers.
[2] More than a 5% of the sample.
[3] Denmark, France, Kazakhstan, Lithuania, Slovak Republic, Slovenia and Spain.

or less similar to each other, different reasons that led a society to have an important number of bilinguals in it, as well as different educational systems, policies, and socioeconomic realities.

The results of this study identify three different situations depending on the trade-off between two factors – the country's historical characteristics related to the language's presence, and the linguistic distance between both languages of bilinguals. Therefore, I observe the following combinations: (i) If the historical context is favorable to language immersion and values the linguistic diversity (this tends to occur when the most of bilinguals speak the same languages due to historical reasons) and the linguistic distance is not very large, bilingualism shows a strong positive effect. This is the case of Spain and Slovenia. (ii) If the historical context is favorable but the linguistic distance is very high, to a greater or lesser extent, bilingualism shows a modest positive effect. We find this situation in Kazakhstan and Lithuania. (iii) If the historical context is not favorable (this typically is because in a country coexist a wide variety of languages) bilingualism shows an insignificant, or even negative, impact. In this case, the higher is the distance between languages, the worse. It occurs in Denmark, France, and The Slovak Republic. I also perform a specific analysis for the particular case of Spain and, I found the same conclusion. Within the country, bilingualism of well-established languages[4] without sharply linguistic distances show a greater positive impact than the other pair of languages does.

These findings have important policy implications, particularly in terms of education. The results highlight the importance of bolstering linguistic immersion in countries where their historical context, and the nature of their languages, is favorable, as well as they suggest the possibility of including bilingual education to reinforce bilingualism early in childhood in the rest of the regions. This paper contributes to the growing literature on bilingualism as well as, the extension of literature on the economics of education, by identifying the conditions that can stimulate skills acquisition via the bilingual advantage.

The structure of the paper is as follows. I begin, in Section 2, with a literature review of the main papers that are related to this theme, as well as explaining the theory that motivates this study. In Section 3, I describe the data. In section 4 I explain the empirical strategy used. In Section 5 I present the results and in section 6 I discuss them. In section 7 I report the results of the Spanish case, and in section 8, I present the conclusions.

2. Literature Review

The literature studying the effects of bilingualism on a range of different aspects from social development and cognitive skills – as examined in psychology and neuroscience – to education and labor market

[4] This is the case of Catalan-Spanish and Galician-Spanish.

outcomes – as analyzed in economics – is extensive. Yet, despite this, the nature of these relationship remains largely unexplored.

Bilingual individuals tend to have greater earnings – nevertheless, there is substantial evidence that this wage premium is not directly related to the fact of their being fluent in two languages, but rather it seems to be attributable to other underlying motives. The earlier literature on the effects of bilingualism is largely focused on migrants learning the host country language, which has clear positive predictions. The literature on language skills among natives is more recent and there is less evidence so far. Chiswick and Miller (1995) emphasize the endogeneity between language and earnings. They estimate the effects of English fluency on wages for a sample of immigrants living in the United States that have a different language of origin. The results are statistically different when using OLS or IV techniques, which indicates that the estimation of the effects of language on earnings with OLS has a selection bias. The authors suggest that this bias is driven by schooling, experience variables, or skills. Similarly, Fry and Lowell (2003) point out that, in the United States, bilingual individuals tend to have higher wages than their monolingual counterparts, but this is because most of them also have higher educational levels. When maintaining other human capital characteristics constant, no significant differences in wages between bilinguals and monolinguals are found. In general, researchers find that bilinguals tend to perform better in school, and this may indicate that bilingualism improves academic ability and, therefore, productivity in the labor market. In essence, what the authors report is that a bilingual person will not be better paid for the fact of being fluent in two languages, but that being bilingual helps this person to develop some other valuable skills. Lang and Siniver's (2009) findings point in the same direction and allow them to claim that the ability to acquire language skills may be an indicative of the ability to acquire other skills, based on their study of the return to English fluency for Russian immigrants and native Israelis in Israel. Given the apparent positive correlation between bilingualism and cognitive skills identified in the literature, in this paper, I seek to test if this relationship holds true by examining data from literacy and numeracy tests completed by bilingual and monolingual individuals.

Studies of the effects of bilingualism on language and cognitive functioning refer to mixed results. Different trials with preschool children find that those in the process of acquiring two languages show lower levels of skill in each language than the monolingual children (Thordardottir et al., 2006; Vagh et al., 2009) and support the idea that bilingualism can harm children's learning potential because it generates confusion (Hakuta, 1986). In contrast, Bialystok et al. (2009) report that bilingualism has a significant positive impact on executive control, which they define as "the set of cognitive skills based on limited cognitive resources for such functions as inhibition, switching attention, and working memory. Executive control emerges late in development and declines early in aging, and supports such activities as high-level thought, multi-tasking, and sustained attention". The authors perform different experiments consisting of games of syllabic structures or vocabulary with 12-month-old infants who were being raised in a bilingual or monolingual environment. They found that bilingual children can learn twice as much about languages as monolingual

children in the same amount of time. The authors conclude that monolinguals have a richer vocabulary and on average are faster than their bilingual counterparts in giving answers in games or tests of this kind. However, bilinguals perform better in fluency and monitoring. Furthermore, several articles highlight the advantages of bilingualism in cognitive aspects over the lifespan – Yang et al. (2011) state that bilingualism is associated with higher executive functioning and attention in children, Costa et al. (2008) extend this conclusion to young adults, and Bialystok et al. (2012) explain that bilingual adults between 30 and 80 years old show a significant advantage in working memory and that bilingualism protects against cognitive decline in old age and postpones the onset of Alzheimer's and dementia.

There is a considerable body of evidence indicating that children learning two languages do not acquire the language aptitudes more slowly than their monolinguals counterparts. The overall language knowledge of bilingual children, combined across both their languages, is equal or even greater than monolingual children's knowledge in terms of vocabulary and grammar development (Pearson et al., 1993; Thordardottir et al.,2006; Hoff et al., 2012). Moreover, Kovelman et al. (2008) explain that exposure to two languages at an early age is positively correlated with reading, phonological awareness, and language competence in both languages. Further evidence is provided by Lleras-Muney and Shertzer (2015), who study the educational outcomes derived from the implementation of the education laws for Americanization, in the early twentieth century. These laws consisted mainly of the suppression of bilingual schooling, establishing English as the unique language of instruction. This was with the objective to improve English language skills and the sense of belonging – that is, feeling American – among immigrants. The authors do not find significant improvement in English literacy skills after switching from bilingual education to the English-only system. Another interesting article in this line is Fouka (2020), who describes the case of the prohibition of German as a language of instruction in the United States after World War I. In the earlier twentieth century, there were many German immigrants in the United States, and bilingual education in these two languages was present in many schools. After the War, some states banned German and moved from bilingual to a monolingual (English only) education. Fouka demonstrates that this had a negative effect on educational outcomes, particularly for individuals from a more German background. In short, what the evidence suggests is that even though bilingual children and adults have somewhat less rich vocabularies in each language than their monolingual counterparts, they show an advantage in cognitive control, which can be applied not only in language processing but also in facilitating acquisition of some other skills. Bilingualism reorganizes specific brain networks enhancing executive control and promotes better cognitive performance throughout the lifespan, even in non-linguistic domains. Bilinguals possess advantages in "mental flexibility", the ability to adapt to ongoing changes and process information efficiently and adaptatively (Peal & Lambert, 1962), a characteristic that promotes general learning and which is highly appreciated in the job market.

The motivation for the current research arises from the evidence presented above: If bilinguals have a cognitive advantage that facilitates learning processes beyond language, then it is reasonable to think that

we should be able to identify this in their other competencies – such as literacy and numeracy skills. Taking a similar line, Aparicio (2018) investigates if the academic performance of bilingual children is better than that of their monolingual counterparts. To perform her analysis, she uses the results on various academic tests taken by bilingual and monolingual children and controls for a wide range of characteristics that can differ between individuals and could affect the results of the tests. Thus, apart from demographic and socioeconomic aspects, she also takes into account home and school inputs. In so doing, she is able to compare bilingual and monolingual children with a similar individual, family, and schooling characteristics. Finally, she concludes that bilingual children outperform their monolingual counterparts.

It is important to take into account that this bilingual advantage might easily be thrown out of balance, given that other factors – above all education, which is typically conditioned by socio-economic status – play a major role in determining skills. Often, being bilingual is indicative of being an immigrant or a descendent of immigrants. A child whose parents are from abroad will probably be raised in the official language of the host country and their family's native language. Unfortunately, being an immigrant, or descendent of immigrants, is still frequently correlated with having a lower socio-economic status, and it can negatively affect the development of these skills. This case is most noticeable in countries like the United States, where a Welfare State does not exist to provide equality of opportunities in education or subsidies for low-income families. In fact, there are multiple studies for the United States in which researchers find a negative correlation between bilingualism and school results. Locay et al., (2013) find significantly lower results on literacy and numeracy test scores of Hispanic children who grew up in the United States but speaking Spanish at home. This negative effect might be explained by the so-called "school readiness gap". Castro et al., (2011) identify this gap in a sample of low-income bilingual children and monolingual middle-class children, and explain that the education and language skills development they present when starting school differs significantly, and this seems to be a consequence of the socio-economic level of their families. This mismatch of knowledge in the starting point predetermines the differential level of acquisition of new competencies during the schooling years. As Haskins et al., (2004) explains, there are no large samples of bilingual individuals of medium or high socio-economic levels in the United States to study, because bilingual homes are "disproportionality low socio-economic status homes". This case is not unique to the United States, in some other countries, children of immigrants, typically bilinguals, also have low levels of academic achievement (Scheele et al., 2010). In contrast, in some other regions, where being bilingual is not related to a low socio-economic status, like Montreal, Quebec, and Wales, bilingualism appears to have a positive impact on school results (Peal & Lambert, 1962; Gathercole, 2010). The same is true for some Post-Soviet countries such as Latvia and Estonia, which also introduced bilingual education together with Russian. The language policies improve student's achievement scores. (Khavenson & Carnoy, 2016).

Although there appears to be a high degree of consensus regarding the positive impact of bilingualism on skills acquisition, there is considerable heterogeneity in the results obtained in studies conducted in different countries. Cobb-Clark et al., (2018) find that the language development of bilingual children is not

significantly different from that of their monolingual counterparts. However, they show the evidence of a positive effect of bilingualism on emotional development and point out that the main causes of these differences would appear to be socio-economic status and educational programs. In this paper, in addition to controlling for these characteristics, I also take into account the heterogeneous effects by country and the type of languages in an effort to identify the determinants of these disparities.

Given that several countries that are multilingual introduced (or removed) bilingualism at school, or changed the language of instruction, several researchers investigated the effect of these language-in-education policies, and they typically find a positive impact of bilingual schooling. However, the benefits of bilingual education programs are still controversial. Anghel et al., (2016) evaluate a program that introduces bilingual English-Spanish education in some schools of the Madrid region and find a negative effect on learning the subject instructed in English, above all in children with less-educated parents. It is important to notice that most of the teachers did not have a native fluency in English and many of the children did not use English outside school (especially the children from families with lower educational levels). This precludes real linguistic immersion and seems to be the most likely reason behind the negative outcomes. On the other hand, Capellari and Di Paolo (2018) find positive wage effects of bilingual schooling. They exploit a reform that changed Catalan schools from monolingual to bilingual education and they find a gain in earnings. As they argue, the rise in the return to education once bilingualism is introduced may reflect the human capital impact of bilingual education, which stimulates the development of cognitive skills. Angrist and Lavy (1997) draw the same conclusion but studying the opposite case. Education reform in Morocco saw bilingual instruction in Arabic and French give way to monolingual Arabic schooling, but the authors identify a wage penalty as a consequence of the loss of French skills, and probably, of cognitive skills too. On this basis, there is enough evidence that bilingual instruction is associated with high levels of academic achievement (Genesee & Lindholm-Leary, 2012) but the success of bilingual education is typically found in countries where bilingual children all speak the same two languages (Hoff, 2013).

Another factor to consider in order to identify the possible causes of the heterogeneity in results is the linguistic distance between the two languages used by bilinguals. Linguistic distance is defined as the dissimilarity between languages – the more two languages differ from each other, the higher the linguistic distance between them. Languages can differ in several aspects such as grammar or pronunciation. (Isphording & Otten, 2013). In the recent years, several papers exploring the effects of linguistic distance have been published and, although there is no consensus on how to calculate this distance, what is clear is that it appears to be an important factor for explaining economic outcomes, such as migration, tourism flows, international trade and the effects of bilingualism (Chiswick & Miller, 1999; Hutchinson, 2002; Lohman, 2011; Isphording & Otten, 2014). In the next section, I explain the approach I adopt to measuring linguistic distance in greater detail.

In line with the theory, my study, which aims to explain the differential impact of bilingualism on the results of The Survey of Adults Skills (literacy and numeracy tests), should reflect this positive correlation and, possibly, the heterogeneity between countries depending on the language of instruction employed in schools and the linguistic distances between the languages spoken.

3. Data and descriptive statistics

3.1. Data

I use data from The Survey of Adult Skills from the *OECD Programme for the International Assessment of Adult Competencies (PIAAC),* waves 2013, 2016, and 2019. The survey is designed with the objective to fully analyze the distribution of skills across key subgroups of the adult population (aged from 16 to 65 years), hence, it does not only provide information on the results of the tests, in addition, it includes the responses to a complete questionnaire of individual characteristics.

The Survey of Adult's Skills aims to capture adult's proficiency in literacy, numeracy and problem-solving in technology-rich environments, which are considered to be the "key information-processing skills", in that they are: (i) Necessary for fully integrating and participating in the labor market, education, and social life, (ii) highly transferable in social contexts and work situations and, (iii) "learnable" and, therefore, subject to policy changes. However, I restrict the analysis to the scores in the literacy and numeracy tests, since the problem-solving test was not applied in all countries. The scores are standardized with a mean score of 500 and a standard deviation of 100, as well as measured using Plausible Values. The background questionnaire contains information on the basic demographic variables, such as gender and age, but it also includes data regarding education, labor experience, immigration status, social background, family's information, and what is more interesting for my research; it contains information about the mother-tongue and language background of the individuals.

One of the major concerns in research on bilingualism is finding an adequate definition of it. Indeed, this definition has varied widely over time. A bilingual individual was early defined as a person with "native-like" control of two languages (Bloomfields, 1935) or a person who regularly alternates two languages (Weinreich's, 1953). The main criticisms of these approaches are on how to measure the proficiency to be considered "native-like", as well as the role of the acquisition's age. Nowadays, there is sufficient consensus on the idea that bilingualism shows significant cognitive advantages if it is acquired early in life (Bialystock et al., 2009; Adesope et al., 2010). For this study, I use the definition proposed by Kohnert (2010); "bilinguals are the individuals who receive regular output in two languages during the most dynamic period of communication development – somewhere between birth and adolescence".

Thanks to the richness of data I can identify the individuals who self-report them as bilinguals when answering in the background questionnaire that they "had learned two languages at home in childhood and

still understand them". In order to avoid misspecification, I exclude the immigrants of the sample; they face a particular challenge when doing these tests because in most cases they are non-native speakers of the host country language, which makes difficult to separately identify the effect of the knowledge of more than one language from other factors related to their cultural origins, even among migrants who are fully proficient in the host country language. Moreover, having migrants in the sample increases the degree of unobserved heterogeneity, due to omitted inputs of the skills' production function such as the education quality in the origin country. For more robustness, in alternative specifications, I keep both immigrants who arrived at the country with 4 or fewer years old and those with 14 or fewer years old. This is because they are supposed to be schooled in the country where they reside – and hence, where they are tested – and this fact makes them more comparable to native-born individuals.

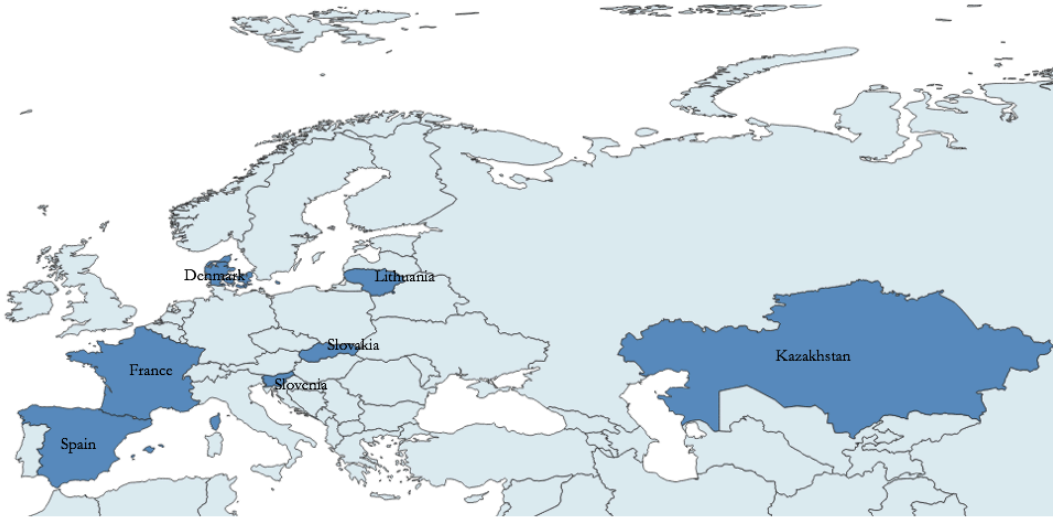

Figure 1:

Countries used in the analysis.

Although the entire survey was conducted in 25 countries, I only keep data for the European and post-soviet countries with more than 5% of a bilingual population in the sample, in order to ensure a sufficiently representative sample. I exclude out from the sample Peru and Mexico because, despite having more than 5% of bilinguals, it is mainly due to the presence of indigenous languages and it can create unobserved heterogeneity among countries. As a result, I keep data from Denmark, France, Kazakhstan, Lithuania, Slovakia, Slovenia, and Spain to perform the analysis. Two main reasons explain why these countries have an important proportion of a bilingual population. On the one hand, we can identify bilingualism due to the co-existence of neighboring – or regional – languages in the country, as a result of a shared history. This is the case of Spain, where most bilinguals speak a regional language such as Catalan or Galician. Slovenia, where there are frontier regions that have as a co-official the neighboring language, i.e. Italian and Hungarian. As well as, in Slovenia, there is a commonly spoken language due to a shared history; the Serbo-Croatan language, which was the official language of the former Yugoslavia, prior to the independence of this country. The same is true for Kazakhstan, with a great number of bilinguals of Kazakh and Russian,

because Russian was the official language of the Soviet-Union, to which this country belonged to. In Lithuania, there is a similar situation with Russian, but the language has not persisted that much, and nowadays it is not considered a co-official language while in Kazakhstan it is. On the other hand, bilingualism could be present in a country due to former massive waves of immigration and generational transmission of the mother-tongue, combined with the use of the official language of the host country. This is especially the case of Denmark and France, in addition to the Slovak Republic to a lesser extent. Denmark has received a significant influx of migrants, especially in the times of economic expansion, prior to the 1973 crisis. France, likewise, has experienced three large historical migratory flows – during the Industrial revolution, the inter-war years, and after World War II –, which have been mainly carried out by people from the French colonies. However, from the seventies, the migration policies become stricter in these countries. Figure 1 illustrates the countries used in the analysis and their geographic location.

As state before, one variable interesting to take into account when analyzing the effects of bilingualism is the linguistic distance between both bilingual's native languages (L1 and L2). Languages can differ in a multitude of dimensions, such as vocabulary, grammar, pronunciation, scripture, and phonetic inventories, due to this fact, it is difficult to find one unique approach for calculating the linguistic distance. In this paper, I use the same measure as Isphording and Otten (2013) in their paper "The Cost of Babylon – Linguistic Distance in Applied Economics", which is derived from the automatic comparison of the pronunciation of words from different languages having the same meaning. This measure is based on the Automatic Similarity Judgment Program (ASJP) developed by the *German Max Planck Institute for Evolutionary Anthropology* and is computed as a function of the phonetic similarity of words (what is called Levenshtein distance). The authors explain that this approach is the most appropriated in order to include linguistic distance into econometric applications. The basic idea is to compare the pronunciation of pairs of words with the same meaning. With this purpose is used the "40-item Swadesh list", which includes 40 words that are considered to be the most common in almost all the world's languages. It includes parts of the human body and some other basic words of the environment. This list was designed by Swadesh in 1952 and it is relevant because it is considered to be universally and culture independently. Although this method only considers the phonetics, a lower Levenshtein distance means a higher probability of sharing other language characteristics such as grammar (Serva, 2011). Therefore, the higher the Levenshtein distance, the more dissimilarity between the languages. Table A1 in the Appendix summarizes some examples.

Apart from that variable, the model will include controls for the basic demographic variables; country, gender, age, and years of schooling. As well as, it will control for parental migration and parental education, for the fact of having both languages different from the language of the test, and for the official – and co-officials, if applicable – languages of the country.

3. 2. Descriptive statistics

The target population of the survey is the non-institutionalized adult population, aged from 16 to 65 years, living in the country in the moment of the realization of the tests, irrespective of nationality, citizenship, and language status. In this case, excluding immigrants. I have a total sample of 35,726 observations[5], from 7 countries. Approximately, 11% of all the individuals in the sample are bilinguals. Table A2 in the Appendix provides detailed data by countries.

Figure 2 shows the Kernel density plot of the scores for literacy and numeracy tests performed by bilinguals and monolinguals. The results are sufficiently normal distributed for both monolingual and bilingual individuals, nonetheless, bilinguals show a higher peak in the center of the distribution. Differences between the distribution of the scores of bilinguals and monolinguals are more pronounced in the numeracy (left) than in the literacy (right) tests. One can observe that the monolingual distribution is shifted to the right, which means that without differentiating by country, monolinguals seems to outperform bilinguals a bit, and this difference is more stressed in the case of numeracy.

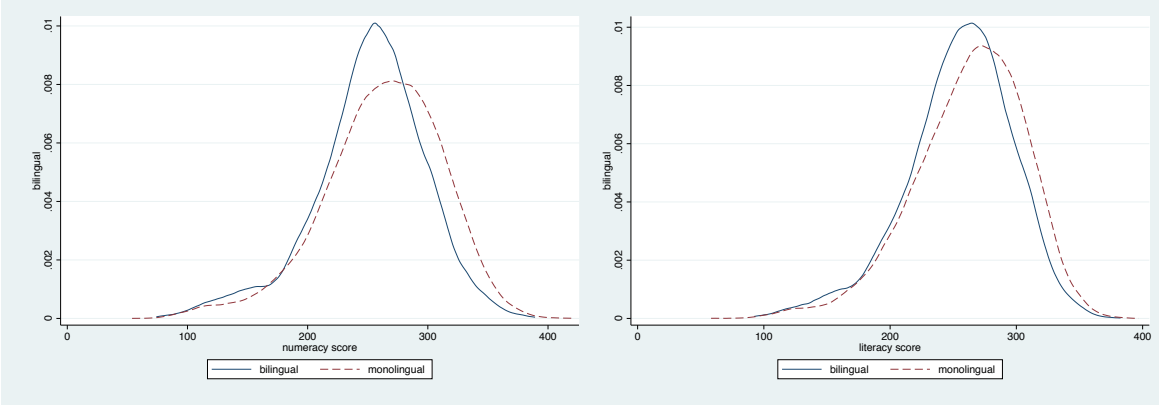

Figure 2:

Kernel density plot of literacy and numeracy scores of bilinguals and monolinguals.

Pooling all countries together might mask the differential effect of bilingualism. On this account, I split the graphics by country (Figures A1 and A2 in the Appendix) to identify the bilingual impact on each. Once more, one can notice greater differences between bilinguals and monolinguals in the results of the numeracy test that in the results of the literacy test. This, in line with the literature, means that being raised as bilingual has not only an impact on literacy skills, the cognitive impact could be even greater in other competences. Observing the differences in all the graphics, one can notice interesting patterns. In Spain and Slovenia, bilingual distribution is shifted to the right in both tests, which means that bilinguals in these two countries do better in literacy and numeracy test than their monolingual counterparts. In Kazakhstan, while it is almost

[5] 36,234 & 36,948 individuals when including immigrants who arrived before the age of 4 and 14, respectively.

impossible to notice any difference between bilinguals and monolinguals in the density plot of literacy scores, bilinguals seem to outperform monolinguals in the numeracy test. For the rest of the countries, monolingual distributions are slightly located to the right. At first sight, the deviation seems to be only important in the Slovak Republic, where monolinguals significantly outperform bilinguals in both tests.

Table A3 in the appendix, reports the descriptive statistics of the main variables used in the empirical analysis, for all countries and with a distinction between bilinguals and monolinguals. In both cases, the sample includes a few more females than males (54% of the total sample, 53% in the case of monolinguals and 56% in the case of bilinguals), and, on average, bilinguals are two years younger. Bilinguals are a bit more likely to have at least one foreign parent, with a higher standard deviation than for the case of monolinguals. However, the mean reflects that in most cases, both parents are native-born for both bilinguals and monolinguals, surely that is because I restrict the sample to a non-immigrant population. For the entire sample, despite the distribution of parental education is roughly equally divided between the three levels, there is slightly more presence of the lower level of education. Curiously, there are more bilingual's parents with the highest or with the lowest level of education than monolingual's parents and more monolingual's parents with the medium level. Besides, there are more monolinguals doing the test in a language different from their native language than bilinguals do, nevertheless, in both cases, they are a minority. The mean for the results on the literacy test is higher for monolinguals, and with a higher standard deviation. In the case of the numeracy test, monolinguals also report a higher mean and a higher standard deviation, in this instance, more stressed. While monolinguals earn a similar score in both tests, bilinguals, in general, do better in literacy. Lastly, the bilinguals' column also reports the descriptive statistics for the linguistic distance variable. While 0 means absolutely no difference between languages, the higher possible value is 105. The average language distance calculated for the entire bilingual sample is 86,46 with a standard deviation of 18.44.

From Table A4 to Table A10, the same results are reported for each country. Based on these characteristics, we can classify countries into two groups. In Denmark, France, Lithuania, and The Slovak Republic, the average scores of literacy and numeracy tests are greater for the monolingual group. While in Kazakhstan, Slovenia, and Spain the bilingual group is the one who achieved better results in both tests. The countries of each group do not only share within them the main reason for having an important percentage of bilingual population, as stated above. Moreover, the descriptive statistics also illustrate some other similar patterns. In the first group, bilinguals tend to report fewer years of schooling than monolinguals, as well as their parents are less likely to be native-born and also shown a lower educational level. In the second group, bilingual individuals and their parents show a greater educational level than their monolingual counterparts - just the opposite. If looking at the mean Levenshtein distance between L1 and L2, one can notice that the smaller is the distance, the greater scores bilinguals reach.

4. Empirical Strategy

In order to analyze PIAAC data, it is constructed a specific statistical model through which the results are allocated to the respondents and, therefore, the scores not only reflect the result obtained by each respondent in the test. Instead, there are used a set of maximum likelihood estimators, obtained from the answers of the individuals in the background questionnaire to come by a distribution that reflects the variability of the possible scores. From these estimations are extracted ten plausible values, which indicate the range of possible values obtained by the population. This distribution is constructed with the scores standardized with a mean of 500 and a standard deviation of 100. This technique aims to avoid underestimation or overestimation of the results. In this study, I regress the estimations using weighted replicate samples and plausible values - a typical procedure when studying surveys with a complex design in the estimation of sampling variances, such as PIAAC or PISA. Subsequently, I estimate the same equations by Ordinary Least Squares using as a dependent variable the mean of all the plausible values and I confirm that coefficients and Standard Errors do not differ significantly from the case when using the replicated weights. This last procedure allows me to measure the marginal effects of being bilingual on the test scores by country after each regression.

4.1 Baseline model

The first linear specification is the following:

$$T_i = \alpha + \beta' X_i + \sum_j \gamma_j I(C_i = j) + \delta BIL_i + \varepsilon_i \quad (1)$$

where *T* is the mean (weighted and unweighted) of plausible values score on the test, for numeracy or literacy, for individual *i*. *BIL* is a dummy variable that takes value 1 if the individual is bilingual and 0 otherwise. I(C=j) is a dummy variable for the country of the individual *i=j* (for the base level I use Spain as a country of reference). *X* is a covariate vector, which will gradually include a set of individual controls. And, ε is the error term.

The baseline specification only includes the basic demographic control variables; age, gender, and years of schooling. I control for the possibility of a non-linear relationship between the age and the results in the test by introducing the squared and cube of the age variable. I will progressively augment the set of controls to account for possible confounders, as the parental background.

Since this equation estimates a homogeneous effect of bilingualism for all countries, the coefficient of interest in this specification is δ, which reflects the impact of being bilingual into the score gained on the test, but I also consider other outcomes such as the coefficient of the control variables.

4.2 Accounting for a heterogeneous effect of bilingualism by country

As already argued, there is sufficient evidence that the impact that bilingualism may have on individuals' skills does not necessarily have to be the same in all countries. A useful way to analyze this differential effect is adding interactions between country and bilingualism in the OLS specification:

$$T_i = \alpha + \beta' X_i + \sum_j \gamma_j I(C_i = j) + \delta BIL_i + \sum_j \lambda_j BIL_i \times I(C_i = j) + \varepsilon_i \qquad (2)$$

After regressing the equation, I perform an interaction test to verify that the interaction between being bilingual and being from each country is statistically significant to explain the test scores. Subsequently, I measure the marginal effects of being bilingual on the test scores by country, which will be the coefficients of interest in this case. Marginal effects are calculated as derivatives of responses and it accounts for the general effect of being bilingual and the interacted effect of each country:

$$\frac{dy}{dx} = \frac{d(T_i)}{d(BIL_i)} = \delta + \lambda_j$$

4.3 Controlling for parental effects

The role of parents may confound the effects of bilingualism on skills acquisition. If one or both parents are from abroad, as well as their educational level could influence whether the individual develops the bilingual advantage. These parents' characteristics are controlled in the model by introducing a categorical variable explaining if one or both parents are from abroad or native-born as well as another variable for their educational level.

With this purpose, I estimate equation 3, which is like the equation 2 with adding these parental controls to the X_i set of covariates.

4.4 Including different languages effects

In order to study the impact of bilingualism, it is necessary to account for language diversity. Although is not very common, it exists the possibility of that some bilingual individual might have the two languages different from the language of the test. We use to find this issue when neither of the two languages is the official in the country or in the region where the test is applied, and this could confound the true effect of bilingualism in skills because this bilingual individual is facing an extra difficulty. If the individual is doing the test in a different language than their native, even if the person knows the language, it probably will negatively affect their results. Consequently, in equation 4 I consider relevant to include a dummy variable controlling for this issue:

$$T_i = \alpha + \beta' X_i + \sum_j \gamma_j I(C_i = j) + \delta BIL_i + \sum_j \lambda_j BIL_i \times I(C_i = j) + \omega DIFFL_i + \sum_j \theta_j DIFFL_i \times I(C_i = j) + \varepsilon_i \quad (4)$$

DIFFL is a dummy variable that takes the value 1 if the individual is doing the test in a non-native language and 0 otherwise. It is introduced in the equation independently as well as interacted by the country.
Furthermore, I include the Levenshtein distance between both languages of bilingual individuals. With this variable, I aim at measuring the effect of bilingualism *ceteris paribus* with respect to the distance between L1 and L2, but also to analyze the independent effect of distance among bilinguals. This procedure enables me to assess whether the dissimilarity between languages is relevant to explain the impact of being bilingual in the development of cognitive skills. I started controlling for the distance between both languages of bilinguals with a differential effect of distance by country, but the interaction test reveals that a differential effect in this case is not significant to explain the bilingual effect on the results of the test. Therefore, in the most complete specification I consider a common effect of linguistic distance across countries:

$$T_i = \alpha + \beta' X_i + \sum_j \gamma_j I(C_i = j) + \delta BIL_i + \sum_j \lambda_j BIL_i \times I(C_i = j) + \rho DL1L2_i + \omega DIFFL_i + \sum_j \theta_j DIFFL_i \times I(C_i = j) + \varepsilon_i \quad (5)$$

The final equation is equation 5, and it includes the variable *DL1L2,* which is the linguistic distance between L1 and L2 of the bilingual individuals. This equation aims to explain the impact of being bilingual on the average score in the literacy and numeracy test, controlling for country fixed effects, the basic individuals and parents' demographic variables, and for the distance between both languages of bilinguals. The equation includes the interaction effects of being bilingual and of doing the test in a non-native language by country. In contrast to the previous estimations, now the coefficients of bilingual (by country) capture the effect keeping fixed distance.

4.5 Robustness checks

Within country analysis

The following equation is estimated separately for each country:

$$T_i = \alpha + \beta' X_i + \delta BIL_i + \rho DL1L2_i + \omega DIFFL_i + \varepsilon_i \quad (6)$$

The X_i covariate includes all the controls explained above (demographic and parental controls) and the dependent variable refers to the literacy or numeracy scores. The estimation is performed separately for each of the 7 countries in the sample with the aim to analyze the bilingual impact on skills independently in each country.

Other potential confounders

In alternative specifications, I include a categorical variable explaining whether one, both or any of the languages that each individual speaks is official or co-official in the country.

I augmented the set of control variables to include household information with the aim to verify if the results are robust to this. More precisely, I add information on the number of children, if any, and a dummy variable that takes value 1 if the individual is living with the partner.

Finally, I augment the sample including migrants who arrived at the host country before the age of 4, and before the age of 14, and I repeat all the estimations including childhood first-generation migrants.

5. Results

The results of the equations presented before estimated by OLS with the non-weighted mean of Plausible Values are reported in Tables 1 and 2. Column 1 of Table 1 reports the baseline estimation, which only includes country fixed effects and demographic controls. The coefficient of interest, which captures the average impact of bilingualism within countries and keeping fixed country effects, age, gender, and schooling years, on the results obtained in the literacy test, is negative and significant. It means that, at first sight, being bilingual harms the literacy scores of adult individuals. Column 1 of Table 2 shows the results of the same estimation for the case of the numeracy test, which are in line with the ones of the literacy test but even more pronounced. These estimations assume a homogeneous effect of bilingualism by countries. In contrast, when considering it heterogeneous (from Column 2 on) the coefficient becomes positive and significant, for both literacy and numeracy tests. Column 2 reports the baseline estimation with assuming a heterogeneous effect of bilingualism by country and, Column 3 also includes controls for parental migration and education.

In these estimations, after allowing for heterogeneous effects by country, the coefficient associated with the bilingualism variable becomes positive and significant, since it captures the effect of bilingualism in the reference country, Spain, where having more than one mother tongue appear to have beneficial effects on skills formation. However, the coefficients of the interactions between bilingualism and country dummies, capturing the differential effect with respect to the reference country, are negative in most of the cases. In order to understand the total impact of bilingualism in each country, it is more useful to look at the marginal effects, that are reported in Tables 3 and 4, for the case of literacy and numeracy tests, respectively.

Therefore, Column 1 of both Tables 3 and 4 shows the same coefficient for all the countries. It is from Column 2, when introducing heterogeneous effects, that the coefficients differ between countries. Columns 2 and 3 report a total significant and negative impact of bilingualism in Denmark, France, Lithuania, and, especially, in the Slovak Republic. In all the cases the impact of bilingualism on numeracy scores is higher than the one on literacy scores. Conversely, being bilingual have a positive impact on skills formation in Kazakhstan, Slovenia, and Spain. For the case of Kazakhstan, this positive effect is only significant in the numeracy test.

Table 1

Literacy score

	(1) litpv	(2) litpv	(3) litpv	(4) litpv	(5) litpv
BIL	-2.295***	5.242***	5.089***	5.088***	13.972***
	(0.620)	(1.488)	(1.504)	(1.508)	(3.179)
BIL x Denmark		-10.454***	-10.997***	-11.146***	-10.666***
		(2.924)	(2.917)	(2.925)	(2.914)
BIL x France		-10.227***	-9.237***	-9.797***	-7.381***
		(2.385)	(2.543)	(2.549)	(2.642)
BIL x Kazakhstan		-3.961**	-3.573*	-4.408**	-1188
		(1.992)	(1.999)	(2.067)	(2.285)
BIL x Lithuania		-9.580***	-9.667***	-9.787***	-7.657***
		(1.902)	(1.918)	(1.920)	(2.038)
BIL x Slovak Rep		-20.121***	-18.835***	-19.571***	-17.958***
		(2.221)	(2.201)	(2.207)	(2.234)
BIL x Slovenia		3.792	3.788	3.432	1.941
		(2.855)	(2.846)	(2.846)	(2.913)
DL1L2					-0.123***
					(0.039)
DIFF				-0.765	-0.727
				(3.423)	(3.429)
DIFF x Denmark				-11.020	-10.889
				(7.644)	(7.655)
DIFF x France				-11.693**	-11.675**
				(4.941)	(4.946)
DIFF x Kazakhstan				-1.769	-2.079
				(3.754)	(3.762)
DIFF x Lithuania				6.014	5.082
				(4.083)	(4.106)
DIFF x Slovak Rep				-12.928***	-12.932***
				(4.008)	(4.013)
DIFF x Slovenia				-2.791	-2.647
				(5.466)	(5.459)
Demographic controls	Yes	Yes	Yes	Yes	Yes
Parents migration	No	No	Yes	Yes	Yes
Parents education	No	No	Yes	Yes	Yes
cons	236.531***	235.608***	229.742***	230.606***	230.586***
	(4.516)	(4.504)	(4.635)	(4.642)	(4.639)
N	35662	35662	34086	34086	34077
adj. R-sq	0.294	0.296	0.305	0.307	0.307
Interaction F-test (BIL x country)	-	19.84***	18.59***	18.90***	16.43***
(DIFF x country)	-	-	-	7.96***	7.21***

Notes: OLS estimations are reported with robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels. All the regressions include country fixed effects and the set of demographic controls includes: gender, years of schooling, age, age^2, and age^3.

Table 2

Numeracy score

	(1) numpv	(2) numpv	(3) numpv	(4) numpv	(5) numpv
BIL	-3.678***	6.579***	6.356***	6.425***	13.380***
	(0.679)	(1.534)	(1.544)	(1.548)	(3.427)
BIL x Denmark		-14.545***	-14.611***	-14.921***	-14.542***
		(3.209)	(3.206)	(3.222)	(3.201)
BIL x France		-15.310***	-14.113***	-14.987***	-12.883***
		(2.647)	(2.807)	(2.813)	(2.920)
BIL x Kazakhstan		-4.473**	-4.046**	-4.913**	-2.389
		(2.013)	(2.010)	(2.072)	(2.325)
BIL x Lithuania		-11.237***	-11.241***	-11.392***	-9.776***
		(2.027)	(2.025)	(2.027)	(2.156)
BIL x Slovak Rep		-29.562***	-27.752***	-28.946***	-27.683***
		(2.487)	(2.435)	(2.445)	(2.473)
BIL x Slovenia		3.410	3.731	3.140	1.925
		(3.009)	(3.001)	(3.000)	(3.068)
DL1L2					-0.096**
					(0.042)
DIFF				2.067	2.097
				(3.310)	(3.315)
DIFF x Denmark				-18.322**	-18.218**
				(7.856)	(7.870)
DIFF x France				-14.572***	-14.577***
				(5.054)	(5.057)
DIFF x Kazakhstan				-4.495	-4.738
				(3.641)	(3.649)
DIFF x Lithuania				-0.055	-0.767
				(4.096)	(4.118)
DIFF x Slovak Rep				-23.028***	-23.031***
				(4.062)	(4.067)
DIFF x Slovenia				-10.077*	-9.968*
				(5.701)	(5.698)
cons	223.760***	222.511***	213.886***	214.976***	214.922***
	(5.017)	(4.989)	(5.094)	(5.095)	(5.094)
Demographic controls	Yes	Yes	Yes	Yes	Yes
Parents migrations	No	No	Yes	Yes	Yes
Parents education	No	No	Yes	Yes	Yes
N	35662	35662	34086	34086	34077
adj. R-sq	0.338	0.342	0.353	0.355	0.355
Interaction F-test (BIL x country)	-	32.48***	30.76***	31.73***	29.51***
(DIFF x country)	-	-	-	11.31***	10.81***

Notes: OLS estimations are reported with robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels. All the regressions include country fixed effects and the set of demographic controls includes: gender, years of schooling, age, age^2, and age^3.

Table 3

Marginal effects of bilingualism by country for literacy score

	(1)	(2)	(3)	(4)	(5)
Denmark	-2.295***	-5.211**	-5.908**	-6.059**	3.306
	(0.620)	(2.518)	(2.501)	(2.502)	(3.880)
France	-2.295***	-4.984***	-4.147**	-4.709**	6.591
	(0.620)	(1.865)	(2.052)	(2.058)	(4.075)
Kazakhstan	-2.295***	1.281	1.517	.680	12.783***
	(0.620)	(1.325)	(1.317)	(1.414)	(4.036)
Lithuania	-2.295***	-4.337***	-4.578***	-4.700***	6.316*
	(0.620)	(1.184)	(1.189)	(1.187)	(3.687)
Slovak Republic	-2.295***	-14.879***	-13.746***	-14.484***	-3.986
	(0.620)	(1.647)	(1.605)	(1.610)	(3.577)
Slovenia	-2.295***	9.034***	8.878***	8.519***	15.913***
	(0.620)	(2.438)	(2.418)	(2.416)	(3.214)
Spain	-2.295***	5.242***	5.089***	5.088***	13.972***
	(0.620)	(1.488)	(1.504)	(1.508)	(3.179)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.

Table 4

Marginal effects of bilingualism by country for numeracy score

	(1)	(2)	(3)	(4)	(5)
Denmark	-3.677***	-7.965***	-8.256***	-8.500***	-1.162
	(0.680)	(2.819)	(2.811)	(2.827)	(4.230)
France	-3.677***	-8.731***	-7.757***	-8.562***	.497
	(0.680)	(2.157)	(2.345)	(2.352)	(4.520)
Kazakhstan	-3.677***	2.107*	2.309*	1.512	10.991**
	(0.680)	(1.304)	(1.288)	(1.378)	(4.337)
Lithuania	-3.677***	-4.657***	-4.885***	-4.967***	3.604
	(0.680)	(1.324)	(1.308)	(1.307)	(4.012)
Slovak Republic	-3.677***	-22.982***	-21.396***	-22.521***	-14.302***
	(0.680)	(1.956)	(1.882)	(1.892)	(3.944)
Slovenia	-3.677***	9.989***	10.086***	9.565***	15.305***
	(0.680)	(2.590)	(2.575)	(2.572)	(3.442)
Spain	-3.677***	6.579***	6.356***	6.425***	13.380***
	(0.680)	(1.534)	(1.544)	(1.548)	(3.427)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.

Table 5

Marginal effects of DIFF by country for literacy and numeracy score

	Literacy		**Numeracy**	
	(4)	(5)	(4)	(5)
Denmark	-11.785*	-11.616*	-16.255**	-16.121**
	(6.841)	(6.851)	(7.134)	(7.147)
France	-12.458***	-12.402***	-12.505***	-12.481***
	(3.574)	(3.576)	(3.837)	(3.837)
Kazakhstan	-2.534*	-2.805*	-2.428*	-2.641*
	(1.536)	(1.540)	(1.513)	(1.517)
Lithuania	5.249**	4.355*	2.012	1.329
	(2.231)	(2.256)	(2.418)	(2.441)
Slovak Republic	-13.693***	-13.659***	-20.961***	-20.935***
	(2.084)	(2.086)	(2.356)	(2.357)
Slovenia	-3.556	-3-374	-8.010*	-7.782*
	(4.269)	(4.256)	(4.653)	(4.647)
Spain	-.765	.727	2.067	2.097
	(3.423)	(3.429)	(3.310)	(3.315)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.

Table 6

Marginal effects of distance at specified values for literacy and numeracy scores.

Margins at distance	Literacy score	Numeracy score
0	269.651***	264.252***
	(3.364)	(3.656)
10	268.424***	263.292***
	(2.985)	(3.243)
20	267.197***	262.331***
	(2.607)	(2.832)
30	265.970***	261.371***
	(2.232)	(2.423)
40	264.743***	260.410***
	(1.861)	(2.019)
50	263.516***	259.449***
	(1.497)	(1.623)
60	262.289***	258.489***
	(1.148)	(1.242)
70	261.062***	257.528***
	(.831)	(.895)
80	259.835***	256.569***
	(.600)	(.641)
90	258.608***	255.607***
	(.572)	(.611)
100	257.381***	254.646***
	(.769)	(.829)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels. This sample only includes bilingual population.

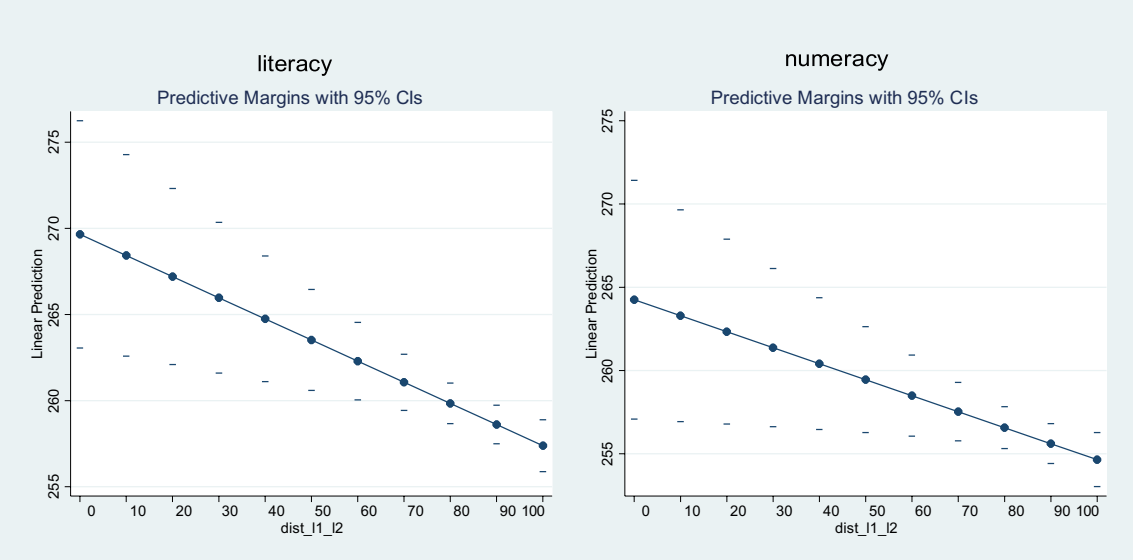

Figure 3:

Interaction effects: literacy and numeracy scores at linguistic distance specified values.

Returning to Tables 1 and 2, I control for the fact of completing the test in a non-native language in Column 4, and I also create an interaction for this variable by country. It is reasonable to think that the possibility of an individual doing the test in a non-native language can confound the results since this person will probably have lower fluency than natives, which may turn into a lower score. All the control variables present the expected sign. The marginal effects by country of the interactions of this regression are reported in Tables 3 and 4, as well as in Table 5 for the fact of doing the test in a different language than the native(s) one(s), for literacy and numeracy test. In this case, the marginal effects of bilingualism in each country are somewhat lower than in the previous estimations but, in essence, reflect the same conclusion; negative and significant impact in Denmark, France, Lithuania, and the Slovak Republic, positive and significant in Slovenia and Spain, positive but not significant in Kazakhstan. We also find a greater impact of bilingualism on the results of the numeracy test. Regarding the heterogeneous effect of doing the test in a non-native language, as expected, it is negative in all the countries, with the exception of Lithuania. In the case of Slovenia and Spain, despite showing a negative effect, it results not significant to explain the results of the test. One interesting finding here is that the fact of not being fluent in the language of the test creates a somewhat higher difficulty in the numeracy test than in the literacy one.

The last column of tables 1 and 2, shows the results of equation 5, the most complete. It controls for all the demographic variables, parental variables, the differential effects of bilingualism in each country, and the possibility of doing the test in a non-native language. Moreover, I also include as additional regressor the linguistic distance between both native languages of the bilinguals (which is equal to zero for monolingual individuals). Initially, I define the equation allowing for heterogeneous effects of linguistic distance by country, by including interactions. However, the corresponding coefficients of these interactions were not significant neither individually nor jointly. For this reason, I include the variable of the linguistic distance in the equation without interactions. The results show that linguistic distance has a negative, significant, and homogeneous impact on the results of the tests, in this case, the impact is stronger in the literacy test. The higher is the distance between both languages of the bilinguals, the worse. Furthermore, looking at the marginal effects of bilingualism (Tables 3 and 4), one can notice how the coefficients increase significantly. When introducing the linguistic distance in the equation, the coefficient of bilingualism is obtained after partialling out the effect of the distance between L1 and L2. In this case, being bilingual does not imply getting lower results in the tests in any country (except in the case of the numeracy test in the Slovak Republic) and it reaffirms the significant bilingual advantage that individuals show in Spain, Slovenia, and Kazakhstan. It suggests that a large linguistic distance harm somehow the positive impact that bilingualism can have on skills formation.

In Table 6, I report the predicted conditional tests for different levels of linguistic distance among bilinguals. The results show a decreasing marginal effect. Figure 3 illustrates this relationship. Both literacy and numeracy scores of bilinguals decrease when the linguistic distance increase.

Kazakhstan is the country of our sample with a higher mean of the linguistic distance between their bilingual population (see Tables A4 to A10 in the Appendix). It explains why this country reports in all the equations

a modest bilingual advantage, but when controlling for the linguistic distance this positive impact of bilingualism becomes highly significant.

Robustness checks:

Within country analysis:

Table A11 in the Appendix presents the results of equation 5 for the literacy test independently estimated for each country. The coefficients differ somewhat in magnitude from the previous regressions, but results are robust to the main conclusion – positive and significant effect of bilingualism in Slovenia and Spain, positive coefficient but without significance in Kazakhstan, mainly an insignificant impact for the rest of the countries, as well as a negative effect of linguistic distance for all the countries but Lithuania.

Controlling for the official language of the country:

Table A12 in the Appendix (A13 for margins) present the estimation using a categorical variable that takes value 0 if the individual is not bilingual or if it is bilingual of two non-official[6] languages in the country, 1 if the individual is bilingual of one official and one non-official language (i.e. English – Spanish bilingual in Spain) and 2 if is bilingual of two official languages (i.e. Catalan – Spanish bilingual in Spain).

Results point out that being bilingual of two official languages (we find more than 1 official language in Spain, Slovenia, and Kazakhstan) has an important significant effect on the scores of both tests. A significant part of the bilingual individuals who are native of two official languages had received a bilingual education and, if not, they work or live in a bilingual environment. This result is important because it suggests that there exists a bilingual advantage that develops its positive effects on individual skills depending on the level of linguistic immersion.

Including household information:

I run new specifications that include control variables for household characteristics, which are other possible confounders when analyzing the scores in tests of competences. The results are presented in Table A14 (marginal effects in A15). The estimated bilingual effect by country is very similar to those reported previously. All the results are robust to the inclusion of household variables.

Including migrants who arrived in the host country before the age of 4 and 14:

Extending the sample with including immigrants who arrived at the country sufficiently young to learn the official language as a native individual does not change the results significantly. There are no important differences in the coefficients, but some of them become more stressed. The unique considerable difference is in the coefficient associated with the control variable of parents' migration, as expected. The fact of having both parents native-born loses all the signification; there are not immigrants with both parents native-born.

[6] I consider co-official and regional languages as officials. I provide a list of the official languages of these countries in the Appendix.

In contrast, having one parent native-born and one foreign-born becomes highly positive and significant. Nevertheless, the coefficients of interest are robust to the inclusion of these immigrants. In any case, despite no considerable differences are found, I consider the most appropriate to do the analysis without these observations because they can still create a selection bias due to socioeconomic or family reasons.

6. Discussion

With the findings reported, I can identify two relevant factors determining the impact of bilingualism. What is suggested here is that we can find different contexts, depending on the combination of these two factors, that can affect the development of bilingual advantage differently. These factors are the following ones:

1) The linguistic immersion degree – which is determined by the number of different languages spoken in the country, as well as the number of speakers of each one.
2) The distance between languages.

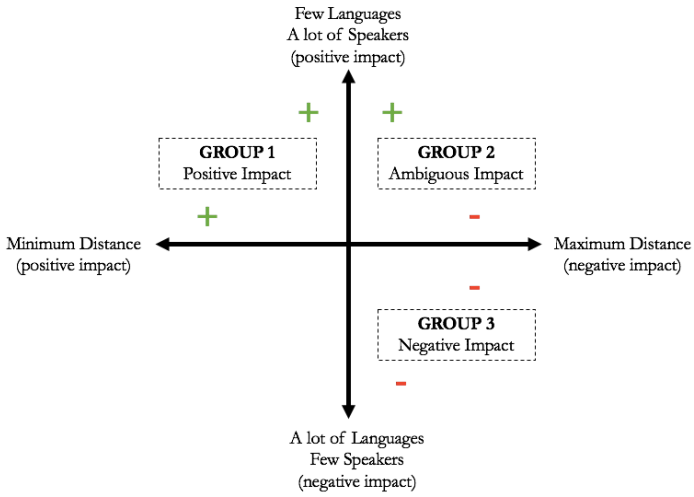

Figure 4:

Trade-off between (1) number of languages and speakers, and (2) distance between languages.

I find a trade-off between the positive impact of factor 1 and the negative impact of factor 2. One the one hand, if there are few languages with a lot of speakers, the individuals of the country are more likely to experience a real bilingual immersion. We use to find this situation in countries with more than one official language, where two (or more) languages are well-established and, typically, also integrated into the educational system. This fact appears to create a positive impact to foster cognitive advantages of bilingualism. On the other hand, the results point out that the more distance between both languages, the less bilingual advantage is reported.

The mean distance between L1 and L2 for the entire bilingual sample from the 7 countries is 86.46, the minimum value is 28.36, and the maximum 104.13. Therefore, I consider "large distance" when it is above this mean. The trade-off between these two factors is represented in Figure 4.

Given this intuition, three groups of countries, with different bilingualism outcomes, are identified:

GROUP 1: Little variety of languages with low linguistic distances.
In this group of countries, both factors are favoring the bilingual advantage and we can find a significantly positive impact of bilingualism on literacy and numeracy scores. This is the case of Slovenia and Spain. These two countries show a robust positive and significant impact of bilingualism in all the regressions. The bilingual advantage can be easily identified in the descriptive statistics, and the positive and significant coefficients associated with these countries in the estimations support this intuition.

Slovenia: The most spoken languages in Slovenia, apart from Slovenian, are: (i) Croatian; 50.44% of the bilingual population in Slovenia speaks Slovenian-Croatian – the distance between this pair of languages is the smaller one reported in the Levenshtein measure sample, only 28.36. And, (ii) Italian; with 21.49% of native speakers, the distance Slovenian-Italian is 88.76. Both languages represent 72% of the Slovenian bilingual population in my sample, the other languages found are spoken by a reduced number of people.

Croatian (or Serbo-Croatian) was one of the official languages of the former Yugoslavia, together with Slovenian, and Macedonian. Slovenia becomes independent from Yugoslavia in 1991, following the so-called armed conflict "Ten-Day War". Until this time the Croatian language was mandatory in all the schools in Slovenia and commonly used throughout the country. Despite nowadays it is not explicitly considered an official language, the use of it persists.

Italian is co-official together with Slovenian in the part of the country that is bordering with Italy. Specifically, the cities of Piran, Izola, Kopler, and Ankaran, which form a considered ethnically mixed area. In this region, the education system is bilingual[7]. This is specified in the constitution of the country of 1991 with the aim to respect the historical community of Italian people there. This also happens with Hungarian, but the representation is lower, only a 3,51% of the bilinguals in my sample for Slovenia. Figure 5 illustrates how the languages most spoken in Slovenia are from neighboring countries (Croatian and Italian), as well as the green circle refers to the bilingual Slovenian-Italian region.

[7] According to the *Eurydice Network*.

Figure 5:

Bilingual influence in Slovenia

Spain: The combination of languages commonly encountered in Spain are: (i) Spanish-Catalan; this is the most frequent combination of languages in Spain, it is spoken by the 50.64% of the Spanish bilingual population of the sample, and present a Levenshtein distance of 72.12 (below the average). (ii) Spanish-Galician; spoken by 26.60% of the bilingual sample in Spain, these two languages show a distance of 54.82, which is significantly low. And there is also some presence of (iii) Spanish-Basque, this pair of languages present a higher distance, 101.71 which is notably above the average, but this combination is not sufficiently representative with a 10.10% of speakers in the sample, in contrast to the global 77.24% that sum the other two pairs.

Galician is a co-official language, together with the official Spanish, in the Galician region, where mainly all the individuals are bilinguals. From 1983, in Galicia, the education is bilingual; some subjects are taught in Spanish and some others in Galician. Moreover, from 2001, the Galician language is also taught in the schools of a part of the Castile-Leon region[8].

Catalan is co-official in the regions of Catalonia and Balearic Islands, together with the Valencian variation in the Community of Valencia. From 1983, there is a non-optional bilingual education in Catalan and Spanish in Catalonia, as well as there is the option of bilingual education in the Balearic Islands, the Community of Valencia, and the part of Aragon that is bordering with Catalonia. In these three cases, even if the family chooses the non-bilingual program, the language is taught.

Basque is co-official in the Basque Country and part of Navarre, the education in this language is like in the case of the Balearic Islands and the Community of Valencia; the family can choose whether the program is bilingual or they only received one subject of the Basque language. Figure 6 illustrates the bilingual regions in Spain.

[8]All the information regarding the language presence in the education system of Spain is extracted from the *Ministerio de Educación y Formación Professional. Gobierno de España.*

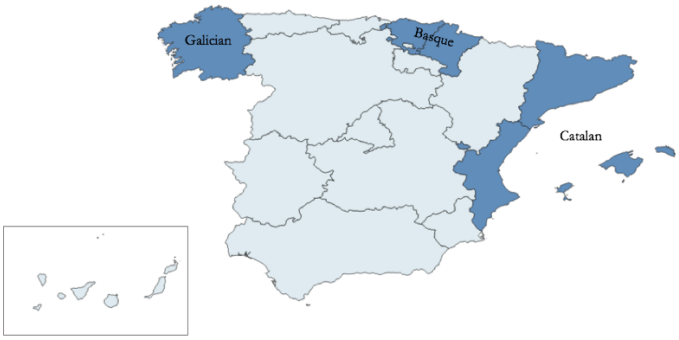

Figure 6:

Bilingual regions in Spain

Slovenia and Spain have several language characteristics in common. In both countries, there is a low variety of languages spoken for a significant percentage of people, and the pair of languages most spoken presents a below-average Levenshtein distance. The reason for what there is a significant proportion of a bilingual population in the country is the same. The most spoken languages are regional or neighboring languages; a consequence of shared history. These languages are (or were) officials in the country and taught in the schools. Furthermore, in these two countries, I find a positive correlation between bilingualism and years of schooling – bilinguals in Spain and Slovenia tend to be more educated than monolinguals. This fact dismisses the possible association of bilingualism with a lower educational and socio-economic level in these countries.

GROUP 2: Little variety of languages with large linguistic distances.

Little variety of languages with a significant number of speakers enhance the positive impact of bilingual advantage, but a high distance between these languages can undermine the positive effects. The results reported in this paper confirms the negative impact of linguistic distance on the scores in the tests. This is the case of Kazakhstan and, to a lesser extent, Lithuania. Kazakhstan presents a modest positive coefficient associated with bilingualism in all the regressions, but it is not until the last one, when controlling for linguistic distance, that it turns highly significant. In Lithuania, the coefficient only turns positive and modestly significant in the last regression, when keeping fixed the effect of the linguistic distance.

Kazakhstan: In Kazakhstan, one combination of languages prevails. The 86,67% of the bilingual population is native of Kazak and Russian. This country was the last one to declare its independence from the Soviet Union, in 1991. Probably, for this reason, it is the post-soviet republic where more people speak Russian. Moreover, the Russian language is official in the country, together with Kazak, and is taught in the schools. The rest of the languages in this country account for only a small part of the bilingual population. However, the bilingual advantage is undercut by the important distance between these two languages – a measured Levenshtein distance of 100.6, one of the largest. It explains why the coefficient becomes highly significant when controlling for the linguistic distance. Even though the impact is not as great as in the

countries of the group 1, bilingual individuals in Kazakhstan outperform monolinguals as well. Correlation between bilingualism and years of education is also positive in Kazakhstan. Figure 7 illustrates how all the Northern part of the country is surrounded by Russia. As in the previous case, the bilingualism in Kazakhstan is influenced by a neighboring country with whom they share history

Lithuania: This country shows the soft version of the Kazakhstan case. The results are in the same line, but less stressed. In the case of Lithuania, the coefficient only became positive and significant (at the 10% level) when controlling for the linguistic distance. In Lithuania, the most spoken language, together with Lithuanian, among bilingual individuals is also Russian – in this case, it represents 70.12% of the bilingual population. There is also some presence of English and Polish; 15.53% and 7.1%, respectively. The Levenshtein distance between Lithuanian and Russian is also large; 91.87. Lithuania, took part in the Soviet Union as well, however, in contrast to Kazakhstan, Lithuania was the first state to declare their independence, in 1990. In addition, previous to the Soviet Union occupation of Lithuania in 1940, in the context of the World War, the country was considered independent from the Russian Empire from 1918. Although there is an important shared history between Lithuania and Russia, it is not as strong as in the case of Kazakhstan and Russia. Due to this fact, the Russian language presence is somewhat lower in Lithuania, in fact, it is not considered an official language and in the educational system, it can be taught, but is not mandatory. In contrast to the previous countries, bilingualism and years of schooling present a slight negative correlation in Lithuania.

Figure 7:

Geographical location of Kazakhstan

Figure 8:

Geographical location of Lithuania

Figure 8 presents the location of Lithuania. Geographically, Lithuania and Russia are not as close as Kazakhstan and Russia. As state above there is also a somewhat presence of Polish language in the country, which is the official language of a bordering country.

GROUP 3: Wide variety of languages with few speakers.

These countries are the ones that have a multilingual environment owing to former immigration waves that lead to a wide variety of spoken languages there. Before controlling for distances, these countries use to show a negative impact of bilingualism in the literacy and numeracy skills. After controlling for the linguistic distance, it becomes insignificant in Denmark and France, and still negative in Slovakia. In most cases, linguistic distances are large in these countries since the languages are not necessarily from neighboring regions.

Denmark: There is an important proportion of a bilingual population in Denmark, but there is not a predominant language among them. This is because Denmark has received a significant influx of migrants in the times of economic expansion, prior to 1973. The countries of origin of these immigrants were especially Turkey, Germany, Poland, and Sweden. The intergenerational transmission of the mother tongue of all these migrants has created a persistence of these languages among the native population which is a descendant of immigrants. The most spoken languages of bilinguals, together with Danish, are Deutsch, Swedish, Turkish, Arabic, English, Norwegian, and Polish, which are more or less equally distributed in terms of the number of native speakers. There is also some presence of other languages such as Spanish or Vietnamese. The calculated mean of all the Linguistic distances among the bilingual population in Denmark is about 82, below the mean but importantly large.

France: The case of France is close to the Danish one. France has received three large migratory waves. The first one was during the Industrial Revolution because the country required a large amount of labor force, the vast majority of them were from Italy, Germany, Belgium, and Spain. The second wave was during the inter-war years and the last one was after World War II, these two lasts migratory flows have been mainly carried out by people from the French colonies. Among all the languages spoken in France, we can find an important presence of Arabic, Italian, Portuguese, Spanish, Turkish, and Polish, as well as some other languages such as English or Deutsch. The calculated mean in this case is about 90.

Slovak Republic: This country has not experienced important historical migration waves as the other two countries, but it also presents a wide variety of languages. A few examples are; Hungarian, Romanian, Polish, Deutsch, Ukrainian, Russian, and English. All together present a mean Levenshtein distance of 85.

In these three countries, there is a negative correlation between bilingualism and years of schooling, which indicates that, despite there are no migrants in the sample, to speak a minority language, in most cases is yet correlated with a lower cultural and socio-economic level.

In sum, I find a bilingual advantage in the countries where being bilingual is not correlated with a lower academic level and there exists bilingual education in the most spoken languages in the country. Real

bilingual immersion, typically fostered by bilingual education, shows a significant positive impact on the literacy and numeracy skills of individuals. Moreover, this positive effect is enhanced if the linguistic distance is not significantly large.

The results give light to the benefits of bilingual culture and education, which is important from a policy point of view. Higher levels of skills imply higher productivity, greater wages, and therefore, a reduction of social and economic inequalities. Other authors already suggest the positive effect of bilingualism in educational outcomes (Aparicio, 2018) and labor market outcomes (Angrist & Lavy, 1997; Capellari & Di Paolo, 2018). This study gives intuition on the possible channel of this positive correlation.

This insight has been already argued using the cases of French-English in Canada and Welsh-English in Wales (Pearl & Lambert, 1962; Gathercole, 2010) – bilingualism has a positive impact on cognitive skills and academic performance if both languages are well-established in the society. Thus, individuals are exposed to two languages in all the possible contexts. Otherwise, in the case of a bilingual individual that lives in a country where only one of their languages is integrated into the society (i.e. it could be the case of a descendent of immigrants living in a country of the group 3) the most likely would be to use each language in a different area – for example, the heritage language in the family sphere and the official language of the country in the social or professional field. This segregation of the two languages precludes an effective bilingual immersion.

The findings suggest the importance of fostering bilingual societies, which is not achieved only by establishing bilingual education, also with applying other linguistic immersion policies, such as increasing the offer of cultural activities in both languages. In the case of societies where bilingualism is not well-established, education policies favoring bilingualism - thus without segregating students by their mother tongue – may have positive effects if there exists a relatively significant level of integration between the different linguistic groups and if the linguistic distance between these languages is not sufficiently large.

7. Spanish case

In order to test the validity of this intuition, I analyze the particular case of Spain. To use data of this country is interesting because in there we can find different linguistic realities. Despite the general effect of bilingualism in Spain has proved to be positive, we might find variability of results if we analyze separately the effects of the different pairs of languages that coexist in the country. In this sense, we can find the following languages combined with the Spanish: Galician and Catalan, which are two well-established languages that do not differ sharply from Spanish, Basque, which shows a certain degree of integration, but the Levenshtein distance between this language and Spanish is significantly large. Moreover, there is also some presence of bilingual individuals in this country who speak minority languages (i.e. English, French, Deutsch, Italian, Arabic, …), probably descendent of immigrants. In sum, within the Spanish sample, we

can identify bilinguals of integrated and non-integrated languages that can report either low or high distance between languages.

With the aim to analyze if there is a differential impact of bilingualism depending on the pair of languages inside the country, I estimate different variations of the following linear specification by Ordinary Least Squares:

$$L_i = \alpha + \beta' X_i + \delta LOCAL_i + \rho DL1L2_i + \omega DIFFL_i + \varepsilon_i$$

Where L_i is the score on literacy test. X_i is a covariate vector, which will gradually include the demographical and parental controls. *DIFFL* is the dummy variable that takes the value 1 if the individual is doing the test in a non-native language. *DLI1L2* is the linguistic distance between L1 and L2 of the bilingual individuals. *LOCAL* is a new dummy variable that takes value 0 if the individual only speaks Spanish, 1 if the individual is a Catalan-Spanish bilingual, 2 if they are a Galician-Spanish bilingual, 3 for a Basque-Spanish bilingual and, 4 if they are bilingual of Spanish and another language. ε is the error term. I only use data for native individuals residing in Spain and, for simplicity, I eliminate other language combinations and non-Spanish monolinguals. Therefore the coefficient of interest is δ, which reflects the impact of bilingualism, distinguishing by the different language combinations and using monolinguals as the reference category.

Table 7

Literacy score in Spain

	(1)	(2)	(3)	(4)
LOCAL=1	5.594***	5.130**	5.214**	-0.643
(Catalan-Spanish)	(2.078)	(2.074)	(2.073)	(5.883)
LOCAL=2	6.176**	7.061**	7.087**	2.631
(Galician-Spanish)	(2.819)	(2.850)	(2.848)	(5.066)
LOCAL=3	-4.682	-7.297*	-7.135	-15.385*
(Basque-Spanish)	(4.534)	(4.540)	(4.538)	(8.986)
LOCAL=4	2.639	3.527	3.419	-0.833
(Other-Spanish)	(3.313)	(3.821)	(3.381)	(5.235)
DIFF			14.270**	14.128**
			(5.688)	(5.689)
DL1L2				0.081
				(0.076)
Demographic controls	Yes	Yes	Yes	Yes
Parental controls	No	Yes	Yes	Yes
Cons	221.733***	204.295***	204.596***	204.106***
	(12.131)	(14.501)	(14.499)	(14.506)
N	4903	4756	4756	4756
Adj. R-sq	0.391	0.396	0.396	0.396

Notes: OLS estimations are reported with robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels. The set of demographic controls includes: gender, years of schooling, age, age^2 , and age^3. The set of parental controls includes: parental migration and parental education.

The results of the different regressions estimated are reported in table 7. The variations in the equation are similar to the previous ones: in the first column, the estimation only includes the demographic controls. In column 2, I extend the set of control variables including also parental education and parental migration. In column 3, I include the DIFF dummy and, column 4 reports the results of the described equation, which include all the variables[9].

From Columns 1 to 3 coefficients do not vary significantly. The results are robust to include the parental controls and the possibility of doing the test in a non-native language. In fact, the test is provided in Spanish throughout the whole country. Moreover, in the regions of Catalonia, the Balearic Islands and Valencia respondents can request for the Catalan version of the test, in the region of Galicia for the Galician version and in the Basque Country for the Basque version. Therefore, as I excluded the bilingual combinations that do not contain Spanish, it is difficult to find someone doing the test in a non-native language.
Bilinguals of Catalan-Spanish and Galician-Spanish show a significant advantage on the literacy scores with respect to their monolingual counterparts, the combination of Basque and Spanish seems to have a slightly negative impact, but almost insignificant and, being bilingual of Spanish and another non-regional language has not a significant effect on the literacy scores. The last column, when controlling for the linguistic distance between L1 and L2 only reports a slightly significant coefficient for the Basque-Spanish combination, which remains negative.

These findings are in line with the results presented above and report the same conclusion. Being bilingual of Catalan or Galician, two well-established languages in Spain for which bilingual education is offered, together with Spanish, seems to have a positive effect that helps individuals to develop the so-called bilingual advantage. Being bilingual of Basque and Spanish may even damage the results on the literacy test. Despite the Basque language is taught in the Basque Country region, the complexity of the language and the large Levenshtein distance that shows with respect to Spanish, difficult the linguistic immersion and can imply an additional challenge for the bilingual individuals of these two languages. Finally, being bilingual of the official language of the country together with a minority language not integrated into the society does not show any relevant impact on skill acquisition.

8. Conclusions

Does bilingual advantage exist? Although there is increasing evidence in the literature, the debate is still open. This paper investigates the effects of bilingualism in literacy and numeracy skills, allowing for heterogeneous effects by country – which are explained by the diverse historical reasons for which a country

[9] The same equations are regressed with numeracy test scores; the results are similar but in general the coefficients are less significant.

has a significant presence of bilingual population – and linguistic distance. I use data from the PIAAC, which contains the scores on literacy and numeracy tests, as well as data on an individual's characteristics, including information of their linguistic background. Thanks to having specific information about multiple languages, I can combine this dataset with the Levenshtein distance – a measure for the linguistic distance between different languages. I show that bilingualism has a positive effect on literacy and numeracy skills in societies where bilingualism is a stable phenomenon and there is linguistic immersion in both languages. Moreover, I find that a large linguistic distance between languages harms this bilingual advantage. Therefore, bilingualism has a positive impact on the results of the tests of skills in Spain and Slovenia, as well as, after controlling for the effect of the linguistic distance, in Kazakhstan and Lithuania too. In Denmark, France, and the Slovak Republic, where the different linguistic groups are not well integrated because there exists a wide range of different languages among bilinguals, this bilingual advantage has not developed. In addition, I also analyze the Spanish case and I find the same conclusion within the country – Catalan-Spanish and Galician-Spanish bilinguals show an advantage in the tests of competencies compared to monolinguals and other bilinguals.

Even though I control for diverse possible confounders such as demographic characteristics, parental and household information, the possibility of doing the test in a non-native language, as well as the official languages of the country; there may continue to exist some Omitted Variable bias, thus creating a certain degree of endogeneity. In addition, despite the fact that mother-tongues are considered to be exogenous variables, they may be capturing other unobserved factors, which could also create some endogeneity. Another concern, that is typical when using survey answers, is the possible presence of measurement errors – some individuals of the sample may state that they are bilingual but in reality, they are not, or in the other way around. The presence of these issues would bias the estimation, thus creating inconsistent OLS coefficients.

However, I consider that the results are sufficiently robust to have some policy implications, particularly in terms of education. The findings point out the importance of bolstering linguistic immersion in the countries where multiple languages coexist. Moreover, they also suggest some considerations for the other countries. Considering that an earlier exposure to two languages confers benefits beyond the ability to communicate with more people – bilingual adults show superior literacy and numeracy skills if this potential advantage is cultivated – introducing bilingual education may have a positive impact in fostering the bilingual advantage if the languages do not differ sharply between them and a relatively high level of integration between the different linguistic groups already exists. Therefore, providing educational programs in minority children's heritage languages and not segregating students will help to develop cognitive skills, which are positively associated with better educational and labor market outcomes. Additionally, this paper opens the possibility of extending the analysis with the wage and labor market effects of bilingualism, which will be the aim of my future research.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80 (2): 207-245.

Anghel, B., Cabrales, A., & Carro, J.M. (2016). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry*, 54(2): 1202-1223.

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, A. (2018). Bilingual children's advantage in academic performance. *Working Paper.* Available at: https://www.dropbox.com/s/oihnxqmo6lhaxcb/Aparicio_submission.pdf

Bialystok, E., Craick, F., Green, D., & Gollan, T. (2009). Bilingual minds. *Psychol Sci Public Interest*, 10(3): 89-129.

Bialystok, E., Craick, F., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4): 240-250.

Bloomfields, L. (1935). Linguistic aspects of science. *Philosophy of Science. The University of Chicago Press*, 2(4): 499-517.

Caminal, R., & Di Paolo, A. (2018). Your language or mine? The non-communicative benefits of language skills. *Economic Inquiry*, 57 (1): 726-750.

Capellari, L., & Di Paolo, A. (2018). Bilingual schooling and earnings: Evidence from a language-in-education reform. *Economics of Education Review,* 64: 90-101.

Castro, D. C., Páez, M. M, Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives*, 5: 15-21.

Chiswick, B., & Miller, P. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13 (2): 246-288.

Chiswick, B., & Miller, P. (1999). Language skills and earnings among legalized aliens. *Journal of Population Economics*, 12: 63-89.

Chiswick, B., & Miller, P. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1): 1-11.

Cobb-Clark, D.A., Harmon, C. & Staneva, A. (2018). The bilingual gap in children's language and Emotional Development. *IZA Institute of Labor Economics,* Discussion Paper No. 11800.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition*, 106: 59-86.

Fouka, V. (2019). Blacklash: The unintended effects of language prohibition in U.S. schools after World War I. *The Review of Economic Studies*, 87(1): 204-239.

Fry, R., & Lowell, B.L. (2003). The value of bilingualism in the U.S. labor market. *Industrial and Labor Relation Review*, 57(1): 128-140.

Gathercole, V.C.M. (2010). Bilingual children: Language and assessment issues for educators. In K. Littleton, C. Wood, & J.K. Staarman (Eds.), "International Handbook of Psychology in Education" (715-740), *Bingley, United Kingdom: Emerland Group*.

Genesee, F., Lindhom-Leary, K. (2012). The education of English language learners. In K. R. Harris, S.

Ginsburgh, V. & Weber, S. (2020). "The economics of language". *Journal of Economic Literature*, 58(2): 348-404.

Grosjean, F. (2010). The extent of bilingualism. In Grosjean, F., "Bilingual: life and reality". *Harvard University Press, Cambridge, MA.*

Hakuta, K. (1986). Mirror of language: The debate on bilingualism. *New York, NY: Basic Books.*

Haskins, R., Greenberg, M., & Fremstad, S. (2004). Federal policy for immigrant children: Room for common ground?. *Future of Children*, 14(2): 1-5.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49(1): 4-14.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language,* 39: 1-27.

Hoff, E., Laursen, B., & Brigdes, K. (2012). Measurement and model building in studying the influence of socioeconomic status on child development. In M. Lewis & L. Mayers (Eds.), "A developmental environmental measurement handbook". Cambridge, England: *Cambridge University Press.*

Hutchison, W.K. (2002). Linguistic distance as a determinant of bilateral trade. *Department of economics, Vanderbilt, Xerox.*

Isphording, I. (2014). Disadvantages of linguistic origin – Evidence from immigrant literacy scores. *Economics Letters,* 123 (2): 236-239.

Isphording, I., & Otten, S. (2013). The costs of Babylon – Linguistic distance in applied economics. *Review of International Economics*, 21(2): 354-369.

Isphording, I., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior & Organization*, 105: 30-50.

Ivlevs, A. & King, R.M. (2014). 2004 minority education reform and pupil performance in Latvia. *Economics of Education Review,* 38: 151-166.

Khavenson, T. & Carnoy, M. (2016). The unintended and intended academic consequences of educational reforms: the case of Post-Soviet Estonia, Latvia and Russia. *Oxford Review of Education*, 42(2): 178-199.

Kohnert, K. (2010). Bilingual children with primary language impairments: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, 43(6): 456-473.

Kovelman, I., Baker, S., Petitto, L-A. (2008). Bilingual and monolingual brains compared: A functional magnetic resonance imaging investigation of syntactic processing and a possible "Neural Signature" of bilingualism. *Journal of Cognitive Neuroscience*, 20 (1): 153-169.

Lang, K., & Siniver, E. (2009). The return to English in a non-English speaking country: Russian immigrants and native Israelis in Israel. *The B.E. Journal of Economic Analysis & Policy*, 9(1): Article 50.

Lleras-Muney, A. & Shertzer, A. (2015). Did the Americanization movement succeed? An evaluation of the effect of English-only and compulsory schooling laws on immigrants. *American Economic Journal: Economic Policy. American Economic Association*, 7(3): 258-290.

Locay, L., Regan, T.L. & Diamond Jr, A.M. (2013). The effects of Spanish-language background on completed schooling and aptitude test scores. *Economic Inquiry,* 51(1): 527-562.

Lohman, J. (2011). Do language barriers affect trade?. *Economics Letters*, 110: 159-162.

Muravyev, A. & Talavera, O. (2016). Can state language policies distort student's demand for education?. *Journal of Comparative Economics*, 44(2): 383-399.

Peal, E., Lamber, W.E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and applied*, 76(27): 1-23.

Pearson, B.Z. (1993). Predictive validity of the Scholastic Aptitude Test (SAT) for Hispanic bilingual students. *Hispanic Journal of Behavioral Sciences*, 15: 342-356.

Pearson, B.Z., Fernández, S.C., & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learnings,* 43: 93-120.

Perani, D., Paulesu, E., Sebastian, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S., Fazio, F., Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain: A Journal of Neurology*, 121(10): 1841-1852.

Scheele, A.F., Leseman, P.P.M., & Mayo, A.Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31: 117-140.

Serva, M. (2011). Phylogeny and geometry of languages from normalized Levenshtein distance. *Cornell University Library*. Available at: http://arxiv.org/abs/1104.4426v3

Thordardottir, E., Rothenberg, A., Rivard, M-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate assessment of two languages?. *Journal of Multilingual Communication Disorders*, 4: 1-21.

Toomet, O. (2011). Learn English, not the local language! Ethnic Russians in the Baltic states. *American Economic Review: Papers and Proceedings*, 101(3): 526-531.

Vagh, S.B., Pan, B.A., Mancilla-Martínez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Society for Research in Child Development*, 80(5): 1545-1563.

Weinreich, M. (1953). Languages in contact: Findings and problems. *The Hague: Mouton.*

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Cambridge University Press,* 14(3): 412-422.

Appendix

Figure A1
Kernel density plot of literacy scores by country.

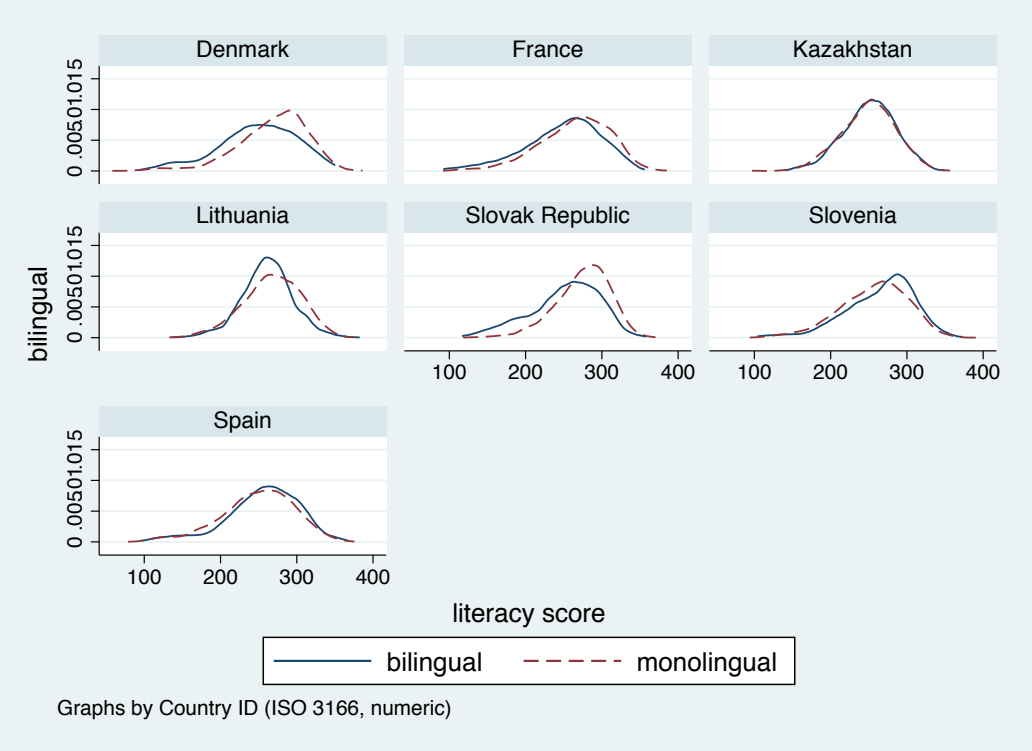

Figure A2
Kernel density plot of numeracy scores by country.

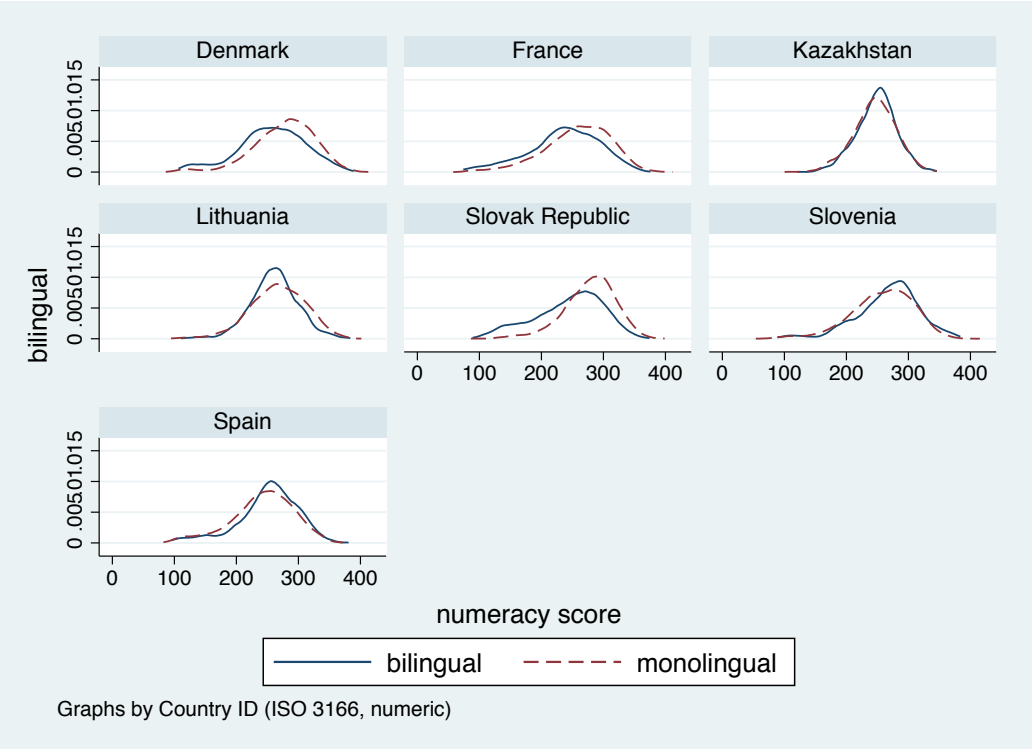

Table A1
Linguistic distance: Computational examples

Word	Spanish	English	Distance
You	tu	yu	1
Not	No	nat	2
Person	persona	pers3n	2
Night	noCe	nEit	3
Mountain	monta5a	maunt3n	5

Source: Isphording, I. (2014).

Table A2
Observations and share of bilinguals by country.

	Entire sample		Including immigrants arrived before the age of 14		Including immigrants arrived before the age of 4		Excluding immigrants	
Denmark	7,280	(6.17%)	6,100	(3.72%)	5,904	(3.17%)	5,773	(2.74%)
France	6,905	(7.60%)	6,406	(6.10%)	6,223	(5.51%)	6,105	(5.27%)
Kazakhstan	3,686	(31.58%)	3,570	(31.48%)	3,503	(31.43%)	3,460	(31.68%)
Lithuania	5,049	(18.99%)	4,994	(18.68%)	4,939	(18.36%)	4,873	(18.24%)
Slovak Republic	5,702	(10.68%)	5,661	(10.63%)	5,621	(10.48%)	5,578	(10.42%)
Slovenia	5,929	(5.84%)	4,886	(5.34%)	4,826	(5.08%)	4,758	(4.85%)
Spain	5,969	(13.33%)	5,331	(12.62%)	5,218	(12.19%)	5,179	(12.07%)
Total	39,878	(12.06%)	36,948	(11.40%)	36,234	(11.06%)	35,726	(10.92%)

Percentage of bilinguals in parenthesis.

Table A3
Descriptive statistics of the main variables for all countries.

	Bilinguals		Monolinguals		Entire sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.56	0.50	0.53	0.50	0.54	0.50
Age	39.70	13.99	41.51	14.38	41.12	14.30
Years of schooling	12.16	2.81	12.06	3.04	12.09	2.98
Both parents foreign-born	0.09	0.28	0.02	0.14	0.03	0.16
One parent foreign-born	0.10	0.30	0.05	0.23	0.06	0.24
Both parents native-born	0.81	0.39	0.93	0.26	0.91	0.28
Neither parent upper secondary education	0.43	0.49	0.40	0.49	0.39	0.49
At least one parent secondary education	0.33	0.47	0.38	0.49	0.38	0.48
At least one parent tertiary education	0.24	0.43	0.22	0.42	0.23	0.42
P.V. Literacy test	258.55	39.11	265.60	41.93	263.98	41.49
P.V. Numeracy test	255.31	43.18	265.59	47.60	263.56	46.67
L1 and L2 different to the language of the test	0.03	0.18	0.05	0.23	0.05	0.22
Distance between L1 and L2	86.46	18.44	-	-	-	-

Table A4
Descriptive statistics of the main variables for Denmark.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.49	0.50	0.44	0.50	0.49	0.5
Age	45.11	15.00	39.96	15.99	45.26	14.95
Years of schooling	12.82	2.68	12.60	2.66	12.83	2.68
Both parents foreign-born	0.01	0.11	0.22	0.41	0.01	0.08
One parent foreign-born	0.04	0.20	0.23	0.42	0.03	0.19
Both parents native-born	0.95	0.22	0.56	0.50	0.95	0.20
Neither parent upper secondary education	0.34	0.47	0.31	0.46	0.34	0.47
At least one parent secondary education	0.38	0.48	0.32	0.47	0.38	0.48
At least one parent tertiary education	0.29	0.45	0.37	0.48	0.29	0.45
P.V. Literacy test	275.33	39.67	272.34	34.96	275.42	39.80
P.V. Numeracy test	284.30	43.35	278.43	40.03	284.48	43.43
Languages different to the language of the test	0.01	0.09	0.02	0.14	0.01	0.09
Distance between L1 and L2	-	-	76.02	15.80	-	-

Table A5
Descriptive statistics of the main variables for France.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.51	0.50	0.56	0.50	0.51	0.50
Age	41.36	14.43	36.98	14.62	41.60	14.38
Years of schooling	11.53	3.47	11.43	3.40	11.54	3.47
Both parents foreign-born	0.06	0.24	0.43	0.50	0.04	0.19
One parent foreign-born	0.08	0.27	0.22	0.42	0.07	0.26
Both parents native-born	0.86	0.35	0.35	0.48	0.89	0.32
Neither parent upper secondary education	0.42	0.49	0.55	0.50	0.42	0.49
At least one parent secondary education	0.37	0.48	0.26	0.44	0.38	0.48
At least one parent tertiary education	0.21	0.41	0.19	0.39	0.21	0.41
P.V. Literacy test	268.05	43.46	265.49	39.72	268.20	43.65
P.V. Numeracy test	262.01	50.06	255.18	47.65	262.40	50.15
Languages different to the language of the test	0.02	0.14	0.02	0.14	0.02	0.15
Distance between L1 and L2	-	-	88.61	10.53	-	-

Table A6
Descriptive statistics of the main variables for Kazakhstan.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.61	0.49	0.60	0.49	0.62	0.49
Age	38.01	12.89	38.05	12.84	38.05	12.70
Years of schooling	12.33	2.27	12.31	2.21	12.26	2.20
Both parents foreign-born	0.05	0.21	0.05	0.21	0.05	0.22
One parent foreign-born	0.08	0.28	0.08	0.27	0.09	0.29
Both parents native-born	0.87	0.34	0.87	0.33	0.86	0.35
Neither parent upper secondary education	0.28	0.45	0.28	0.45	0.28	0.45
At least one parent secondary education	0.44	0.50	0.44	0.50	0.47	0.50
At least one parent tertiary education	0.28	0.50	0.27	0.44	0.25	0.43
P.V. Literacy test	250.08	35.25	250.90	34.59	249.37	35.77
P.V. Numeracy test	247.39	33.40	248.05	32.33	245.45	34.62
Languages different to the language of the test	0.26	0.44	0.03	0.18	0.36	0.48
Distance between L1 and L2	-	-	98.7	8.79	-	-

Table A7
Descriptive statistics of the main variables for Lithuania

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.60	0.49	0.62	0.48	0.60	0.49
Age	42.66	14.17	43.87	13.86	42.40	14.23
Years of schooling	13.35	2.68	13.33	2.43	13.35	2.74
Both parents foreign-born	0.02	0.13	0.04	0.19	0.01	0.11
One parent foreign-born	0.06	0.24	0.08	0.28	0.05	0.23
Both parents native-born	0.92	0.27	0.88	0.32	0.93	0.25
Neither parent upper secondary education	0.40	0.49	0.41	0.49	0.39	0.49
At least one parent secondary education	0.19	0.39	0.21	0.41	0.19	0.39
At least one parent tertiary education	0.41	0.49	0.38	0.49	0.42	0.49
P.V. Literacy test	266.43	37.33	261.96	32.94	267.43	38.17
P.V. Numeracy test	266.08	47.90	261.25	37.97	267.15	45.04
Languages different to the language of the test	0.06	0.24	0.07	0.26	0.06	0.23
Distance between L1 and L2	-	-	89.85	9.09	-	-

Table A8
Descriptive statistics of the main variables for Slovak Republic.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.53	0.50	0.54	0.50	0.52	0.50
Age	39.17	14.27	37.84	14.17	39.33	14.27
Years of schooling	12.77	2.76	11.42	2.81	12.92	2.71
One parent foreign-born	0.01	0.09	0.02	0.12	0.01	0.08
Both parents foreign-born	0.05	0.22	0.07	0.26	0.05	0.22
One parent native-born	0.94	0.24	0.91	0.29	0.94	0.23
Neither parent upper secondary education	0.30	0.46	0.50	0.50	0.28	0.45
At least one parent secondary education	0.59	0.49	0.43	0.50	0.61	0.49
At least one parent tertiary education	0.11	0.32	0.07	0.26	0.12	0.32
P.V. Literacy test	272.85	37.62	251.55	45.40	275.32	35.80
P.V. Numeracy test	274.29	45.11	243.77	55.01	277.84	42.42
Languages different to the language of the test	0.06	0.24	0.01	0.12	0.06	0.24
Distance between L1 and L2	-	-	85.70	20.21	-	-

Table A9
Descriptive statistics of the main variables for Slovenia.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.51	0.50	0.49	0.50	0.51	0.50
Age	41.01	14.23	38.91	14.59	41.12	14.20
Years of schooling	10.52	1.94	10.93	1.94	10.50	1.93
One parent foreign-born	0.04	0.20	0.32	0.47	0.03	0.17
Both parents foreign-born	0.08	0.27	0.19	0.39	0.07	0.26
One parent native-born	0.88	0.33	0.49	0.50	0.90	0.30
Neither parent upper secondary education	0.31	0.46	0.23	0.42	0.32	0.46
At least one parent secondary education	0.50	0.50	0.54	0.50	0.50	0.50
At least one parent tertiary education	0.19	0.39	0.23	0.42	0.19	0.39
P.V. Literacy test	259.93	43.23	272.57	41.25	259.28	43.23
P.V. Numeracy test	262.33	48.43	276.42	45.02	261.61	48.49
Languages different to the language of the test	0.02	0.12	0.01	0.11	0.01	0.13
Distance between L1 and L2	-	-	57.90	30.55	-	-

Table A10
Descriptive statistics of the main variables for Spain.

	Entire sample		Bilinguals		Monolinguals	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.50	0.50	0.50	0.50	0.50	1.50
Age	40.52	13.80	40.05	13.63	40.59	40.59
Years of schooling	11.21	3.58	11.65	3.51	11.16	11.16
Both parents foreign-born	0.01	0.08	0.00	0.06	0.01	0.08
One parent foreign-born	0.02	0.14	0.03	0.18	0.02	0.14
Both parents native-born	0.97	0.16	0.96	0.19	0.97	0.16
Neither parent upper secondary education	0.72	0.45	0.70	0.47	0.73	0.44
At least one parent secondary education	0.14	0.35	0.17	0.38	0.14	0.35
At least one parent tertiary education	0.13	0.34	0.15	0.35	0.13	0.34
P.V. Literacy test	253.62	46.15	261.41	44.2	252.65	46.23
P.V. Numeracy test	247.17	48.59	256.71	45.52	245.97	48.75
Languages different to the language of the test	0.03	0.17	0.01	0.09	0.03	0.17
Distance between L1 and L2	-	-	72.39	14.58	-	-

List of official languages in each country:

Country	**Official and co-official languages:**
Denmark	Danish
France	French
Kazakhstan	Kazakh and Russian
Lithuania	Lithuanian
Slovak Republic	Slovak
Slovenia	Slovenian, Hungarian and Italian
Spain	Spanish, Catalan, Galician and Basque

Table A11
Independent regression for literacy score in each country.

	Denmark litpv	France litpv	Kazakhstan litpv	Lithuania litpv	Slovak Rep. litpv	Slovenia litpv	Spain litpv
BIL	3.580	1.413	4.972	-5.324	-0.305	12.999**	20.952***
	(12.752)	(19.484)	(10.860)	(13.117)	(5.468)	(5.402)	(7.372)
DL1L2	-0.085	-0.038	-0.040	0.005	-0.158**	-0.083	-0.224**
	(0.164)	(0.216)	(0.110)	(0.144)	(0.063)	(0.082)	(0.100)
DIFF	-3.987	-9.503**	-2.427	4.365*	-13.139***	-1.478	0.636
	(7.345)	(3.820)	(1.505)	(2.633)	(2.041)	(4.450)	(3.371)
Gender (female)	-2.200***	-0.239	1.083	1.592	0.040	-0.603	-6.042***
	(0.841)	(0.950)	(1.229)	(1.010)	(0.869)	(1.065)	(1.005)
Years of schooling	6.981***	5.769***	2.211***	4.460***	4.650***	8.976***	6.484***
	(0.195)	(0.174)	(0.287)	(0.214)	(0.193)	(0.335)	(0.161)
Age	-2.065**	1.746*	-2.470*	-2.870***	-2.093**	-5.125***	-2.188**
	(0.927)	(0.972)	(1.301)	(1.070)	(0.919)	(1.152)	(0.991)
Age^2	0.043*	-0.062**	0.070**	0.035	0.042*	0.111***	0.060**
	(0.024)	(0.025)	(0.034)	(0.027)	(0.024)	(0.030)	(0.026)
Age^3	-0.000**	0.001**	-0.001**	-0.000	-0.000	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
One parent foreign-born	19.263***	5.651**	-6.063*	-3.412	5.519	6.002*	10.759
	(6.108)	(2.804)	(3.098)	(4.387)	(4.470)	(3.434)	(7.873)
Both parents native-born	19.280***	5.003**	-9.675***	-6.978*	-1.899	3.647	17.530**
	(5.774)	(2.475)	(2.442)	(4.006)	(4.156)	(3.029)	(7.242)
At least one parent secondary educ	1.257	3.562***	12.753***	0.837	15.356***	4.658***	6.330***
	(1.054)	(1.204)	(1.444)	(1.534)	(1.123)	(1.384)	(1.437)
At least one parent tertiary educ	12.039***	16.536***	17.793***	10.939***	21.470***	19.341***	11.774***
	(1.187)	(1.458)	(1.757)	(1.436)	(1.709)	(1.759)	(1.614)
cons	211.614***	190.051***	245.531***	270.359***	237.891***	237.523***	199.092***
	(11.480)	(11.244)	(15.356)	(12.914)	(11.151)	(13.284)	(13.165)
N	5725	5032	3367	4791	5529	4613	5020
adj. R-sq	0.350	0.368	0.070	0.184	0.259	0.302	0.398

Notes: OLS estimations are reported with robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.

Table A12
OLS estimations including official languages.

	litpv	numpv
BIL – 1 official	14.766***	11.273**
	(5.183)	(5.646)
BIL – 2 official	12.710***	13.170***
	(2.948)	(3.306)
BIL – 1 of x Denmark	-12.380**	-12.645**
	(4.952)	(5.315)
BIL – 2 of x Denmark	-	-
BIL – 1 of x France	-8.924*	-10.438**
	(4.724)	(5.071)
BIL – 2 of x France	-	-
BIL – 1 of x Kazakhstan	3.041	7.581
	(5.241)	(5.363)
BIL – 2 of x Kazakhstan	-2.606	-5.019**
	(2.377)	(2.455)
BIL – 1 of x Lithuania	-10.086**	-8.407*
	(4.436)	(4.695)
BIL – 2 of x Lithuania	-	-
BIL – 1 of x Slovak Rep	-18.879***	24.386***
	(4.558)	(4.878)
BIL – 2 of x Slovak Rep	-	-
BIL – 1 of x Slovenia	-1.101	3.561
	(5.405)	(5.739)
BIL – 2 of x Slovenia	8.836**	4.765
	(4.174)	(4.565)
DIFF	-4.315***	-6.552***
	(0.987)	(1.035)
DL1L2	-0.111***	-0.095
	(0.035)	(0.040)
Demographical controls	Yes	Yes
Parental controls	Yes	Yes
cons	230.457***	214.834**
	(4.634)	(5.094)
N	34077	34077
adj. R-sq	0.307	0.354

Table A13
Marginal effects of bilingualism of 1 or 2 official languages by country.

	litpv	numpv
Denmark - 1 official	2.387	-1.372
	(3.671)	(4.145)
France - 1 official	5.842	0.835
	(3.768)	(4.359)
Kazakhstan - 1 official	17.808***	18.854***
	(4.364)	(4.640)
Lithuania - 1 official	4.680	2.866
	(3.432)	(3.933)
Slovak Republic -1 official	-4.113	-13.113***
	(3.330)	(3.867)
Slovenia - 1 official	13.666***	14.834***
	(3.228)	(3.495)
Spain - 1 official	14.766***	11.273**
	(5.183)	(5.646)
Kazakhstan - 2 official	10.105***	8.150*
	(3.840)	(4.339)
Slovenia - 2 official	21.546***	17.935***
	(4.923)	(5.528)
Spain - 2 official	12.710***	13.170***
	(2.948)	(3.306)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.

Table A14
OLS estimations including controls for household characteristics.

	litpv	numpv
BIL	9.293**	9.348**
	(4.151)	(4.255)
BIL x Denmark	-4.718	-8.858**
	(3.801)	(3.976)
BIL x France	-4.770	-10.572***
	(3.795)	(4.068)
BIL x Kazakhstan	1.627	0.343
	(3.060)	(3.048)
BIL x Lithuania	-6.082**	-8.023***
	(2.876)	(2.978)
BIL x Slovak Rep	-15.805***	-25.280***
	(2.982)	(3.173)
BIL x Slovenia	9.546**	9.275**
	(3.746)	(3.758)
DL1L2	-0.087*	-0.067
	(0.050)	(0.051)
DIFF	-4.235	-1.042
	(4.516)	(4.184)
DIFF x Denmark	-11.289	-22.576*
	(11.541)	(11.946)
DIFF x France	-11.660*	-14.018**
	(6.772)	(6.564)
DIFF x Kazakhstan	2.626	-0.837
	(4.871)	(4.564)
DIFF x Lithuania	8.938*	3.284
	(5.412)	(5.316)
DIFF x Slovak Rep	-9.430*	-19.412***
	(5.155)	(5.060)
DIFF x Slovenia	-1.790	-3.393
	(7.760)	(7.371)
Gender (female)	-1.384***	-9.618***
	(0.487)	(0.526)
Years of schooling	5.829***	7.086***
	(0.095)	(0.102)
Age	7.389***	6.689***
	(0.970)	(1.045)
Age^2	-0.155***	-0.132***
	(0.022)	(0.024)
Age^3	0.001***	0.001***
	(0.000)	(0.000)
One parent foreign-born	-0.789	0.867
	(1.952)	(2.088)
Both parents native-born	-3.921**	-2.375
	(1.734)	(1.857)
At least one parent secondary educ	5.838***	8.010***
	(0.581)	(0.628)
At least one parent tertiary educ	11.727***	14.719***
	(0.740)	(0.795)
Number of children	-1.068***	-1-117***
	(0.254)	(0.276)
Living partner	1.490***	3.755***
	(0.742)	(0.804)
cons	83.526***	69.526***
	(13.696)	(14.751)
N	20627	20627
adj. R-sq	0.316	0.388

Notes: OLS estimations are reported with robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels. All the regressions include country fixed effects.

Table A15
Marginal effects of bilingualism by country.

	litpv	numpv
Denmark	4.575	0.490
	(4.740)	(4.843)
France	4.523	-1.223
	(5.375)	(5.664)
Kazakhstan	10.919**	9.691*
	(5.160)	(5.188)
Lithuania	3.211	1.325
	(4.796)	(4.912)
Slovak Republic	-6.512	-15.931***
	(4.621)	(4.664)
Slovenia	18.838***	18.624***
	(4.888)	(4.539)
Spain	9.293**	9.348**
	(4.151)	(4.255)

Notes: Robust Standard Errors in brackets. ***,** and * indicate significance at the 1%, 5% and 10% levels.