"Linguistic skills and the intergenerational transmission of language"

Caminal R, Cappellari L and Di Paolo A

IREA
Research Institute of Applied Economics

AQR
Regional Quantitative Analysis Research Group

Abstract

We investigate the pattern of intergenerational transmission of language in a bilingual society. We consider the case of Catalonia, where the two main speech communities, Spanish and Catalan, are of similar sizes, both languages are official, and each one enjoys the protection of a different layer of government. However, whereas all native Catalan speakers are bilingual, only a fraction of native Spanish speakers are fully proficient in Catalan. In this environment, transmission decisions are far from trivial. We provide a simple theory showing that changing the costs of acquiring a second language affects language skills as well as the language parents speak to their children. Empirically, we exploit the natural experiment generated by a language-in-education reform that introduced Catalan–Spanish bilingualism at school to estimate the causal effects of language skills on intergenerational transmission. Results show that the increased proficiency in Catalan among native Spanish speakers induced by the reform augmented their propensity to speak Catalan to their offspring. The causal effect of the policy-induced increase in language proficiency on language transmission is not confounded by spurious trends, potential changes in language identity, and linguistically mixed partnership formation.

JEL Classification: I28, J13, J24, J62, Z13.

Keywords: language, intergenerational transmission, language proficiency, language-in-education reform.

Ramon Caminal: Institute of Economic Analysis, CSIC, and Barcelona GSE. Email: ramon.caminal@iae.csic.es

Lorenzo Cappellari: Catholic University of Milan and IZA. Email: lorenzo.cappellari@unicatt.it

Antonio Di Paolo: AQR-IREA, University of Barcelona, Av. Diagonal 690 (08034), Barcelona, Spain. Email: antonio.dipaolo@ub.edu

1. Introduction

Language skills impinge on many social and economic outcomes. In particular, it has been extensively documented that immigrants proficient in the host-country language enjoy enhanced employability and substantial wage premia. Bleakely and Chin (2004) and Chiswick and Miller (2007) are just two prominent examples of a fairly large and geographically diverse literature.[1] Thus, promoting additional language skills alleviates wage inequality and is likely to raise efficiency owing to the positive externality generated by the acquisition of a second language (Selten and Pool, 1991; Church and King, 1993).

A few studies have also examined the transmission of language skills between first- and second-generation immigrants. More specifically, Bleakley and Chin (2008) and Casey and Dustmann (2008) for the cases of the US and the UK, respectively, found that the parents' English-speaking proficiency yields significant, positive results for children's English proficiency. These findings imply that the family background influences the individual's language repertoire; that is, parents' decisions are not fully offset by other sources of skills, such as the school. Therefore, language-related inequalities are likely to persist across generations.[2]

This evidence also raises some important questions. First, through what mechanism are language skills inherited? The children's language skills may be affected by the language spoken within the family. Alternatively, the parents' proficiency in the host-country language may have an impact on residential and school choices. Whereas the former is an intimate decision, the latter could be affected by various policy

---

[1] In a similar vein, starting with Frankel and Rose (2002), it has been shown that the knowledge of foreign languages significantly influences the pattern of international trade (see also Mélitz, 2008, and Egger and Lassmann, 2015).

[2] More recently, Kuziemko (2014) demonstrated the existence of reversed intergenerational transmission, with the proficiency of immigrant parents being negatively affected by their children's proficiency. That is, parents lean on, rather than learn from, their children.

instruments. Second, what are the parents' motivations behind their language choices? On the one hand, they may be inclined to transmit the language that embodies their cultural heritage or their ethnic identity. On the other hand, they may consider passing on the host-country language to foster their children's integration, shelter them from discrimination, and ultimately improve their economic outcomes in the long-run.[3] Third, what are the determinants of the intergenerational transmission of language in bilingual societies? Previous studies have focused on immigrants in societies with a predominant language. Clearly, parents also face non-trivial language choices in environments in which there is more than one language that cannot be considered a "minority" language. There, the key issues are not so much related to the speed of assimilation, but to the economic value of bilingualism, the level of endogamy of different speech communities, and the long-run contest between competing languages for dominance. Understanding the pattern of language transmission in bilingual societies involves a whole new set of issues.

In this paper, we address some of these questions. We study the pattern of intergenerational transmission of language in an asymmetric bilingual society. In such an environment, the behavior of the speakers of the stronger language is as relevant as the behavior of the speakers of the weaker language. We pay special attention to the role of language policies and their long-lasting effects on language-related inequality.

We focus on the Spanish region of Catalonia. Catalonia is a particularly interesting case study for two reasons. First, it is a pure example of an asymmetric bilingual society (with two large speech communities) in which communication is not at stake owing to the universal knowledge of one of the languages. In particular, Spanish

---

[3] Related to this point, Algan *et al*. (2013) shown that the economic incentives affects the transmission of cultural identity, captured by naming patterns. Specifically, they found that the employment penalty associated with Arabic names in France reduces the probability of assigning an Arabic name to the new-borns.

and Catalan are both official languages in Catalonia and have coexisted for ages. Moreover, they have similar numbers of native speakers. However, they do not hold symmetrical positions. Spanish is a very strong global language, with more than 400 million native speakers. In contrast, Catalan is only spoken in the territories of the former Crown of Aragón, and less than six million people use it on a daily basis. Thus, the differential incentives to learn Spanish and Catalan are huge. As a result, all native Catalan speakers are fully bilingual (high proficiency in both Catalan and Spanish), whereas only a fraction of native Spanish speakers are proficient in Catalan. Second, in the 1980s the regional authorities implemented a durable language-in-education reform that introduced Catalan as a medium of instruction (together with Spanish) in Catalan schools, which significantly enhanced the Catalan proficiency among native Spanish speakers.[4]

The universal knowledge of Spanish implies that any real effects generated by any new language skills induced by the policy reform must be associated with the role of languages beyond communication. Furthermore, the large-scale reform, which was unexpected and gradually implemented, represented a "natural experiment". In particular, we are able to exploit the heterogeneous effect of compulsory exposure to the reform by native language to identify the causal effects of skills on the language spoken with their children.

For the case of Catalonia, and despite the fact that the ability to communicate has never been at stake, there is ample empirical evidence on the value of bilingualism. In particular, it has been shown that knowledge of Catalan fosters employability (Rendon, 2007) and increases earnings (Di Paolo and Raymond, 2012; Cappellari and Di Paolo,

---

[4] This reform was a crucial step in the process of promoting the knowledge and use of Catalan after Franco's regime, when the public use of the language was banned.

2018). It also affects social outcomes; in particular, it increases the frequency of mixed marriages (Caminal and Di Paolo, 2017).[5]

Thus, the fact that bilingualism is rewarded by the market and, moreover, improves the quality of other social interactions is likely to affect parents' decisions regarding language transmission. Indeed, whereas virtually all native Catalan speakers use Catalan with their children, the behavior of native Spanish speakers shows great variability, with approximately a third speaking Catalan with their children.

We first present a simple choice-theoretic model of the intergenerational transmission of language in a bilingual society, which provides testable predictions and helps frame the data analysis. The model portrays the optimal choice of the language(s) parents use with their children as balancing their own language preferences with the welfare of their children. Individuals are assumed to have a preferred language, but they are willing to (at least, partially) switch to their second language, provided that a more mixed linguistic repertoire has a sufficiently positive effect on the welfare of their children. Parents with better proficiency in their second-language experience a lower cost of switching. Consequently, the main prediction of the model (and the main testable implication) is that an exogenous increase in the proficiency in the second language is expected to increase the intensity with which such a language is transmitted to the next generation. Thus, policies that change current language skills are also expected to have longer-term effects.

We next exploit language-in-education policy changes to estimate the causal effect of parental language skills on the choice of language spoken with children for the

---

[5] There is a growing literature in economics showing that languages are much more than neutral communication devices. A number of studies have shown that linguistic diversity (taken as a proxy of ethnic diversity) in a political union is associated with more conflict and lower redistribution. See, for instance, Alesina *et al.* (2003) and Desmet *et al.* (2013). More recently, different characteristics of languages have been associated with different economic behaviors regarding various intertemporal decisions, as well as female labor participation (Chen, 2013; Gay *et al.*, 2013; Galor e*t al.*, 2016).

case of Catalonia. We can access unique survey information on the linguistic and socio-economic characteristics of the respondents, their language background, as well as the language spoken with their children.

The language-in-education reform of 1983 followed on from the approval of the Language Normalization Act (LNA) and induced a sharp change in the educational system: starting from that academic year, students in any grade of primary and compulsory secondary school were taught using both Catalan and Spanish. Therefore, individuals that were at school in 1983 or who started school after that year were partially or completely exposed to the bilingual regime, whereas older individuals were entirely educated in the monolingual regime (in Spanish). The sociolinguistic characteristics of the Catalan population in the early 1980s were crucially affected by the massive immigration flows from the rest of Spain that took place in the 1960s and early 1970s. Since the population of Catalan origin was already proficient in oral Catalan, bilingualism at school boosted Catalan skills among the population of Spanish origin. We leverage year of birth and native language to construct an instrumental variable (IV) for the Catalan skills of native Spanish speakers, which exploits potential exposure to bilingualism at school during compulsory education.[6] The validity of the instrument relies on the assumption that direct cohort effects in language transmission in addition to the ones operating via language fluency are common between Catalan and Spanish speakers, which we provide corroborating evidence for.

We show that parents' language skills matter in the intergenerational transmission of language. The reform had a significant impact on proficiency in Catalan for Spanish natives, increasing it by on average around one point on a 10-point scale (15% increase) for those who were exposed to bilingualism for the whole of their compulsory

---

[6] Clots-Figueras and Masella (2013) used the exposure to the language-in-education reform as an exogenous determinant of national identity in Catalonia.

education, relative to the benchmark of no exposure. Using the proposed IV, we find that a unit increase in proficiency increases the intensity of intergenerational Catalan use for Spanish native speakers by 40% of a point on the 1–5 scale, corresponding to an 18% reduction in the average gap in Catalan use between native Spanish and native Catalan speakers, which is in line with the main prediction of the theoretical model.

Our results are consistent with the idea that policies that reduce the cost of acquiring a second language can significantly change language skills with spill-overs onto the next generation. In the last part of the paper we provide several robustness checks to corroborate the validity of our results. The evidence from falsification exercises indicates that our exclusion restriction is not capturing spurious trends across the cohorts. The results are also stable when we adopt alternative definitions of native languages.

The theoretical model emphasizes the role of the parents' language skills in intergenerational language transmission. However, it also considers possible alternative mechanisms that could potentially explain the empirical evidence. First, the educational reform might have changed not only skills but also the preference for the language to be transmitted to the children. To address the likelihood of this alternative channel, we exploit information on the respondents' self-identification language. We argue that if changes in preferences are a relevant channel, then results should be affected by the exclusion of "language switchers" (*i.e.* individuals whose self-identification language is different from their mother tongue). We show that they are not, which suggests that the potential changes in language preferences induced by the reform are not the main driver of the more intense use of Catalan by native Spanish speakers with their children.

Second, it could also be the case that enhanced language skills induced by the reform did not have a direct effect on language transmission, but rather the effect was

through partnership formation. That is, native Spanish speakers with better Catalan proficiency are more likely to marry a native Catalan speaker (Caminal and Di Paolo, 2017). Hence, the more intense use of Catalan with the children could be owing to the differential behavior of mixed couples. However, when we exclude linguistically mixed couples, the results also remain unaffected.

The remainder of the paper proceeds as follows: In Section 2 we provide a brief historical account of the sociolinguistic dynamics of Catalonia. Section 3 presents the theoretical model. Section 4 contains a description of the data used in the empirical analysis. The empirical framework is detailed in Section 5. Section 6 discusses the main results and Section 7 examines the evidence from the robustness checks. Section 8 concludes.

2. Historical background

Catalan and Spanish (Castilian) are both Romance languages. Catalan is the variety of Vulgar Latin that developed in the territory from the 9th century. Spanish arrived in Catalonia as early as the 15th century and consolidated its position among the elite in the 18th century. In the 19th and early 20th centuries, the expansion of elementary education increased the knowledge of Spanish among the general population.

Franco's regime (1939–1977) represented a serious setback for Catalan. It imposed Spanish as the only official language and the only language to be used in education. Catalan was excluded from most social activities but, nevertheless, it was still intensively used in the private sphere and, more importantly, transmitted across generations. The massive migration flows from the south of Spain to Catalonia in the 1960s and early 1970s dramatically changed the demographics of the region. In the mid-1970s, 40% of the Catalan population was born outside Catalonia. Urban

segregation of immigrants and the ban on the use of Catalan in schools and media meant that, for the first time in history, a very large fraction of the Catalan population was monolingual in Spanish (Woolard and Gahng, 1990; Siguan, 1991).

The political environment drastically changed with the advent of democracy. In particular, new language-in-education policies were introduced after the approval of the Language Normalization Act (LNA) in 1983. The primary goal of the LNA was for all students to be fully bilingual in both Catalan and Spanish by the end of their compulsory education. It also established an education system in which students were not separated on the basis of their native language. With the implementation of this reform, education slowly evolved from a system in which Catalan was excluded to one in which Catalan was the main language of instruction in compulsory education.[7] The reform led to a significant improvement of the Catalan skills of native Spanish speakers, whereas the Spanish skills of both native Spanish and Catalan speakers remained essentially unchanged (at excellent levels).[8]

Nowadays, the position of the two languages in Catalan society is still asymmetric, but the situation is clearly more complex. Spanish is the dominant language in Spain, the official language throughout all its territory, and backed by the central government. Moreover, it is one of the most important global languages with almost 400 million native speakers and probably more than 200 million second-language speakers. Clearly, the private and social incentives to learn Spanish are huge. Catalan enjoys strong support from the regional government. Besides Catalonia, it is also spoken in other territories of the ancient Kingdom of Aragón, mainly Valencia and the Balearic Islands. Overall, less than six million people speak Catalan on a daily basis.

---

[7] See Clots-Figueras and Masella (2013), or Cappellari and Di Paolo (2018), for more details about the LNA reform.
[8] See Section 4 for some descriptive figures about language proficiency by native language.

Consumption of media and cultural goods in Catalonia is largely in Spanish, with the exception of radio and books where the share of Catalan fluctuates around 50% and 30%, respectively. Local and regional politics are conducted mostly in Catalan, but in other branches of government (such as judicature) Catalan is almost non-existent.

Three facts deserve special consideration. First, all Catalan citizens are fully proficient in Spanish, but not all are proficient in Catalan. This asymmetry is not likely to die out any time soon. Even though basic skills in Catalan are widespread among the younger generations, the distribution of Catalan skills still exhibits high variance. Thus, in terms of proficiency, Spanish is still the dominant language, but less so than a few decades ago. Second, the labor market seems to consistently reward bilingual individuals, as reported by Rendon (2007), Di Paolo and Raymond (2012), and Cappellari and Di Paolo (2018). Third, bilingualism also provides non-monetary private and social rewards. In particular, increased fluency in Catalan among native Spanish speakers has increased their chances of finding a partner from the other language community (Caminal and Di Paolo, 2017).

3. Theoretical framework

3.1 The benchmark model

In this section, we lay out a simple choice-theoretic model of language transmission in an asymmetric bilingual society. The model aims at capturing some of the crucial trade-offs involved in such a decision with no claim of generality. On the contrary, since the goal is to provide a plausible interpretation of the empirical results, we will make special assumptions that aim at representing the specific case of Catalonia as closely as possible. The main focus is on the language(s) that parents use with their children. Such

parental choice will be an important (but not the only) determinant of the children's future linguistic repertoire, which in turn will influence their wellbeing.

We analyze a society with two languages, A and B. The initial number of native speakers of each language is similar.[9] However, the status of these two languages is not symmetric. Language A is a strong global language and hence the incentives to learn it are huge. In contrast, the incentives to learn language B are only linked to its domestic use, and hence are much more limited.

The model considers two stages and two generations, old and young, of the same size. In the first stage, each member of the old generation (the parent) is matched to a member of the young generation (the child), forming a family. Parents differ in their linguistic background (skills and preferences). Each parent chooses the language(s) used in the family, which determines the child's linguistic background at the beginning of the second stage. During the second stage, the members of the young generation can make costly investments that further improve their proficiency in these languages.

We describe parents' linguistic choices as well as children's posterior decisions as concerning continuous variables. This facilitates the analysis as well as the connection between the model and the data. Parents are assumed to have rational expectations about their children's choices. Thus, we use backwards induction and first examine the second stage. The family linguistic heritage (native language(s)) of a member of the young generation is represented by a pair (x_A, x_B), where $x_i \in [0,1]$ is the intensity of the use of language i within the family, $i = A, B$, and the resource constraint is $x_A + x_B = 1$. Thus, $x_i = 1$ denotes an individual whose unique native language is i. The family is not the unique source of language skills. The influence of other factors (including the school) are modelled as the individual's optimal choice, balancing costs and benefits.

---

[9] Hence the labels that are typically used to reflect the relative size of their speech communities, "majority" or "minority", cannot be applied here.

We let $I_i \in [0,1]$ be the individual's proficiency in language i, with which they arrive to adulthood (and to the labor market). Individuals can improve their proficiency in language i by an amount $I_i - x_i \geq 0$ by incurring a cost $C_i(I_i - x_i)$, where $C_i()$ is an increasing and strictly convex function. The benefits of proficiency I_i are denoted by $B_i(I_i)$, which is an increasing and strictly concave function.[10] Notice that, for simplicity, the benefits of acquiring a language only depend on the level of proficiency in that particular language and not on the proficiency in the other language or on the distribution of language skills in the society. In particular, even in the case that everyone is fully proficient in language A and hence communication is not at stake, we assume that learning language B still provides additional benefits. The asymmetry between these two languages will be captured by a single parameter, γ. In particular, for a fixed value of proficiency, z, we let $B_A(z) = \gamma B_B(z) \equiv \gamma B(z)$, $\gamma > 1$. Thus, learning language A generates higher benefits. For simplicity, we assume that cost functions are identical: $C_A(I - x) = C_B(I - x) \equiv C(I - x)$.[11]

Given (x_A, x_B), an individual is expected to choose her proficiency in both languages (I_A, I_B) in order to maximize the difference between benefits and costs. Hence, in the case of an interior solution, the optimal choice, I_i^*, will be given by $B_i'(I_i^*) = C_i'(I_i^* - x_i)$. Two technical remarks are in order. First, notice that $\frac{dI_i^*}{dx_i} \in (0,1)$. Hence, $C_i'(I_i^* - x_i)$ decreases with x_i. Second, if $\gamma B'(1) \geq C'(1)$, then for all x_A, the optimal level of proficiency in language A is $I_A^* = 1$. That is, individuals of all backgrounds will end up being fully proficient in language A. We will maintain this

---

[10] Throughout the analysis we also assume that this function, as well as the rest of the functions introduced below, are twice continuously differentiable.

[11] It could make sense to assume that learning language A is cheaper, because of its more frequent use in the media and in non-domestic interactions. Introducing this additional asymmetry would reinforce our results.

assumption throughout this section, and hence $I_A^*(x_B) = 1$.[12] In contrast, $I_B^*(x_B)$ will typically be an increasing function of x_B. More specifically, we assume that $B'(0) - C'(0) > 0$ so that $I_B^*(x_B) > 0$.

It will be useful to summarize the implications of different inherited linguistic capacities for the welfare of a member of the young generation. Since $x_A + x_B = 1$, then such a family background can be represented by a single variable, x_B. In particular, we let $\psi(x_B)$ be the indirect utility function of a member of the young generation with background x_B:

$$\psi(x_B) = \gamma B(1) - C(x_B) + B\big(I_B^*(x_B)\big) - C(I_B^*(x_B) - x_B) \tag{1}$$

Note that $\psi'(x_B) = -C'(x_B) + C'(I_B^* - x_B)$ and $\psi''(x_B) < 0$. Moreover, $\psi'(0) > 0$. Consequently, the value of x_B that maximizes ψ, $\bar{x}_B$, is always strictly positive and lower than one.[13] Hence, the most profitable linguistic background includes a positive skill in language B, the relatively weaker language. The reasons are twofold. First, bilingualism is positively rewarded in society, even though everyone is fully proficient in language A. Second, the learning cost function is convex, which means that a mixed linguistic background minimizes total costs.

We next turn to the first stage and examine the linguistic decisions of the members of the old generation. The asymmetry between the two languages is also reflected in the distribution of parents' linguistic skills, which we take as exogenous. All parents are fully proficient in their native language, but only native B speakers are fully proficient in language A, whereas not all native A speakers are proficient in language B.

---

[12] This assumption is not necessary for the main comparative static result discussed below.
[13] If the costs of learning language A are lower, then $\bar{x}_B$ is higher, and it can even be equal to 1. In addition, $\bar{x}_B$ decreases with γ, but $\bar{x}_B > 0$ even if $\gamma = 1$.

In particular, we denote by parameter $\mu \in [0,1]$ the level of proficiency in language B of a native A speaker: $\mu = 1$ indicates that the parent is fully bilingual and $\mu = 0$ is purely monolingual. In the case of native B speakers, the level of proficiency in the second language is $\mu = 1$.

Parents also have heterogeneous preferences about the language they use, in particular with their children. The intensity of these preferences is denoted by a parameter $\lambda \in [1,2]$. For now, suppose that individuals have a preference towards their native language (to be relaxed). In particular, the higher the value of λ, the stronger the preference for using their native language with their children. Moreover, since the minimum value of λ is equal to 1, then the net cost of switching to the second language is proportional to $\lambda - \mu$, which is always positive.

A native A speaker with parameters (λ, μ) chooses the fraction of time, $x_B \in [0,1]$, that she uses language B with her child in order to maximize her objective function:

$$U = -(\lambda - \mu)f(x_B) + \psi(x_B) \tag{2}$$

where $f(.)$ is strictly increasing and convex. The first term represents the disutility of using her second language (B). Notice that a higher value of x_B is associated with lower utility. Moreover, for a given value of x_B, both a higher value of λ and a lower value of μ imply lower utility. Since they are both entered linearly in the objective function, preferences and language skills are modelled as perfect substitutes. The second term represents the economic reward of their children's (endogenous) linguistic repertoire. If the solution is interior, then it is given by the following first order condition:

$$-(\lambda - \mu)f'(x_B) + \psi'(x_B) = 0. \tag{3}$$

That is, the optimal value of x_B, x_B^*, balances the parent's disutility from using the second language and the children's payoff derived from the family linguistic background (preferences versus economic incentives.) Notice that the cost of switching decreases with language skills. That is, using the implicit function theorem, we derive the main comparative static result: $\frac{dx_B^*}{d\mu} > 0$, which is the main hypothesis tested in this paper:

Prediction: An exogenous increase in the proficiency in language B among native A speakers induces a more intensive use of language B with their children.

For completeness, we now consider the choice of a parent who is a native B speaker. In this case, her utility function can be written as:

$$(\lambda - 1)f'(1 - x_B) + \psi'(x_B) = 0 \tag{4}$$

Unsurprisingly, *ceteris paribus*, native B speakers uses language B more often than native A speakers. In fact, $x_B^* = 1$ provided $(\lambda - 1)f'(0) + \psi'(1) \geq 0$.

Our formulation of the parent's objective function is related to two different strands of the literature on intergenerational transmission. First, the language choice influences the position of the children in the labor market, which is reminiscent of the literature on the intergenerational transmission of inequality (see, for instance, Solon, 2014). The main difference is that in this literature parents pay for the physical resources (time and money) required to build the children's stock of human capital,

whereas in our case the costs are exclusively associated with linguistic preferences. Second, the language choice is also determined by parental preferences and, in this sense, is related to the literature on the intergenerational transmission of cultural traits (see, for instance, Bisin and Verdier, 2001). However, our simple model focuses on the transmission of language skills and ignores transmission of values or linguistic preferences.

Both the assumptions and predictions of the model fit the Catalan case reasonably well. In particular: if we identify language A with Spanish and language B with Catalan, then (i) the probability of being bilingual (high proficiency in both languages) is higher for native Catalan speakers, (ii) bilingualism is rewarded in the labor market, despite the fact that everyone is proficient in Spanish, and (iii) native Spanish speakers transmit Catalan to their children, although less intensively than native Catalan speakers (descriptive statistics corroborating these statements are provided in Section 4).

This model cannot adequately represent societies with an undisputed dominant language (for instance, English in the US). In those cases, a bilingual immigrant worker, who has learnt the dominant language in addition to her native language, could be worse off than a monolingual, native speaker of the dominant language as far as their labor market position is concerned (Chiswick and Miller, 2017). The potential premium associated to the knowledge of the immigrant's native language is very small, and probably dominated by the possibility of being discriminated against on the basis of a peculiar accent. Thus, exclusively transmitting the dominant language to the next generation (language shift) is often an optimal strategy. In a more balanced bilingual society, discrimination on the basis of accents is much less likely (since it will be so pervasive) and in contrast the economic value of bilingualism is likely to be higher.

3.2. Alternative mechanisms

a) The role of mixed couples

Suppose now that the language used with the children is cooperatively chosen by two parents; that is, two different individuals with potentially different language skills and preferences.[14] If they are both native A speakers, then the value of x_B that maximizes the sum of the utilities of both parents is clearly an intermediate value of their individual optimal choices. In the case of a mixed couple (each member has a different native language) the solution is also intermediate, but since the optimal value of x_B for a native B speaker will tend to be higher, then a native A speaker in a mixed couple is expected to choose language B more intensively than in a homogeneous couple owing to a stronger *averaging* effect.

If parental decisions are taken by two heterogeneous individuals then it could make sense to model language transmission as a two-variable choice: the fraction of time each parent uses language B with their children (which can be chosen cooperatively or non-cooperatively). It is still natural to assume significant spill-overs: each parent cares about both variables, but presumably with different intensities. It is straightforward to see that this alternative formulation opens the door to a *substitutability* effect: The native A speaker in a mixed couple has incentives to use language B less intensively than the partner in an attempt to minimize the total costs associated with the use of a second language. If such a *substitutability* effect dominates the above *averaging* effect then a native A speaker would be expected to use language B less intensively in a mixed couple. Of course, which effect dominates is an empirical matter.

---

[14] Thus, a family is formed by two members of the old generation and an undetermined number of members of the young generation.

b) Preference switching

In the previous discussion, we suggested that preferences could be determined by the individual's native language. Although the native language is likely to be a fundamental determinant of linguistic preferences, individuals are exposed to other influences that may also affect their preferences.[15] The benchmark model also provides an immediate comparative static result that could be relevant in the interpretation of the empirical results: if native A speakers reduce their relative preference for using language A with their children (lower value of λ) or, more drastically, replace language A by language B as their preferred language, then they will use language B more intensively with their children. Hence, if the same exogenous (policy) change that alters language skills can also affect language preferences, then we would need to pay attention to the relative strength of each of these channels.

4. Data description

The data are drawn from the Survey of Language Use of the Catalan Population, which is produced by the Catalan Statistical Institute (IDESCAT), is representative of the Catalan population, and is available for survey years 2008 and 2013. The survey contains unique information on sociolinguistic characteristics, such as native language (first language spoken at home during childhood), self-identification language, proficiency in both Catalan and Spanish, parental native language, (current or former) partner's language and, most importantly, the language normally spoken with the first three children. It also provides detailed sociodemographic characteristics (gender, year of birth, place of birth, place of residence, education, etc.) and parental background

---

[15] Moreover, some individuals have more than one native language, in which case the explanatory value of the native language would be ambiguous.

(father's and mother's place of birth, language, and the level of education attained). All variables are self-reported by the respondent.

We focus on individuals who were entirely educated in Catalonia and exclude from the estimation sample those born in other Spanish regions who migrated to Catalonia at age 6 or older. Foreigners are also excluded from the estimation sample. The first birth-cohort that experienced some exposure to bilingualism in compulsory education contains those born in 1970, who attended the last grade of compulsory school in 1983. We build our sample symmetrically around this pivotal cohort and include in the analysis individuals born from 1950 up to 1990, implying that the youngest respondent is aged 18 and the oldest 63 at the date of interview.[16]

Virtually all selected respondents report that either Spanish or Catalan is their native language or the language they speak with children, and we deleted the few (120) observations for which this is not the case. Roughly 60% of respondents who meet the previous criteria have at least one child and can contribute to the analysis. After applying all the data selections described above and eliminating some observations with missing values in the relevant variables, we obtain a sample of 3,667 observations (out of about 14,000 cases available in the two waves of the survey).

Table 1 reports descriptive statistics for the variables used in the empirical analysis. Native language (of the individuals, their parents, and eventually their partner) is originally classified into three categories (not considering other languages): 1) only Catalan, 2) Catalan and Spanish, and 3) only Spanish. There are very few cases of mixed native language (4.2%). We consider individuals to be native Spanish speakers if they declare their native language to be only Spanish (the results are robust for this choice, as we show later). According to this definition, 45% of individuals are native

---

[16] We use data for older cohorts to conduct placebo exercises, see Section 6.

Spanish speakers. Although the survey reports separate information on several dimensions of proficiency (understanding, speaking, writing, and reading), here we focus on oral proficiency in Catalan, since oral fluency is the most relevant domain of language proficiency for language transmission from parents to children. Self-reported speaking skills in Catalan are coded on a 0–10 scale, with higher levels referring to more advanced language skills. Native Catalan speakers are virtually fully proficient in their language (77% report a value of 10), whereas the average oral skills in Catalan are 1.9 points lower for their native Spanish speaker counterparts (the difference is statistically significant).

Although we have separate information on the language spoken with each of the first three children, we only consider the first child, both to preserve sample size (respondents with at least two children are 64% of the selected sample) and because there is little between-children variation in language spoken with children by birth order. The language spoken with children is recorded into five categories, namely: only Spanish (1), more Spanish than Catalan (2), equal Spanish and Catalan (3), more Catalan than Spanish (4), and only Catalan (5). The large majority of native Catalan speakers transmit their native language to the firstborn child. Conversely, a substantial amount of variation is observed among native Spanish speakers. Only 39% of them speak only Spanish to their firstborn child, whereas 25% speak only Catalan, and the remainder mixes both languages.

5. Empirical framework

We now set out an empirical framework to test the main prediction of the theoretical model of Section 3 by estimating the effect of parental linguistic skills on the choice of language parents speak with their children. As we know from the previous sections,

native Catalan speakers are fully proficient in Catalan and almost exclusively use Catalan with their children. On the other hand, native Spanish speakers are not in general fully proficient and exhibit more variability in their choice of language transmission. We have information on native language (a dummy variable S_i that equals 1 if individual i reports Spanish as the native language), the language spoken with the firstborn child (an ordinal variable L_i taking values from 1 to 5 on a monotonic scale of intensity of use of Catalan), and the self-assessed degree of oral proficiency in Catalan (a variable C_i that takes values from 0 to 10 for increasing levels of proficiency).

Assuming that the native language is exogenous, the equation of interest is as follows:

$$L_i = \alpha + \beta' X_i + \delta C_i + \theta_{\tau(i)} + \varepsilon_i \ if \ S_i = 1 \tag{5}$$

where $\tau(i)$ is person i's year of birth and θ is a year of birth fixed effect to control for secular trends, the vector X_i contains individual's age polynomial, child's age polynomial, gender dummy, and a wave dummy. In this model, the parameter δ represents the effect of Catalan linguistic skills on native Spanish parents' decision on how intensively to use Catalan with their children. This is the key parameter measuring the impact of skills on the intergenerational transmission of language: the prediction to be tested is $\delta > 0$.

Our theoretical framework clearly highlights the direction of causality between parental skills and intergenerational transmission. However, a causal interpretation of the OLS estimate of equation (1) requires orthogonality between the linguistic skills (C_i) and unobservables that may affect the choice of the language spoken, summarized in the error term ε_i. This is clearly hard to claim. For instance, Spanish natives with a taste for

Catalan culture will invest in their own linguistic skills and will also be more likely to use Catalan with their children. As long as preferences are unobservable, the simple OLS estimate of (5) will deliver biased and inconsistent parameter estimates.[17]

We address these concerns by exploiting the plausibly exogenous variation in linguistic skills induced by the LNA reform of 1983. As documented by Caminal and Di Paolo (2017) and Cappellari and Di Paolo (2018), cohorts educated in both Catalan and Spanish are more proficient in Catalan than older cohorts educated in Spanish only, and this effect crucially depends on language background, since native Catalan speakers were already highly proficient in both languages even before the implementation of the reform. However, using reform-induced variation across birth cohorts of native Spanish speakers only will not be sufficient to generate a valid IV because that variation would be confounded with secular trends in the language spoken with children. A solution is to also include native Catalan speakers in the analysis and to assume that any secular trend in intergenerational language use is common between the two groups. This framework can be exploited in a difference-in-differences (DiD) setup to generate an instrumental variable for Catalan proficiency, where the compliers are Spanish native speakers in "treated" cohorts.

This approach is in the spirit of the identification strategy proposed by Bleakley and Chin (2004, 2008, and 2010). They estimate the assimilation effects of English proficiency among US immigrants, exploiting the well-established fact of the existence of a "critical period" of language acquisition (*i.e.*, immigrants who arrive in the host country at a very young age assimilate the language more easily). Their identifying variable is the interaction between age at arrival and a dummy that takes the value one if

---

[17] Moreover, the self-reported nature of the language fluency variable could be another source of bias if individuals tends to over-report their language skills. This potential measurement error is likely to bias the estimates obtained by OLS downward.

the immigrant comes from a non-English-speaking country. Under the assumption that the non-language assimilation effects of early migration are the same for immigrants arriving from English-speaking countries as for those from non-English-speaking countries, the differential effect of age at arrival for those who migrated from a non-English-speaking country should be purged of non-language-related effects and thus would represent a valid exclusion restriction.

In our case, we exploit the fact that oral language skills are also acquired within the family at an early age. Hence, the language-in-education reform did not exert any significant effect on the oral proficiency of native speakers. Moreover, the Spanish skills of native Catalan speakers remained very high and stable over cohorts. The underlying assumption of our identification strategy is that both language communities were subject to the same general cohort effects in intergenerational language use. Therefore, specific cohort effects experienced by native Spanish speakers after the policy change should be attributed to their improved skills in Catalan. Under this assumption, the IV estimator identifies the causal effect of oral proficiency in Catalan on language transmission among native Spanish speakers affected by the LNA reform (a Local Average Treatment Effect, LATE).

After including Catalan native speakers in the analysis, equation (1) becomes:

$$L_i = \alpha + \beta' X_i + \delta C_i + \pi S_i + \theta_{\tau(i)} + \varepsilon_i, \qquad (6)$$

which is estimated on the full sample and where we have added a control for native language. Because there is essentially no variation in Catalan skills for Catalans, the associated parameter (δ) estimated from equation (2) is virtually equivalent to its counterpart in equation (1), and also equivalent to that that would be estimated by

22

entering the $S_i C_i$ interaction in equation (2) in place of C_i, as in each case the source of variation used in estimation is the same.

To construct the IV, one option is to separate cohorts on the basis of the year of birth, individuals born after 1970 representing the treated cohorts. However, not all treated cohorts received the same amount of bilingual education during their compulsory schooling, and a binary pre-post indicator would miss this key aspect of the analysis. One reason for differential exposure is the sharp implementation of the reform across all educational levels in 1983, implying that for cohorts already attending school in that academic year the actual amount of exposure depends on the remaining number of years before completion of compulsory education. Secondly, a subsequent reform that took place in 1990 extended compulsory education from 8 to 10 years, translating into longer exposure to bilingual schooling for cohorts born in 1983 and onwards. Therefore, to capture differential treatment intensity within the group of treated cohorts, we construct the exposure indicator as a stepwise function of the birth year:

$$E_{\tau(i)} = \begin{cases} 10 \ if \ \tau(i) \geq 1983 \\ 8 \qquad\quad if \quad 1977 \leq \tau(i) < 1983 \\ \tau - 1969 \quad if \quad 1970 \leq \tau(i) < 1977 \\ 0 \qquad\quad if \quad \tau(i) < 1970 \end{cases} . \tag{7}$$

This is exactly the variable that Clots-Figueras and Masella (2013) used to measure the degree of exposure to the educational reform during compulsory education, and that they used as an exogenous determinant of political identity in Catalonia. In our case, using the exposure indicator, the first stage DiD equation for instrumenting skills in Catalan is:

$$C_i = \mu + \varphi' X_i + \omega S_i + \rho(E_{\tau(i)} S_i) + \sigma_{\tau(i)} + u_i \tag{8}$$

where $\sigma_{\tau(i)}$ is a year of birth fixed effect. Replacing the prediction from the first stage equation ($\hat{C}_i$), the IV estimator for the effect of skills on intergenerational language use can be obtained from the following equation:

$$L_i = \alpha + \beta' X_i + \pi S_i + \delta_{IV} \hat{C}_i + \theta_{\tau(i)} + \varepsilon_i \tag{9}$$

As discussed earlier on in the section, the consistency of δ_{IV} is based on the assumption that non-language cohort effects that could be captured by the exposure variable are common between Spanish and Catalan speakers, and are thus partialled out by the (common) year of birth fixed effects included in the main equation ($\theta_{\tau(i)}$). Since δ_{IV} parameterises a causal relationship, we can now re-frame the test of the theoretical model prediction as $\delta_{IV} > 0$.

6. Results

In Table 2 we report the OLS estimates of the impact of linguistic skills on language transmission choices for native Spanish speakers, corresponding to parameter δ in equation (1).[18] The positive parameter estimate indicates that linguistic skills in Catalan are associated with more intense use of Catalan with children. In particular, in column (1) a unit change in the linguistic skill indicator (C_i, measured on a 0–10 scale) is associated with a 0.281 increase in the intensity of Catalan use with children (L_i, measured on a 1–5 scale). The estimated effect is highly significant. It indicates that one standard deviation increase in oral proficiency in Catalan is associated with an increase

---

[18] Complete results are reported in Table A1 in the Appendix. Results are unaffected if we use models for ordered data in place of OLS.

of 0.46 standard deviation points in the intensity of the use of Catalan with the first child. Alternatively, considering that the average of the Catalan use indicator is 4.8 for native Catalans and 2.6 for native Spanish speakers, the result indicates that a unit increase in linguistic skills reduces the mean difference in Catalan use in the two groups by 13%. Adding controls for parents' and individual's characteristics in columns (2) to (4) does not affect the substance of the evidence.

As discussed in Section 5, the causal parameter of interest in equation (1) is not identified as long as unobserved determinants of linguistic skills also affect the choice of language spoken with children. We can address this issue by means of the proposed IV estimator, which exploits the differential change in linguistic skills between native Spanish and Catalan speakers induced by the 1983 reform. We therefore begin by including also the information on native Catalan speakers in the analysis and estimate equation (2) by OLS using the full sample. Results are presented in Table 3 and show that the point estimate of the impact of linguistic skills is virtually the same as the one estimated in the sample with only native Spanish speakers (see Table 2A in the Appendix for the complete results). This is because there is almost no variation in linguistic skills among native Catalans, and hence the underlying parameter estimates in Tables 2 and 3 are essentially the same.

Using the full sample, we can now move to the IV estimates. The upper panel in Table 4 reports the first stage DiD estimates of the impact of bilingual compulsory education on skills in Catalan.[19] The positive estimate of ρ indicates that for native Spanish speakers 1 year of exposure to bilingual education increases reported oral proficiency by 0.114 points on the 0–10 scale. The predicted proficiency score among Spanish native speakers educated before the reform is 7.45, implying that someone

---

[19] The full results of the first and second stages are reported in the Appendix in Tables A3 and A4, respectively.

starting school after the reform (and therefore experiencing 10 years of exposure to bilingual school) experiences a proportional increase in proficiency of approximately 15% relative to the benchmark of no exposure (= $10 \times 0.114/7.45$). The effect is relevant and also statistically significant; in particular, the F-statistic for the instrument shows that it is a rather strong predictor of proficiency. Reading across columns, we note that the first stage estimate is rather stable after introducing additional controls, which speaks in favor of the exogeneity of the proposed IV.

The lower panel of Table 4 reports the IV estimates for the effect of linguistic skills on intergenerational language use. Under the assumption that cohort effects in language transmission, other than the ones operating through skills, are common across speech communities, the IV estimates capture the causal effect of the increase in oral skills in Catalan of native Spanish speakers (induced by the LNA reform) on the language spoken to the first child. The estimated effect is larger than in the OLS counterpart: for example, in column (1) a unit increase in proficiency increases the intensity of intergenerational Catalan use for native Spanish speakers by 40% of a point on the 1–5 scale, corresponding to an 18% reduction in the average gap of Catalan use between native Spanish and Catalan speakers (= $0.404/2.2$). In the likely case of absence of defiers to the reform, the IV estimate can be interpreted as a LATE, the causal effect of skills for the sub-group of compliers, native Spanish speakers whose proficiency increased because of the reform. The fact that the IV estimate is larger than the OLS counterpart suggests that the intergenerational transmission decision may be more reactive to increases in linguistic skills for compliers compared with the general population.[20]

---

[20] However, the finding of a higher impact of language proficiency on language transmission estimated by IV might be also due to the presence of measurement error in the language skills variable, which tends to bias OLS downward (under classical measurement error).

7. Robustness

7.1 Validity of the instrument

We now subject our findings to a number of robustness checks. We begin in Table 5 by addressing the sensitivity of our results to different functional form specifications of the DiD first stage regression. Specifically, rather than a linear specification (as in the baseline specification, reported in column (1)), we run our IV using either a quadratic function of exposure (column (2)) or dummies for each year of exposure (column (3)). The point estimate of the parameter of interest obtained under alternative functional forms is virtually the same as in the more parsimonious specification based on linear effects in the first stage.[21]

Second, we address the concern that the DiD first stage regression is not capturing a genuine effect of bilingual schooling on skills, but rather a generalized trend in linguistic skills affecting younger cohorts of native Spanish speakers, a circumstance that would invalidate the causal interpretation of the IV estimates. To gain insights into this potential issue, we proceed by estimating the reduced form effect of the language-in-education reform on intergenerational transmission and compare it with analogous estimates obtained from placebo reforms. To mimic placebo reforms, we construct a sample of cohorts that never actually received bilingual education, and pretend that some of them were exposed to bilingualism at school. More specifically, we exclude from our original estimating sample cohorts of individuals treated by the reform (those born from 1970 to 1990) and include older cohorts of non-treated (1949–1945, stopping in 1945 owing to limitations in the availability of information of year of birth).[22] Next,

---

[21] We also checked that the results are stable to different specifications of the functional form of the effect of individual and child's age, as shown in Tables A5 and A6 in the Appendix.

[22] See Table A7 in the Appendix for descriptive statistics referring to the placebo sample.

for the resulting sample of non-treated individuals born in 1945–1969, we assume that the reform was implemented in a placebo year, and re-estimate the reduced form effect for this placebo reform. We conduct the exercise for placebo years 1963 through to 1970. The results presented in Table 6 show that the reduced form effect of the interaction between compulsory exposure and the dummy for being a native Spanish speaker is positive and statistically significant, as expected, with 1 year of exposure to bilingual education raising the probability of transmitting Catalan to children by 4.6 percentage points (p.p.) in the group of native Spanish speakers. Instead, the coefficients of placebo exposure interacted with the dummy for being a native Spanish speaker are generally insignificant and negative, which goes against the idea that our exposure variable is capturing spurious positive trends across cohorts. Thus, the evidence suggests that our first stage regression is plausibly capturing the effect of the policy change, supporting our interpretation of the findings from the IV model.

Having obtained reassuring evidence with respect to the main component of our exclusion restriction (the compulsory exposure variable), we scrutinize its second component, the native language variable. This variable is self-reported and we are concerned about the possibility that the misreporting error could be endogenously driven by the reform. The reason is that some treated native Spanish speakers could define themselves as a native Catalan speaker just because they were exposed to Catalan at school. Therefore, we repeat our estimations using alternative proxies for native language, namely the language of the parents (a dummy that takes the value 1 if both parents have Spanish as their native language) and the origins of the parents (a dummy that takes the value 1 if both parents were born outside Catalonia). The results are reported in Table 7 and show that the point estimate when we use these two alternative proxies for native language (especially regional origins) is somewhat higher than the

baseline, but the overall results remain unchanged.[23] Moreover, in the last column of the same table we also show the results obtained after dropping observations with mixed native language, *i.e.* individuals who claim their native language is Spanish and Catalan. Again, the main coefficients are undistinguishable from the baseline, which again points to the overall robustness of our results.

7.2 Alternative mechanisms

Both our theoretical model and our empirical results emphasize the importance of language skills in changing the incentives to transmit the native language to the children. However, there may be alternative mechanisms that are consistent with our empirical findings.

The first potential mechanism is that the reform may have also changed the language preferences of the compliers. Language preferences are likely to be closely related (although not perfectly correlated) to ethnic or national identity. Thus, the evidence reported in Clots-Figueras and Masella (2013), indicating that the 1983 reform promoted the Catalan identity, provides a motivation to explore such an alternative mechanism. In particular, our findings could be driven by a change in language preferences. Our theoretical setup clearly indicated that a lower relative preference for using the native language (lower value of λ), or a switch in the preferred language, had the same qualitative effects as an improvement in proficiency in the second language. That is, native Spanish speakers who increased their skills in Catalan owing to the exposure to the reform, may have also reduced their preferences towards the use of Spanish with their children. Fortunately, the survey does contain a question about the

---

[23] This is possibly because the new subpopulations of compliers resulting from these alternative variables to define language background are more sensitive to the reform than those captured by the original variable for native language (as confirmed by the higher coefficient of the instrument in the first stage).

respondents' language of self-identification, which is coded in the same way as the native language. We believe that those respondents with a language of self-identification that is different from their native language are the ones that might have experienced a change in preferences. We provide some suggestive evidence about the role of changes in language preferences induced by the reform by re-estimating the model without "language switchers" (individuals whose self-identification language is different from their mother tongue.) The results, reported in the second column of Table 8, are very stable with respect to the baseline in column 1, indicating that changes in language preferences are not the main driver of the results.

The second alternative mechanism we consider is endogenous partnership formation. Indeed, one of the extensions of our benchmark theoretical model suggests that transmission choices may depend on the language characteristics of the partner (mixed versus homogeneous couples). Empirically, the potential relevance of such an alternative mechanism is confirmed by Caminal and Di Paolo (2017), who found that the increase in proficiency in Catalan among native Spanish speakers induced by the reform raised their propensity to find a Catalan-speaking partner and to more intensively use Catalan with the partner. Therefore, it is possible that the effects we found are simply reflecting that those native Spanish speakers who become proficient in Catalan after its introduction as a medium of instruction more intensively use this language with their children simply because they are more likely to be in a linguistically mixed couple. To appreciate the relevance of this alternative mechanism as a plausible interpretation of our empirical results, in the third column of Table 8 we repeat the analysis excluding mixed couples (*i.e.* individuals whose partner has a different mother tongue). This exclusion does not alter the substance of our findings, providing suggestive evidence

that changing assortative patterns induced by the reform are not the main driver of the estimated effects.

8. Concluding remarks

We have argued that policies that change second-language skills may alter the pattern of intergenerational transmission of language. We have proposed a simple theoretical model of an asymmetric bilingual society, in which parents choose the language they use with their children by trading off their language preferences with the economic incentives of their offspring (bilingualism is rewarded in the labor market). As parents improve their proficiency in the second language, the costs of switching to their second language fall and, as a result, they are more inclined to (also) transmit their second language to their children. Empirically, we exploit a natural experiment generated by a policy change implemented in Catalonia in the 1980s. Our results indicate that increased proficiency in Catalan among native Spanish speakers induced by the policy change raised their propensity to transmit Catalan to their children. We also discuss alternative and plausible interpretations of our results. In particular, the change in the pattern of intergenerational transmission could also be generated by a change in language preferences or by a higher propensity to form mixed couples. We provide suggestive evidence that these two alternative mechanisms are not the main drivers of our results.

Better second-language skills improve the chances that the second language is transmitted to the next generation (intergenerational spillover effects of second-language acquisition). Thus, language-related inequalities may persist across generations. We knew that such a channel operates among immigrants in countries with a well-defined dominant language. This paper shows that this channel can also function in bilingual societies, as long as transmitting the second language is also associated with

monetary or non-monetary rewards. It also highlights that language policies that affect the skills of the current generation are likely to have long-lasting effects.

References

Alesina, A, A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003), Fractionalization. Journal of Economic Growth 8, 155-194.

Algan, Y., Mayer, T., and Thoenig, M. (2013). The economic incentives of cultural transmission: Spatial evidence from naming patterns across France. CEPR Discussion Paper 9416.

Bisin, A. and T. Verdier (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97: 298-319.

Bleakly, H. and A. Chin (2004). Language Skills and Earnings: Evidence from childhood immigrants. *Review of Economics and Statistics,* 86: 481-496.

Bleakley, H., and Chin, A. (2008). What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *Journal of Human Resources*, 43(2): 267-298.

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

Casey, T. and C. Dustmann (2008). Intergenerational Transmission of Language Capital and Economic Outcomes. *The Journal of Human Resources*, 43 (3): 660-687.

Caminal, R. and Di Paolo, A. (2017). Your Language or Mine? *Economic Inquiry*, on-line first (DOI: 10.1111/ecin.12542).

Cappellari, L. and Di Paolo, A. (2018). Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform. *Economics of Education Review*, 64: 90-101.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2): 690-731.

Chiswick, B. and P. Miller (2007). "*The Economics of Language: International Analyses*". London: Routledge.

Chiswick, B. and Miller, P. (2017). Do native-born bilinguals in the US earn more? *Review of the Economics of the Household*, forthcoming (https://doi.org/10.1007/s11150-017-9398-5).

Church, J., and I. King (1993). Bilingualism and network externalities. *Canadian Journal of Economics*, 26: 337-345.

Clots-Figueras, I. and P. Masella (2013). Education, Language and Identity. *The Economic Journal*, 123: 332-357.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012), The political economy of linguistic cleavages. Journal of Development Economics 97, 322-338.

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

Egger, P. H., and A. Lassmann (2015). The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. *The Economic Journal*, 125(584): 699-745.

Frankel, J. and A. Rose (2002). An estimate of the effect of currencies on trade and income. *Quarterly Journal of Economics*, 117: 437-466.

Galor, O., O Özak, and A. Sarid (201), Geographical Origins and Economic Consequences of Language Structures. SSRN 2820889.

Gay, V., E. Santacreu-Vasut, and A. Shoham (2013), The Grammatical Origins of Gender Roles, BEHL wp 2013-03.

Kuziemko, I. (2014). Human Capital Spillovers in Families: Do Parents Learn from or Lean on Their Children? *Journal of Labor Economics*, 32(4): 755-786.

Mélitz, J. (2008), Language and Foreign Trade. *European Economic Review,* 52: 667-699.

Rendon, S. (2007). The Catalan premium: language and employment in Catalonia. *Journal of Population Economics*, 20(3): 669-686.

Selten, R., and J. Pool (1991). The distribution of foreign language skills as a game equilibrium, in R. Selten (ed.) *Game Equilibrium Models*, vol. 4, pp. 64-84, Berlin: Springer-Verlag.

Siguan, M. (1991). The Catalan Language in the Educational System of Catalonia. *International Review of Education*, 37 (1): 87-98.

Solon, G. (2014). Theoretical models of inequality transmission across multiple generations. *Research in Social Stratification and Mobility*, *35*: 13-18.

Woolard, K.A., and T.J. Gahng (1990). Changing Language Policies and Attitudes in Autonomous Catalonia. *Language and Society*, 19 (3): 311-330.

Tables
Table 1: descriptive Statistics by Native Language

Variable	pooled sample		native Catalan speakers		native Spanish speakers	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
language spoken with first child: only Catalan	0.610	0.488	0.912	0.283	0.246	0.431
language spoken with first child: more Catalan than Spanish	0.047	0.213	0.028	0.166	0.070	0.256
language spoken with first child: equal Catalan and Spanish	0.088	0.283	0.036	0.187	0.149	0.356
language spoken with first child: more Spanish than Catalan	0.070	0.255	0.010	0.099	0.143	0.350
language spoken with first child: only Spanish	0.185	0.388	0.013	0.113	0.392	0.488
age of the first child	16.747	10.413	17.454	10.500	15.895	10.245
oral skills in Catalan	8.714	2.154	9.578	0.911	7.672	2.692
oral skills in Spanish	9.558	0.943	9.415	1.094	9.729	0.683
native Spanish speaker	0.454	0.498	0.000		1.000	
Catalan (or Catalan and Spanish) as self-identification language	0.696	0.460	0.982	0.131	0.349	0.477
partner Catalan (or Catalan and Spanish) native speaker	0.580	0.494	0.769	0.422	0.351	0.477
both parents Spanish speakers	0.408	0.491	0.033	0.178	0.859	0.348
both parents born outside Catalonia	0.356	0.479	0.043	0.203	0.732	0.443
years of compulsory language exposure	1.660	2.852	1.446	2.702	1.918	3.003
wave 2013	0.554	0.497	0.546	0.498	0.563	0.496
age	45.563	8.495	46.591	8.528	44.325	8.290
male	0.456	0.498	0.470	0.499	0.439	0.496

Table 1 (continued): descriptive Statistics by Native Language

Variable	pooled sample Mean	S.D.	native Catalan speakers Mean	S.D.	native Spanish speakers Mean	S.D.
father's place of birth: Barcelona	0.207	0.405	0.274	0.446	0.126	0.332
father's place of birth: Girona	0.071	0.257	0.126	0.332	0.005	0.073
father's place of birth: Tarragona	0.036	0.186	0.062	0.241	0.005	0.069
father's place of birth: Southern Catalonia (Terres de l'Ebre)	0.059	0.235	0.100	0.300	0.009	0.095
father's place of birth: Western Catalonia (Ponent)	0.071	0.257	0.119	0.324	0.014	0.117
father's place of birth: Central Catalonia	0.058	0.234	0.098	0.297	0.010	0.101
father's place of birth: Pyrenees and Aran Valley	0.034	0.182	0.060	0.238	0.003	0.055
father's place of birth: Balearic Islands and Valencia	0.010	0.099	0.009	0.094	0.011	0.104
father's place of birth: Basque Country and Galicia	0.015	0.122	0.006	0.077	0.026	0.159
father's place of birth: other Spanish regions	0.422	0.494	0.130	0.337	0.773	0.419
father's place of birth: other places	0.011	0.104	0.009	0.094	0.013	0.114
miss father's place of birth	0.006	0.079	0.007	0.083	0.005	0.073
mother's place of birth: Barcelona	0.217	0.412	0.287	0.453	0.132	0.338
mother's place of birth: Girona	0.072	0.259	0.124	0.329	0.010	0.098
mother's place of birth: Tarragona	0.042	0.201	0.071	0.257	0.007	0.085
mother's place of birth: Southern Catalonia (Terres de l'Ebre)	0.059	0.236	0.104	0.305	0.006	0.077
mother's place of birth: Western Catalonia (Ponent)	0.065	0.247	0.110	0.313	0.011	0.104
mother's place of birth: Central Catalonia	0.059	0.236	0.099	0.299	0.011	0.104
mother's place of birth: Pyrenees and Aran Valley	0.032	0.175	0.054	0.227	0.004	0.065
mother's place of birth: Balearic Islands and Valencia	0.019	0.137	0.019	0.138	0.019	0.135
mother's place of birth: Basque Country and Galicia	0.017	0.128	0.008	0.089	0.027	0.162
mother's place of birth: other Spanish regions	0.406	0.491	0.108	0.311	0.765	0.424
mother's place of birth: other places	0.007	0.082	0.008	0.089	0.005	0.073
miss mother's place of birth	0.005	0.070	0.006	0.077	0.004	0.060
(highest) parental education: no education	0.220	0.414	0.144	0.351	0.311	0.463
(highest) parental education: primary	0.524	0.499	0.555	0.497	0.487	0.500
(highest) parental education: secondary	0.168	0.374	0.204	0.403	0.125	0.331
(highest) parental education: tertiary	0.064	0.244	0.074	0.262	0.051	0.219
missing parental education	0.024	0.153	0.022	0.148	0.026	0.159
individual's place of birth: Barcelona	0.485	0.500	0.385	0.487	0.606	0.489
individual's place of birth: Girona	0.089	0.285	0.126	0.332	0.045	0.208
individual's place of birth: Tarragona	0.058	0.234	0.063	0.243	0.053	0.224
individual's place of birth: Southern Catalonia (Terres de l'Ebre)	0.066	0.248	0.114	0.318	0.008	0.088
individual's place of birth: Western Catalonia (Ponent)	0.096	0.295	0.129	0.335	0.057	0.232
individual's place of birth: Central Catalonia	0.085	0.278	0.110	0.313	0.054	0.225
individual's place of birth: Pyrenees and Aran Valley	0.043	0.203	0.067	0.250	0.014	0.119
individual's place of birth: Balearic Islands and Valencia	0.004	0.059	0.001	0.039	0.006	0.077
individual's place of birth: Basque Country and Galicia	0.002	0.049	0.000		0.005	0.073
individual's place of birth: other Spanish regions	0.071	0.258	0.004	0.067	0.152	0.359
individual's place of residence: Barcelona city	0.128	0.335	0.101	0.301	0.162	0.368
individual's place of residence: Barcelona's metropolitan area	0.317	0.466	0.204	0.403	0.454	0.498
individual's place of residence: Girona	0.113	0.317	0.146	0.353	0.073	0.260
individual's place of residence: Tarragona	0.077	0.267	0.080	0.272	0.074	0.262
individual's place of residence: Southern Catalonia (Terres de l'Ebre)	0.073	0.260	0.120	0.325	0.017	0.129
individual's place of residence: Western Catalonia (Ponent)	0.132	0.339	0.143	0.350	0.119	0.324
individual's place of residence: Central Catalonia	0.093	0.291	0.112	0.316	0.070	0.256
individual's place of residence: Pyrenees	0.065	0.247	0.093	0.291	0.031	0.174
completed education: primary or no education	0.293	0.455	0.250	0.433	0.345	0.476
completed education: secondary	0.455	0.498	0.447	0.497	0.464	0.499
completed education: tertiary	0.238	0.426	0.289	0.454	0.175	0.380
completed education: other education levels	0.014	0.119	0.013	0.115	0.016	0.124
number of observations	3,667		2,004		1,663	

Table 2: OLS regressions (outcome = language spoken with the first child), only Spanish speakers

	(1)	(2)	(3)	(4)
oral skills in Catalan	0.281***	0.278***	0.244***	0.242***
	(0.011)	*(0.011)*	*(0.013)*	*(0.013)*
parents' controls	NO	YES	NO	YES
individual controls	NO	NO	YES	YES
adjusted R-squared	0.254	0.260	0.293	0.295
number of observations	1663	1663	1663	1663

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's and child's age and individual's year of birth dummies. Regression in column (2) contains controls for paternal and maternal place of birth (with missing indicators) and highest parental education (with missing indicators). Regression in column (3) includes controls for individual's place of birth, place of residence and completed education. Complete results are reported in Tables A1 in the Appendix.

Table 3: OLS regressions (outcome = language spoken with the first child), pooled sample of Spanish and Catalan speakers

	(1)	(2)	(3)	(4)
native Spanish speaker	-1.647***	-1.413***	-1.468***	-1.334***
	(0.047)	*(0.069)*	*(0.049)*	*(0.068)*
oral skills in Catalan	0.264***	0.258***	0.237***	0.233***
	(0.009)	*(0.009)*	*(0.009)*	*(0.009)*
parents' controls	NO	YES	NO	YES
individual controls	NO	NO	YES	YES
adjusted R-squared	0.567	0.572	0.584	0.587
number of observations	3667	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's and child's age and individual's year of birth dummies. Regression in column (2) contains controls for paternal and maternal place of birth (with missing indicators) and highest parental education (with missing indicators). Regression in column (3) includes controls for individual's place of birth, place of residence and completed education. Complete results are reported in Tables A2 in the Appendix.

Table 4: 2SLS regressions, pooled sample of Spanish and Catalan speakers

	(1)	(2)	(3)	(4)
FIRST STAGE — *outcome: oral skills in Catalan*				
native Spanish speaker	-2.102***	-1.735***	-1.613***	-1.473***
	(0.104)	(0.117)	(0.103)	(0.116)
exposure × native Spanish speaker	0.114***	0.118***	0.109***	0.110***
	(0.025)	(0.026)	(0.024)	(0.025)
adjusted R-squared	0.222	0.235	0.281	0.284
F-test of excluded instruments	21.78	21.41	20.40	19.70
[p-value]	*[0.000]*	*[0.000]*	*[0.000]*	*[0.000]*
2SLS — *outcome = language spoken with the first child*				
native Spanish speaker	-1.379***	-1.244***	-1.270***	-1.179***
	(0.243)	(0.186)	(0.186)	(0.175)
oral skills in Catalan	0.404***	0.369***	0.376***	0.355***
	(0.121)	(0.116)	(0.119)	(0.121)
parents' controls	NO	YES	NO	YES
individual controls	NO	NO	YES	YES
adjusted R-squared	0.540	0.555	0.559	0.568
number of observations	3667	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's and child's age and individual's year of birth dummies. Regression in column (2) contains controls for paternal and maternal place of birth (with missing indicators) and highest parental education (with missing indicators). Regression in column (3) includes controls for individual's place of birth, place of residence and completed education. The F-test on excluded instruments refers to the Angrist-Pischke multivariate F-test on the interactions between years of exposure to Catalan at compulsory schooling and the indicator for Spanish as native language. Complete results of the first-stage regressions are reported in Table A3 in the Appendix.

Table 5: sensitivity to alternative definitions of the exclusion restriction

	(1)	(2)	(3)
FIRST STAGE — *outcome: oral skills in Catalan*			
native Spanish speaker	-2.102***	-2.130***	-2.143***
	(0.104)	*(0.108)*	*(0.111)*
exposure × native Spanish speaker	0.114***	0.232***	
	(0.025)	*(0.085)*	
exposure^2 × native Spanish speaker		-0.016	
		(0.011)	
exposure = 0 × native Spanish speaker			*ref. cat.*
exposure = 1 × native Spanish speaker			0.591***
			(0.111)
exposure =2 × native Spanish speaker			-0.083
			(0.114)
exposure = 3 × native Spanish speaker			1.027***
			(0.115)
exposure = 4 × native Spanish speaker			0.662***
			(0.110)
exposure = 5 × native Spanish speaker			0.709***
			(0.118)
exposure = 6 × native Spanish speaker			1.009***
			(0.118)
exposure = 7 × native Spanish speaker			0.963***
			(0.116)
exposure = 8 × native Spanish speaker			0.676**
			(0.249)
exposure = 10 × native Spanish speaker			1.356***
			(0.449)
adjusted R-squared	0.222	0.223	0.223
F-test of excluded instruments	21.78	13.33	25.60
[p-value]	*[0.000]*	*[0.000]*	*[0.000]*
2SLS — *outcome = language spoken with the first child*			
native Spanish speaker	-1.379***	-1.297***	-1.382***
	(0.243)	*(0.233)*	*(0.209)*
oral proficiency in Catalan	0.404***	0.447***	0.403***
	(0.121)	*(0.116)*	*(0.103)*
adjusted R-squared	0.540	0.520	0.540
number of observations	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's and child's age and individual's year of birth dummies. The F-test on excluded instruments refers to the Angrist-Pischke multivariate F-test on the interactions between years of exposure to Catalan at compulsory schooling and the indicator for Spanish as native language.

Table 6: falsification analysis (Baseline and Placebo Reduced Form Equations)

	real reform in 1983	placebo reform in:							
		1970	1969	1968	1967	1966	1965	1964	1963
native Spanish speaker	-2.228***	-2.298***	-2.287***	-2.266***	-2.243***	-2.218***	-2.187***	-2.168***	-2.152***
	(0.052)	(0.092)	(0.094)	(0.094)	(0.095)	(0.096)	(0.098)	(0.103)	(0.112)
exposure × native Spanish speaker	0.046***								
	(0.014)								
exposure* × native Spanish speaker		0.010							
		(0.014)							
exposure* × native Spanish speaker			0.007						
			(0.014)						
exposure* × native Spanish speaker				0.002					
				(0.014)					
exposure* × native Spanish speaker					-0.003				
					(0.014)				
exposure* × native Spanish speaker						-0.007			
						(0.014)			
exposure* × native Spanish speaker							-0.013		
							(0.015)		
exposure* × native Spanish speaker								-0.015	
								(0.015)	
exposure* × native Spanish speaker									-0.017
									(0.017)
adjusted R-squared	0.472	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
number of observations	3667	3037	3037	3037	3037	3037	3037	3037	3037

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's and child's age and individual's year of birth dummies. The regression in the first column is reduced form equation for the baseline sample. Regressions in columns 2-9 are based on a subsample of never-treated individuals (born between 1945 and 1969, in Catalonia or migrated before age 6); placebo compulsory exposure (exposure*) is imputed "as if" the reform was implemented from 13 to 20 years before the real reform of 1983 (i.e. in 1970 instead of 1983 for column 2, in 1969 for column 3, and so forth).

Table 7: alternative proxies for native language

	(1) native language	(2) parents' language	(3) parents' origins	(4) no mixed language
proxy for native language:				
OLS — *outcome = language spoken with the first child*				
native Spanish speaker	-1.647***	-1.428***	-1.194***	-1.694***
	(0.047)	(0.043)	(0.053)	(0.049)
oral skills in Catalan	0.264***	0.302***	0.335***	0.265***
	(0.009)	(0.009)	(0.009)	(0.009)
adjusted R-squared	0.567	0.524	0.472	0.586
FIRST STAGE — outcome: oral skills in Catalan				
native Spanish speaker	-2.102***	-1.980***	-1.900***	-2.107***
	(0.104)	(0.119)	(0.126)	(0.103)
exposure × native Spanish speaker	0.114***	0.125***	0.166***	0.109***
	(0.025)	(0.028)	(0.028)	(0.025)
F-test of excluded instruments	21.78	19.87	36.04	19.08
[p-value]	*[0.000]*	*[0.000]*	*[0.000]*	*[0.000]*
adjusted R-squared	0.222	0.192	0.166	0.224
2SLS — *outcome = language spoken with the first child*				
native Spanish speaker	-1.379***	-1.182***	-0.949***	-1.416***
	(0.243)	(0.231)	(0.185)	(0.256)
oral skills in Catalan	0.404***	0.441***	0.485***	0.409***
	(0.121)	(0.122)	(0.111)	(0.126)
adjusted R-squared	0.540	0.496	0.439	0.557
number of observations	3667	3622	3662	3515

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for child's age and individual's year of birth dummies. Column (1): native language = 1 if the individual's native language is only Spanish, 0 otherwise. Column (2): parents' language = 1 if both parents are native Spanish speakers, 0 otherwise. Column (3): parents' origins = 1 if both parents were born outside Catalonia, 0 otherwise. Results in column (4) are obtained by excluding individuals with mixed native language (i.e. native language = Catalan AND Spanish).

Table 8: alternative mechanisms (identity changes and partnership formation)

	(1)	(2)	(3)
OLS — _outcome = language spoken with the first child_			
native Spanish speaker	-1.647***	-2.502***	-2.391***
	(0.047)	(0.067)	(0.078)
oral proficiency in Catalan	0.264***	0.141***	0.175***
	(0.009)	(0.010)	(0.011)
adjusted R-squared	0.567	0.743	0.748
FIRST STAGE — _outcome: oral proficiency in Catalan_			
native Spanish speaker	-2.102***	-3.096***	-2.697***
	(0.104)	(0.133)	(0.126)
exposure × native Spanish speaker	0.114***	0.196***	0.141***
	(0.025)	(0.035)	(0.032)
adjusted R-squared	0.222	0.364	0.308
2SLS — _outcome = language spoken with the first child_			
native Spanish speaker	-1.379***	-1.742***	-1.704***
	(0.243)	(0.195)	(0.265)
oral proficiency in Catalan	0.404***	0.418***	0.454***
	(0.121)	(0.070)	(0.108)
adjusted R-squared	0.540	0.649	0.648
number of observations	3667	3039	2549

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's age and individual's year of birth dummies. Results in column (1) are obtained using the entire sample. Results in column (2) are obtained by excluding individuals with self-identification language different from their native language. Results in column (3) are obtained by excluding individuals whose partner's language is different from their native language.

Appendix
Table A1: complete OLS results, only Spanish speakers

	(1)	(2)	(3)	(4)
intercept	-7.049***	-12.766**	-14.759***	-17.958***
	(2.185)	(4.709)	(3.496)	(5.040)
oral skills in Catalan	0.281***	0.278***	0.244***	0.242***
	(0.011)	(0.011)	(0.013)	(0.013)
wave 2013	-0.656***	-1.211***	-1.483***	-1.742***
	(0.217)	(0.335)	(0.279)	(0.388)
age	0.377	0.426	0.407	0.461
	(0.274)	(0.300)	(0.304)	(0.316)
age^2	-0.009	-0.008	-0.006	-0.006
	(0.007)	(0.007)	(0.008)	(0.008)
age^3	0.000*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
male	0.020	0.004	-0.004	-0.022
	(0.086)	(0.085)	(0.083)	(0.085)
age of the first child	0.047	0.042	0.044	0.042
	(0.029)	(0.029)	(0.030)	(0.029)
age of the first child^2	-0.005***	-0.005**	-0.005**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
age of the first child^3	0.000***	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
father's place of birth: Barcelona		*reference category*		
father's place of birth: Girona		0.719		0.088
		(0.433)		(0.403)
father's place of birth: Tarragona		-0.412		-0.554
		(0.668)		(0.607)
father's place of birth: Southern Catalonia (Terres de l'Ebre)		0.202		-0.105
		(0.504)		(0.436)
father's place of birth: Western Catalonia (Ponent)		0.291		0.097
		(0.356)		(0.353)
father's place of birth: Central Catalonia		0.905*		0.497
		(0.448)		(0.426)
father's place of birth: Pyrenees and Aran Valley		0.164		0.043
		(0.534)		(0.575)
father's place of birth: Balearic Islands and Valencia		0.734**		0.679*
		(0.343)		(0.353)
father's place of birth: Basque Country and Galicia		0.136		-0.066
		(0.302)		(0.301)
father's place of birth: other Spanish regions		0.077		0.002
		(0.112)		(0.117)
father's place of birth: other places		0.241		0.089
		(0.327)		(0.333)
miss father's place of birth		-0.463		-0.426
		(0.525)		(0.564)
mother's place of birth: Barcelona		*reference category*		
mother's place of birth: Girona		0.295		-0.265
		(0.479)		(0.414)
mother's place of birth: Tarragona		-0.274		-0.340
		(0.279)		(0.245)
mother's place of birth: Southern Catalonia (Terres de l'Ebre)		0.821		0.014
		(0.538)		(0.518)
mother's place of birth: Western Catalonia (Ponent)		-0.428		-0.756**
		(0.347)		(0.373)
mother's place of birth: Central Catalonia		0.384		0.070
		(0.378)		(0.385)
mother's place of birth: Pyrenees and Aran Valley		0.653		0.531
		(0.695)		(0.659)
mother's place of birth: Balearic Islands and Valencia		-0.041		0.059
		(0.362)		(0.353)
mother's place of birth: Basque Country and Galicia		-0.112		-0.179
		(0.259)		(0.302)
mother's place of birth: other Spanish regions		-0.096		-0.192*
		(0.099)		(0.098)
mother's place of birth: other places		0.466		0.201
		(0.636)		(0.689)
miss mother's place of birth		0.390		0.383
		(0.494)		(0.481)
adjusted R-squared	0.254	0.260	0.293	0.295
number of observations	1663	1663	1663	1663

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A1 (continued): complete OLS results, only Spanish speakers

	(2)	(3)	(4)
(highest) parental education: no education		*reference category*	
(highest) parental education: primary	0.143		0.145
	(0.093)		*(0.086)*
(highest) parental education: secondary	0.150		0.083
	(0.103)		*(0.106)*
(highest) parental education: tertiary	-0.272		-0.350*
	(0.175)		*(0.198)*
missing parental education	0.313**		0.285
	(0.141)		*(0.173)*
individual's place of birth: Barcelona		*reference category*	
individual's place of birth: Girona		0.689***	0.720***
		(0.223)	*(0.205)*
individual's place of birth: Tarragona		0.273	0.337
		(0.210)	*(0.220)*
individual's place of birth: Southern Catalonia (Terres de l'Ebre)		0.536	0.497
		(0.334)	*(0.424)*
individual's place of birth: Western Catalonia (Ponent)		0.227	0.306
		(0.168)	*(0.182)*
individual's place of birth: Central Catalonia		0.419*	0.350
		(0.237)	*(0.228)*
individual's place of birth: Pyrenees and Aran Valley		-0.359	-0.462
		(0.373)	*(0.424)*
individual's place of birth: Balearic Islands and Valencia		-0.203	-0.330
		(0.472)	*(0.505)*
individual's place of birth: Basque Country and Galicia		0.482	0.607
		(0.673)	*(0.721)*
individual's place of birth: other Spanish regions		-0.212*	-0.156
		(0.113)	*(0.124)*
individual's place of residence: Barcelona city		*reference category*	
individual's place of residence: Barcelona's metropolitan area		0.048	0.064
		(0.150)	*(0.152)*
individual's place of residence: Girona		0.262	0.292
		(0.263)	*(0.260)*
individual's place of residence: Tarragona		-0.122	-0.087
		(0.178)	*(0.175)*
individual's place of residence: Southern Catalonia (Terres de		1.022***	1.066***
		(0.236)	*(0.254)*
individual's place of residence: Western Catalonia (Ponent)		0.210	0.267
		(0.174)	*(0.184)*
individual's place of residence: Central Catalonia		0.488**	0.520**
		(0.231)	*(0.230)*
individual's place of residence: Pyrenees		1.035***	1.060***
		(0.264)	*(0.278)*
completed education: primary or no education		*reference category*	
completed education: secondary		0.173**	0.196**
		(0.072)	*(0.075)*
completed education: tertiary		0.285**	0.359**
		(0.119)	*(0.133)*
completed education: other education levels		0.400	0.448
		(0.361)	*(0.391)*
adjusted R-squared	0.573	0.586	0.588
number of observations	1663	1663	1663

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A2: complete OLS results, pooled sample of Spanish and Catalan speakers

	(1)	(2)	(3)	(4)
intercept	-3.246***	-4.100***	-7.284***	-7.489***
	(1.062)	(1.370)	(1.509)	(1.688)
native Spanish speaker	-1.647***	-1.413***	-1.468***	-1.334***
	(0.047)	(0.069)	(0.049)	(0.068)
oral skills in Catalan	0.264***	0.258***	0.237***	0.233***
	(0.009)	(0.009)	(0.009)	(0.009)
wave 2013	-0.451***	-0.477***	-0.847***	-0.842***
	(0.162)	(0.167)	(0.186)	(0.192)
age	0.218	0.248*	0.262	0.272*
	(0.145)	(0.145)	(0.157)	(0.159)
age^2	-0.004	-0.005	-0.004	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)
age^3	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
male	-0.023	-0.021	-0.016	-0.021
	(0.038)	(0.038)	(0.038)	(0.039)
age of the first child	0.018	0.016	0.022	0.021
	(0.019)	(0.019)	(0.018)	(0.018)
age of the first child^2	-0.003**	-0.002**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
age of the first child^3	0.000**	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
father's place of birth: Barcelona		*reference category*		
father's place of birth: Girona		0.026		-0.303***
		(0.089)		(0.097)
father's place of birth: Tarragona		-0.192		-0.249**
		(0.124)		(0.121)
father's place of birth: Southern Catalonia (Terres de l'Ebre)		0.071		-0.190*
		(0.083)		(0.110)
father's place of birth: Western Catalonia (Ponent)		0.154**		0.030
		(0.070)		(0.075)
father's place of birth: Central Catalonia		0.146		-0.078
		(0.102)		(0.112)
father's place of birth: Pyrenees and Aran Valley		0.015		-0.144
		(0.114)		(0.128)
father's place of birth: Balearic Islands and Valencia		0.210		0.135
		(0.157)		(0.179)
father's place of birth: Basque Country and Galicia		-0.211		-0.323
		(0.210)		(0.200)
father's place of birth: other Spanish regions		-0.138*		-0.189**
		(0.076)		(0.078)
father's place of birth: other places		-0.101		-0.201
		(0.200)		(0.188)
miss father's place of birth		-0.148		-0.225
		(0.216)		(0.229)
mother's place of birth: Barcelona		*reference category*		
mother's place of birth: Girona		0.186**		-0.127
		(0.079)		(0.079)
mother's place of birth: Tarragona		0.202*		0.121
		(0.115)		(0.124)
mother's place of birth: Southern Catalonia (Terres de l'Ebre)		0.228**		-0.119
		(0.085)		(0.109)
mother's place of birth: Western Catalonia (Ponent)		0.045		-0.139
		(0.081)		(0.107)
mother's place of birth: Central Catalonia		0.126		-0.111
		(0.095)		(0.108)
mother's place of birth: Pyrenees and Aran Valley		0.155		-0.016
		(0.128)		(0.132)
mother's place of birth: Balearic Islands and Valencia		0.036		0.009
		(0.149)		(0.146)
mother's place of birth: Basque Country and Galicia		0.014		-0.057
		(0.164)		(0.186)
mother's place of birth: other Spanish regions		-0.043		-0.127**
		(0.057)		(0.058)
mother's place of birth: other places		0.210		0.098
		(0.272)		(0.284)
miss mother's place of birth		0.101		0.008
		(0.245)		(0.246)
adjusted R-squared	0.567	0.572	0.584	0.587
number of observations	3667	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A2 (continued): complete OLS results, pooled sample of Spanish and Catalan speakers

	(3)	(4)	(5)
(highest) parental education: no education	*reference category*		
(highest) parental education: primary	0.117**		0.108**
	(0.056)		(0.050)
(highest) parental education: secondary	0.134**		0.101*
	(0.051)		(0.051)
(highest) parental education: tertiary	-0.078		-0.111
	(0.078)		(0.083)
missing parental education	0.125		0.127
	(0.076)		(0.080)
individual's place of birth: Barcelona	*reference category*		
individual's place of birth: Girona		0.257**	0.471***
		(0.116)	(0.130)
individual's place of birth: Tarragona		0.265**	0.287**
		(0.102)	(0.117)
individual's place of birth: Southern Catalonia (Terres de l'Ebre)		0.168*	0.307*
		(0.098)	(0.164)
individual's place of birth: Western Catalonia (Ponent)		0.281***	0.322***
		(0.082)	(0.101)
individual's place of birth: Central Catalonia		0.153	0.237*
		(0.114)	(0.127)
individual's place of birth: Pyrenees and Aran Valley		-0.171	-0.114
		(0.111)	(0.145)
individual's place of birth: Balearic Islands and Valencia		-0.219	-0.263
		(0.362)	(0.384)
individual's place of birth: Basque Country and Galicia		0.399	0.501
		(0.643)	(0.683)
individual's place of birth: other Spanish regions		-0.264**	-0.177
		(0.107)	(0.116)
individual's place of residence: Barcelona city	*reference category*		
individual's place of residence: Barcelona's metropolitan area		-0.009	0.010
		(0.096)	(0.097)
individual's place of residence: Girona		0.231*	0.279**
		(0.134)	(0.137)
individual's place of residence: Tarragona		-0.111	-0.072
		(0.106)	(0.105)
individual's place of residence: Southern Catalonia (Terres de l'Ebre)		0.374***	0.434***
		(0.104)	(0.096)
individual's place of residence: Western Catalonia (Ponent)		0.035	0.058
		(0.104)	(0.110)
individual's place of residence: Central Catalonia		0.319**	0.339**
		(0.129)	(0.133)
individual's place of residence: Pyrenees		0.529***	0.542***
		(0.116)	(0.119)
completed education: primary or no education	*reference category*		
completed education: secondary		0.140***	0.127***
		(0.041)	(0.042)
completed education: tertiary		0.225***	0.229***
		(0.056)	(0.059)
completed education: other education levels		0.285	0.300
		(0.184)	(0.190)
adjusted R-squared	0.572	0.584	0.587
number of observations	3648	3648	3648

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A3: complete First-Stage results

	(1)	(2)	(3)	(4)
intercept	31.936***	32.551***	19.445***	21.329***
	(2.618)	(2.658)	(2.954)	(2.750)
native Spanish speaker	-2.102***	-1.735***	-1.613***	-1.473***
	(0.104)	(0.117)	(0.103)	(0.116)
exposure × native Spanish speaker	0.114***	0.118***	0.109***	0.110***
	(0.025)	(0.026)	(0.024)	(0.025)
wave 2013	3.115***	3.185***	1.916***	2.083***
	(0.413)	(0.435)	(0.393)	(0.398)
age	0.248	0.224	0.351	0.317
	(0.351)	(0.349)	(0.343)	(0.341)
age^2	-0.018*	-0.018*	-0.015	-0.015
	(0.010)	(0.010)	(0.009)	(0.009)
age^3	0.000*	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
male	-0.266***	-0.262***	-0.181***	-0.200***
	(0.058)	(0.057)	(0.058)	(0.058)
age of the first child	-0.020	-0.017	0.007	0.007
	(0.031)	(0.030)	(0.028)	(0.028)
age of the first child^2	-0.001	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
age of the first child^3	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
father's place of birth: Barcelona		*reference category*		
father's place of birth: Girona		0.261*		-0.030
		(0.146)		(0.170)
father's place of birth: Tarragona		0.019		-0.132
		(0.175)		(0.192)
father's place of birth: Southern Catalonia (Terres de l'Ebre)		0.158		-0.227
		(0.141)		(0.168)
father's place of birth: Western Catalonia (Ponent)		0.275**		-0.016
		(0.119)		(0.125)
father's place of birth: Central Catalonia		0.317**		-0.110
		(0.124)		(0.153)
father's place of birth: Pyrenees and Aran Valley		0.392**		-0.051
		(0.175)		(0.220)
father's place of birth: Balearic Islands and Valencia		-0.359		-0.465
		(0.423)		(0.432)
father's place of birth: Basque Country and Galicia		0.038		-0.162
		(0.390)		(0.399)
father's place of birth: other Spanish regions		-0.032		-0.102
		(0.106)		(0.104)
father's place of birth: other places		-0.065		-0.103
		(0.306)		(0.311)
miss father's place of birth		-0.202		-0.294
		(0.505)		(0.443)
mother's place of birth: Barcelona		*reference category*		
mother's place of birth: Girona		0.065		-0.330**
		(0.133)		(0.163)
mother's place of birth: Tarragona		0.322**		-0.053
		(0.145)		(0.199)
mother's place of birth: Southern Catalonia (Terres de l'Ebre)		0.239*		-0.291*
		(0.140)		(0.156)
mother's place of birth: Western Catalonia (Ponent)		0.060		-0.339***
		(0.115)		(0.115)
mother's place of birth: Central Catalonia		-0.048		-0.440***
		(0.124)		(0.142)
mother's place of birth: Pyrenees and Aran Valley		-0.131		-0.505**
		(0.200)		(0.239)
mother's place of birth: Balearic Islands and Valencia		0.030		0.044
		(0.232)		(0.212)
mother's place of birth: Basque Country and Galicia		-0.101		-0.236
		(0.289)		(0.254)
mother's place of birth: other Spanish regions		-0.101		-0.220*
		(0.137)		(0.126)
mother's place of birth: other places		-0.175		-0.406
		(0.339)		(0.340)
miss mother's place of birth		-0.111		-0.325
		(0.620)		(0.618)
adjusted R-squared	3667	3667	3667	3667
number of observations	0.222	0.235	0.281	0.284

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A3 (continued): complete First-Stage results

	(2)	(3)	(4)
(highest) parental education: no education		*reference category*	
(highest) parental education: primary	0.461***		0.315***
	(0.102)		*(0.097)*
(highest) parental education: secondary	0.727***		0.405***
	(0.115)		*(0.112)*
(highest) parental education: tertiary	0.897***		0.482***
	(0.135)		*(0.135)*
missing parental education	0.432*		0.385
	(0.253)		*(0.256)*
individual's place of birth: Barcelona		*reference category*	
individual's place of birth: Girona		0.298*	0.469**
		(0.150)	*(0.190)*
individual's place of birth: Tarragona		0.744***	0.810***
		(0.187)	*(0.212)*
individual's place of birth: Southern Catalonia (Terres de l'Ebre)		0.349	0.618***
		(0.208)	*(0.215)*
individual's place of birth: Western Catalonia (Ponent)		0.351**	0.537***
		(0.172)	*(0.192)*
individual's place of birth: Central Catalonia		0.308*	0.572***
		(0.170)	*(0.183)*
individual's place of birth: Pyrenees and Aran Valley		0.181	0.504*
		(0.176)	*(0.278)*
individual's place of birth: Balearic Islands and Valencia		-0.887	-0.840
		(0.650)	*(0.642)*
individual's place of birth: Basque Country and Galicia		-0.017	0.106
		(0.973)	*(0.842)*
individual's place of birth: other Spanish regions		-0.658***	-0.525**
		(0.195)	*(0.205)*
individual's place of residence: Barcelona city		*reference category*	
individual's place of residence: Barcelona's metropolitan area		-0.264***	-0.202*
		(0.095)	*(0.101)*
individual's place of residence: Girona		0.159	0.234
		(0.190)	*(0.193)*
individual's place of residence: Tarragona		-0.377*	-0.289
		(0.192)	*(0.195)*
individual's place of residence: Southern Catalonia (Terres de		0.231	0.391
		(0.202)	*(0.234)*
individual's place of residence: Western Catalonia (Ponent)		0.097	0.156
		(0.149)	*(0.156)*
individual's place of residence: Central Catalonia		0.247	0.335*
		(0.171)	*(0.169)*
individual's place of residence: Pyrenees		0.304**	0.356**
		(0.144)	*(0.154)*
completed education: primary or no education		*reference category*	
completed education: secondary		0.710***	0.648***
		(0.111)	*(0.105)*
completed education: tertiary		1.170***	1.060***
		(0.105)	*(0.103)*
completed education: other education levels		0.766**	0.687*
		(0.345)	*(0.348)*
adjusted R-squared	0.235	0.281	0.284
number of observations	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A4: complete 2SLS results, pooled sample of Spanish and Catalan speakers

	(1)	(2)	(3)	(4)
intercept	-3.341	-3.548	-4.023*	-4.091*
	(2.701)	(2.562)	(2.416)	(2.466)
native Spanish speaker	-1.379***	-1.244***	-1.270***	-1.179***
	(0.243)	(0.186)	(0.186)	(0.175)
oral skills in Catalan	0.404***	0.369***	0.376***	0.355***
	(0.121)	(0.116)	(0.119)	(0.121)
wave 2013	-0.980**	-0.908*	-1.200***	-1.174***
	(0.469)	(0.471)	(0.347)	(0.393)
age	0.222	0.253*	0.249*	0.264*
	(0.146)	(0.144)	(0.149)	(0.151)
age^2	-0.002	-0.003	-0.002	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)
age^3	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
male	0.015	0.009	0.010	0.004
	(0.047)	(0.046)	(0.041)	(0.043)
age of the first child	0.018	0.016	0.019	0.018
	(0.019)	(0.018)	(0.019)	(0.019)
age of the first child^2	-0.002*	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
age of the first child^3	0.000*	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
father's place of birth: Barcelona		*reference category*		
father's place of birth: Girona		-0.004		-0.300***
		(0.096)		(0.098)
father's place of birth: Tarragona		-0.194		-0.232**
		(0.123)		(0.118)
father's place of birth: Southern Catalonia (Terres de l'Ebre)		0.058		-0.160
		(0.082)		(0.113)
father's place of birth: Western Catalonia (Ponent)		0.126**		0.035
		(0.064)		(0.073)
father's place of birth: Central Catalonia		0.110		-0.066
		(0.109)		(0.106)
father's place of birth: Pyrenees and Aran Valley		-0.029		-0.138
		(0.113)		(0.124)
father's place of birth: Balearic Islands and Valencia		0.255*		0.197
		(0.155)		(0.173)
father's place of birth: Basque Country and Galicia		-0.213		-0.302
		(0.205)		(0.198)
father's place of birth: other Spanish regions		-0.131*		-0.175**
		(0.072)		(0.074)
father's place of birth: other places		-0.091		-0.185
		(0.195)		(0.187)
miss father's place of birth		-0.121		-0.185
		(0.231)		(0.243)
mother's place of birth: Barcelona		*reference category*		
mother's place of birth: Girona		0.181**		-0.085
		(0.082)		(0.085)
mother's place of birth: Tarragona		0.166		0.128
		(0.117)		(0.123)
mother's place of birth: Southern Catalonia (Terres de l'Ebre)		0.200**		-0.088
		(0.094)		(0.104)
mother's place of birth: Western Catalonia (Ponent)		0.035		-0.099
		(0.080)		(0.094)
mother's place of birth: Central Catalonia		0.132		-0.057
		(0.093)		(0.105)
mother's place of birth: Pyrenees and Aran Valley		0.168		0.044
		(0.123)		(0.128)
mother's place of birth: Balearic Islands and Valencia		0.029		0.000
		(0.150)		(0.146)
mother's place of birth: Basque Country and Galicia		0.028		-0.027
		(0.153)		(0.167)
mother's place of birth: other Spanish regions		-0.030		-0.100*
		(0.055)		(0.060)
mother's place of birth: other places		0.232		0.149
		(0.256)		(0.270)
miss mother's place of birth		0.115		0.050
		(0.239)		(0.238)
adjusted R-squared	0.540	0.555	0.559	0.568
number of observations	3667	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A4 (continued): complete 2SLS results, pooled sample of Spanish and Catalan speakers

	(2)	(3)	(4)
(highest) parental education: no education		*reference category*	
(highest) parental education: primary	0.117**		0.108**
	(0.056)		*(0.050)*
(highest) parental education: secondary	0.065		0.069
	(0.077)		*(0.063)*
(highest) parental education: tertiary	0.056		0.055
	(0.095)		*(0.067)*
missing parental education	-0.172		-0.166*
	(0.124)		*(0.097)*
individual's place of birth: Barcelona		*reference category*	
individual's place of birth: Girona		0.216*	0.414***
		(0.117)	*(0.137)*
individual's place of birth: Tarragona		0.161	0.188
		(0.149)	*(0.165)*
individual's place of birth: Southern Catalonia (Terres de l'Ebre)		0.120	0.235
		(0.108)	*(0.167)*
individual's place of birth: Western Catalonia (Ponent)		0.232**	0.256**
		(0.095)	*(0.109)*
individual's place of birth: Central Catalonia		0.109	0.167
		(0.119)	*(0.128)*
individual's place of birth: Pyrenees and Aran Valley		-0.196*	-0.173
		(0.109)	*(0.152)*
individual's place of birth: Balearic Islands and Valencia		-0.094	-0.159
		(0.338)	*(0.356)*
individual's place of birth: Basque Country and Galicia		0.408	0.492
		(0.686)	*(0.697)*
individual's place of birth: other Spanish regions		-0.158	-0.101
		(0.144)	*(0.141)*
individual's place of residence: Barcelona city		*reference category*	
individual's place of residence: Barcelona's metropolitan area		0.026	0.033
		(0.099)	*(0.098)*
individual's place of residence: Girona		0.208	0.249*
		(0.133)	*(0.133)*
individual's place of residence: Tarragona		-0.060	-0.038
		(0.122)	*(0.114)*
individual's place of residence: Southern Catalonia (Terres de		0.342***	0.388***
		(0.099)	*(0.094)*
individual's place of residence: Western Catalonia (Ponent)		0.018	0.036
		(0.107)	*(0.111)*
individual's place of residence: Central Catalonia		0.285**	0.299**
		(0.124)	*(0.128)*
individual's place of residence: Pyrenees		0.482***	0.496***
		(0.108)	*(0.109)*
completed education: primary or no education		*reference category*	
completed education: secondary		0.042	0.049
		(0.091)	*(0.087)*
completed education: tertiary		0.066	0.102
		(0.143)	*(0.135)*
completed education: other education levels		0.181	0.218
		(0.190)	*(0.190)*
adjusted R-squared	0.555	0.559	0.568
number of observations	3667	3667	3667

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include individual's year of birth dummies.

Table A5: sensitivity to age polynomials' order (individuals)

OLS — outcome = language spoken with the first child				
native Spanish speaker	-1.648***	-1.647***	-1.649***	-1.649***
	(0.047)	(0.047)	(0.047)	(0.047)
oral skills in Catalan	0.263***	0.264***	0.264***	0.264***
	(0.009)	(0.009)	(0.009)	(0.009)
adjusted R-squared	0.567	0.567	0.567	0.567
BIC	10888.9	10891.9	10889.1	10873.7
FIRST STAGE — outcome: oral skills in Catalan				
native Spanish speaker	-2.106***	-2.102***	-2.106***	-2.106***
	(0.104)	(0.104)	(0.104)	(0.104)
exposure × native Spanish speaker	0.118***	0.114***	0.118***	0.118***
	(0.025)	(0.025)	(0.025)	(0.025)
adjusted R-squared	0.224	0.222	0.224	0.224
BIC	15155.1	15136.8	15155.1	15155.1
2SLS — outcome = language spoken with the first child				
native Spanish speaker	-1.356***	-1.379***	-1.390***	-1.341***
	(0.236)	(0.243)	(0.251)	(0.244)
oral skills in Catalan	0.416***	0.404***	0.399***	0.425***
	(0.118)	(0.121)	(0.125)	(0.121)
adjusted R-squared	0.535	0.540	0.542	0.530
BIC	11159.1	11116.7	11093.6	11168.7
individual's age polynomial order	4	3	2	1

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for child's age and individual's year of birth dummies. Number of observations = 3667.
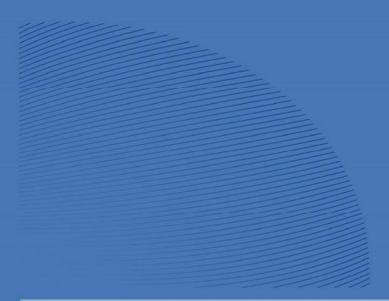
Table A6: sensitivity to age polynomials' order (child)

OLS — outcome = language spoken with the first child				
native Spanish speaker	-1.650***	-1.647***	-1.643***	-1.644***
	(0.048)	*(0.047)*	*(0.047)*	*(0.047)*
oral skills in Catalan	0.263***	0.264***	0.264***	0.264***
	(0.009)	*(0.009)*	*(0.009)*	*(0.009)*
adjusted R-squared	0.568	0.568	0.567	0.567
BIC	10813.80	10808.20	10807.30	10800.70
FIRST STAGE — outcome: oral skills in Catalan				
native Spanish speaker	-2.103***	-2.102***	-2.102***	-2.105***
	(0.105)	*(0.104)*	*(0.105)*	*(0.105)*
exposure × native Spanish speaker	0.114***	0.114***	0.115***	0.117***
	(0.024)	*(0.025)*	*(0.025)*	*(0.025)*
adjusted R-squared	0.222	0.222	0.223	0.223
BIC	15152.897	15136.764	15128.855	15129.298
2SLS — outcome = language spoken with the first child				
native Spanish speaker	5.166***	5.170***	5.263***	5.355***
	(0.872)	*(0.864)*	*(0.851)*	*(0.833)*
oral skills in Catalan	0.403***	0.404***	0.417***	0.424***
	(0.123)	*(0.121)*	*(0.119)*	*(0.115)*
adjusted R-squared	0.540	0.540	0.534	0.531
BIC	11111.3	11116.7	11153.3	11172.0
child's age polynomial order	4	3	2	1

Note: standard errors clustered by year of birth within parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. All regressions include dummies for wave and gender, cubic polynomials for individual's age and individual's year of birth dummies. Number of observations = 3667.

Table A7: descriptive Statistics by Native Language for the placebo sample of never-treated individuals (born between 1945 and 1969, in Catalonia or migrated before age 6)

Variable	(pooled) sample of never treated individuals (1945-1969) Mean	S.D.	native Catalan speakers, never treated sample (1945-1969) Mean	S.D.	native Spanish speakers, never treated sample (1945-1969) Mean	S.D.
language spoken with first child: only Catalan	0.186	0.472	0.0142	0.1182	0.4496	0.4977
language spoken with first child: more Catalan than Spanish	0.048	0.184	0.0087	0.0930	0.1074	0.3098
language spoken with first child: equal Catalan and Spanish	0.072	0.259	0.0272	0.1628	0.1407	0.3479
language spoken with first child: more Spanish than Catalan	0.035	0.213	0.0234	0.1513	0.0525	0.2230
language spoken with first child: only Spanish	0.659	0.384	0.9265	0.2611	0.2498	0.4331
age of the first child	24.06	10.22	24.919	10.163	22.754	10.168
oral skills in Catalan	8.697	2.208	9.5316	0.9741	7.4205	2.8616
oral skills in Spanish	9.519	1.019	9.3956	1.1465	9.7075	0.7455
native Spanish speaker	0.395	0.489	0.0000		1.0000	
Catalan (or Catalan and Spanish) as self-identification language	0.746	0.435	0.9836	0.1271	0.3809	0.4858
partner Catalan (or Catalan and Spanish) native speaker	0.635	0.481	0.8083	0.3938	0.3700	0.4830
both parents Spanish speakers	0.347	0.476	0.0281	0.1652	0.8361	0.3703
both parents born outside Catalonia	0.327	0.469	0.0449	0.2071	0.7560	0.4296
years of compulsory language exposure	0.000		0.0000		0.0000	
wave 2013	0.519	0.500	0.5158	0.4999	0.5237	0.4996
age	52.33	7.67	53.320	7.719	50.829	7.351
male	0.463	0.499	0.4635	0.4988	0.4621	0.4988
number of observations	3,037		1,836		1,201	