Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2025/09 1/51 pág.
Working Paper 2025/09 1/51 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2025/04 1/51 pág.
Working Paper 2025/04 1/51 pag.

Bilingual Education and Identity

Ramon Caminal, Antonio Di Paolo and Ada Ferrer-i-Carbonell

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública
UNIVERSITAT DE BARCELONA

WEBSITE: www.ub.edu/irea/ • CONTACT: irea@ub.edu

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

IREA Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. For that reason, IREA Working Papers may not be reproduced or distributed without the written consent of the author. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of IREA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

Abstract

We present new evidence on the impact of a reform that introduced Catalan-Spanish bilingual education in Catalonia on identity formation. Specifically, we revisit the findings of Clots-Figueras and Masella (2013, The Economic Journal) by examining how exposure to Catalan as medium of instruction affects identity and political preferences. To do so, we use more recent data from repeated cross-sections and multiple alternative sources. Furthermore, we explore an overlooked dimension of identity: self-identification language. At the aggregate level, we find a small but negative effect of bilingual education on the likelihood of identifying as exclusively Catalan. Our results are robust to a battery of sensitivity checks and falsification tests. However, they differ significantly from those of Clots-Figueras and Masella. Our replication of their results reveals a lack of robustness, primarily due to their definitions of identity, as well as to other aspects of their model specification. Our analysis of heterogeneous effects shows that the small negative impact of the reform on identifying as "only Catalan" is entirely driven by individuals from non-Catalan backgrounds, whether in terms of native language or parental origins. For this group, exposure to bilingual education also reduces the likelihood of adopting Catalan as the language of self-identification and support for the independence of Catalonia. These findings suggest that the language-in-education reform might have triggered a backlash effect.

Keywords: language-in-education reform, bilingual education, identity, language, political preferences.

JEL Classification: I28, J15, Z13.

Authors:

Ramon Caminal, Institute of Economic Analysis, CSIC, and BSE

Antonio Di Paolo, Universitat de Barcelona & AQR-IREA

Ada Ferrer-i-Carbonell, Institute of Economic Analysis, CSIC, and BSE

Acknowledgements: We acknowledge financial support from the Institut d'Estudi de l'Autogovern through the grant PRE166/23/000004, the Ministry of Science and Innovation (grants PID2021-124256OB-I00, PID2021-122575NB-I00 and PID2020-118653GB-I00), Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S), and AGAUR-Generalitat de Catalunya through grants 2021SGR00416 and 2021SGR00415. The authors have no conflicts of interest to disclose. We acknowledge comments received during presentations at EALE, EEA-ESEM, AIEL, LEER workshop on Education Economics, LESE Conference. The usual disclaimers apply.

1. Introduction

This paper aims to investigate the impact of bilingual education on identity, with a specific focus on a reform that introduced Catalan as medium of instruction in Catalonia—a topic previously explored by Clots-Figueras and Masella (2013). We build on their work by using a more recent and comprehensive data from repeated cross-sections and multiple alternative sources. Additionally, we expand the analysis by employing alternative definitions of identity and considering different sets of controls and sample restrictions.

A large body of research has documented the relevance of group identity in a variety of economic and social environments (Akerlof and Kranton, 2010, Shayo, 2020, Charness and Chen, 2020). Consequently, understanding how individuals get to identify with various social groups is of utmost importance. From a historical perspective, an essential feature of the construction of nation-states in the XIX and XX centuries was the homogenization of a diverse population; in particular, the adoption of a unified national identity. This was achieved primarily through the introduction of state-controlled mass education, the implementation of a unique "national language", compulsory military service, and the building of infrastructure (See Alesina et al., 2021, and the historical references they cite[1]). Therefore, mass education and language have been identified as crucial determinants of a unified national sentiment.

The comparison between France and Spain can serve as an example. In the XIXth century, the Spanish government failed to allocate sufficient funds towards mass education due to divergent interests between the elite of the country and the regional elites (Hauk and Ortega, 2019). The primary enrollment rate in France in the 1880s was already 75%, whereas it was only 42% in Spain (Soysal and Strang, 1989).[2] Such limited investment efforts are probably behind the current relative vitality of some minority languages (Galizian, Euskera, and Catalan). In contrast, French soon became the dominant language, pushing the rest to the brink of extinction.

The implementation of a sole national language has sometimes been contested. The best documented case is the ban of German in US schools occurred after World War I. Before 1917, bilingual education was prevalent in areas of the Midwestern United States with a large fraction of German immigrants. After 1917, some states banned German as a

[1] The incentives for investing in nation-building were both political: responding to internal and external threats (Alesina et al., 2020), and economic: the creation of large, unified domestic markets (Alesina et al., 2000).

[2] Recent research has provided systematic evidence on the role of education and language on nation-building. See, in particular, Bandiera et al. (2018) for the US, You (2018) for China, and Blanc and Kubo (2021) for France.

language of instruction.[3] Fouka (2020) showed how such a policy shift, rather than resulting in faster assimilation, "contributed to a cultural backlash and greater isolation of the German community from mainstream American culture". The increased influence of school on identity was compensated for by the extra effort the parents made in transmitting their ethnic identity to their children.[4]

In the present context, the Catalan case is particularly interesting from a research standpoint for at least two reasons. First, despite the difference in strength between the two languages, the weaker one (Catalan) possesses sufficient significance to influence numerous aspects of social and political life. Among non-official languages at the country level,[5] Catalan is one of the most widely spoken in Europe.[6] This opens the possibility of a polarized conflict around language policies.

Second, in the early 1980s the regional government implemented a large-scale language-in-education reform following the approval of the Language Normalization Act, LNA (*Llei de Normalització Lingüística)* of 1983. This reform came shortly after the end of Franco's regime, during which the use of Catalan was prohibited. The policy had two main objectives: 1) to promote the social use and prestige of Catalan and 2) to ensure full bilingualism at the end of compulsory education, regardless of the students' linguistic background. The second goal stemmed from the unequal distribution of language skills in the Catalan society in the early 1980s: while all native Catalan speakers were proficient in Spanish, a large fraction of the native Spanish speakers were either monolingual or passively bilingual.[7]

As a result of the language-in-education reform, the school was transformed from a monolingual system, in which Spanish was the only language, to a bilingual one, in which Catalan became the main language of instruction. Since exposure to the reform through compulsory education levels only depends on the birth year, such a policy shift can be viewed as a quasi-experiment, and as such, it can serve as the foundation for a credible identification strategy.

---

[3] These events coincided with the *Americanization campaign, 1914-1924,* which is considered the period of the emergence of the American national identity (Ricento, 2003).

[4] Similarly, Cinnirella and Schueler (2018) showed that increased centralized education spending in Prussia (1886-1911, as part of a Germanization policy) led to higher overall support for pro-nationalist parties, but backfired among the Polish minority, who strengthened their Polish identity. On the resistance to education by culturally marginalized groups see also Carvalho et al. (2024).

[5] We are ignoring the microstate of Andorra, where Catalan is indeed official.

[6] Joshua Fishman, one of the world's foremost sociolinguists, chose Catalan as one of the remarkable success stories of reversed language shift (Fishman, 1991).

[7] The relative size of linguistic groups is very relevant to understand language policy choices. In the case of Indonesia in the 1980's, Bazzi et al (2019) showed that fractionalization favors the adoption of the (neutral) national language, whereas the presence of a small number of large language groups generates the opposite effect, as ethnic attachment intensifies. See also Laitin and Ramachandrian (2023).

Thus, one of the expected effects of the reform was the expansion of language skills, especially for individuals with a Spanish-speaking background (Caminal and Di Paolo, 2019). The economic theories of identity formation (pioneered by Akerlof and Kranton, 2000) suggest that such a reduction in the cost of acquiring the Catalan language, one of the markers of the Catalan identity, is likely to stimulate the adoption of such an identity. Thus, the main research question is whether bilingual education can be an effective instrument of "nation-building", or at least, whether it could contribute to redefining the borders of linguistic/ethnic groups.[8] Naturally, this is an empirical question.

Clots-Figueras and Masella (2013), CF-M hereafter, investigated the effect of the same Catalan language-in-education reform on identity.[9] They used data from a single cross-sectional survey of the *Centro de Investigaciones Sociologicas* (CIS), conducted in 2001. Their variable of interest was the number of (potential) years of exposure to bilingual education during compulsory schooling, which exclusively depends on the birth year. Their analysis seems to indicate that the language-in-education reform fostered the adoption of the Catalan identity, as measured by the response to a question regarding the feeling of belonging. Furthermore, they argued that exposure to bilingual education had an impact on voting behavior, boosting support for parties favoring greater decentralization. Some authors (See, for instance, Ginsburgh and Weber, 2020) cite this paper as evidence that the origins of the current conflict over Catalan secession can be traced back to the educational reform.[10]

In this paper, we reevaluate the impact of the language-in-education reform on Catalan identity and political preferences, with the aim of providing more compelling results. We use the measure of compulsory language exposure proposed by CF-M. Thus, we rely on Intention-to-Treat effects throughout the paper. However, our approach combines several important innovations. First, as main data source, we exploit repeated cross-section data covering the period comprised between 2014 and 2020. We mostly rely on the Barometer of Political Opinion (BPO), a survey carried out by the Center of Opinion Studies (COS). The use of multiple waves covering several years allows us to disentangle the age and cohort effects. By using repeated cross-sections, we can observe individuals born in the same year but with different ages when responding to the survey. On the contrary, CF-M employed a single cross-section, implying that birth cohort variation

[8] Research on the formation of "regional" identity is scarce. An important exception is Dehdari and Gehring (2022). They show how negative experiences with nation-states, like war, occupation, and repression, affected the formation of a specific identity in Alsace-Lorraine.

[9] See also Aspachs et al. (2008), who examined the impact of language exposure throughout schooling on identity formation in Catalonia and the Basque Country. The key difference between these cases is that, in Catalonia, all public and state-subsidized private schools were subject to the policy change. In contrast, the Basque Country maintained both a (Spanish) monolingual system and a Basque-Spanish bilingual system after the reform (also implemented in 1983), allowing parents to choose the type of school for their children.

[10] Clearly, alternative explanations are also plausible. In particular, as discussed later in the background section, the ruling of the Spanish Constitutional Court of 2010 (See Casas et al, 2024) might help explaining the sharp increase in the support for Catalan secession.

(used to capture the effect of the reform) coincides with age variation. In their baseline specification, they include a fourth-order age polynomial together with the exposure variable, which implies that, by using a single cross-section, identification is based on a strongly non-linear functional form of the age effect. Moreover, we also exploit alternative data sources to validate the main findings.[11]

Second, we explore alternative definitions of the main outcome variable(s). CF-M used a self-reported measure of the feeling of belonging, defined on a 1 to 5 scale: (1) Only Spanish, (2) More Spanish than Catalan, (3) as Spanish as Catalan, (4) More Catalan than Spanish, and (5) Only Catalan. They consider two main identity outcomes in their analysis. In the first (which they called *Identity*), they use the full ordinal scale as dependent variable. In the second (which they call it *IdentityD*), they create a binary variable, coded as 1 if the respondent selected (3), (4) or (5), and zero otherwise. Unfortunately, *IdentityD* exhibits very little variability: 95% of the responses in our sample, and 87% in CF-M's sample, are coded as 1. Moreover, grouping the indicators of the Catalan identity ((4) and (5)), with (3), the neutral identity option, blurs the interpretation of the results. In particular, an increase in I*dentityD* (and similarly for the *Identity* variable) could result from a shift from a strong Spanish identity to the neutral category, which may reflect a reduction in polarization, and hence a lower likelihood of conflict.

In this paper, we use the same ordinal scale of the question about the feeling of belonging, but adopt more transparent definitions of Catalan identity: (a) *National Identity*, which takes the value 1 if the response is (5), and zero otherwise, and (b) *Weak National Identity*, which takes the value 1 if the response is (4) or (5), and zero otherwise. In addition, similarly to CF-M, we examine the impact of the reform on an alternative outcome, using individuals' self-reported support for Catalan independence.

Third, we pay close attention to linguistic variables. In particular, we account for the respondents' native language, alongside parental geographical origins, as a crucial feature of the family background, whereas CF-M only consider the latter variable. Indeed, the language transmitted within the family is closely related to the intergenerational transmission of national identity and political preferences. In addition, we examine the presence of heterogeneous effects of the language reform on identity formation, not only in terms of the geographic parental origin (as done by CF-M), but also in terms of the respondents' native language. Finally, we use the language of self-identification as an alternative outcome. Even though languages are closely associated with ethnicity and

[11] Specifically, the Survey of the Language Use of the Population (SLUP), conducted by the Catalan Institute of Statistics, focusing on linguistic issues; as well as the Survey of the Political Environment (SPE) and the Omnibus, both carried out by COS. We also use the same CIS data employed in CF-M to replicate their findings. However, no other dataset provided by the CIS enables similar analysis due to the unavailability of some key variable (i.e. place of birth and year of arrival to Catalonia for individuals born elsewhere).

political identities, their connection is not always straightforward. Indeed, indicators of linguistic and national identity, as well as political preferences, exhibit a positive, albeit imperfect correlation. In the context of our case study, a reform centered around the languages of instruction will likely have a greater impact on linguistic identity than on other identity dimensions. Therefore, by studying the impact of exposure to bilingual education on linguistic identity, we offer new evidence on this previously unexplored aspect of identity formation.

Our findings differ markedly from those presented by CF-M. In our preferred specification for estimating the aggregate effect of the reform, the coefficient for the exposure variable is very close to 0 for all outcomes. The only statistically significant result is a negative coefficient for *National Identity* (-0.5 percentage points), which is very modest in size. Specifically, ten years of exposure to the reform —corresponding to the entire duration of compulsory education — reduces the likelihood of identifying as "Only Catalan" by 5 percentage points. Given that 38% of individuals unaffected by the reform report feeling "Only Catalan", this suggests that full exposure to bilingual education at school reduces that likelihood by approximately 13%. In contrast, the estimates for the linguistic identity (adopting Catalan as the language of self-identification) or support for Catalonia's independence are virtually 0 and statistically insignificant. The finding of null, and possibly negative (though small) aggregate effects, is consistent across multiple data sources and alternative outcomes. This result also holds up under several robustness checks and falsification tests. In contrast, the influence of the family background on identity appears very strong and robust.

Our results may not come as a big surprise. After all, the conditions under which such a reform took place were quite different from those existing during the construction of the Western nation-states. First, Catalonia operates within the Spanish political and judicial system. In fact, the Spanish Constitutional Court validated the Catalan government's language-in-education policies in 1994. Second, in the 1980s about half of the Catalan population were native Spanish speakers, meaning they could hardly be characterized as a minority. Third, the education reform received wide popular and political support. Notably, as discussed in the next section, the 1983 LNL was passed unanimously in the Catalan Parliament.

We place particular emphasis on replicating the results presented by CF-M and on assessing the stability of their findings. In our replication exercise, we check for the stability and consistency of the results across several dimensions (definitions of the outcome variables, estimation sample, and model specification). The most relevant finding is that the estimated coefficient of the exposure variable is markedly lower and loses statistical significance by using as outcome our definition of national identity (i.e. feeling "only Catalan"). Moreover, the same coefficient is virtually 0 across all outcomes — even those examined by CF-M — when simply changing the age polynomial from fourth to second order. Overall, the findings from our data and the replication of CF-M's

study strongly challenge their main conclusion. Contrary to promoting the Catalan identity, the introduction of Catalan as medium of instruction in compulsory education (driven by the language reform) appears to have no sizeable aggregate effect on identity formation. If anything, the reform seems to have discouraged the adoption of a Catalan identity. Such a discrepancy may stem from their attempt to capture the effect of compulsory language exposure using only a single cross-sectional dataset, and a misleading definition of national identity with little variation.

Finally, using our preferred specification and main dataset, we also explore the potential for heterogeneous effects. This is particularly important because the small aggregate effects might conceal opposing effects on different population groups. Specifically, we estimate how the consequences of exposure to bilingualism varies by gender, the respondent's native language, and their parents' place of birth. While the latter two variables are positively correlated with each other, they are not perfectly so. A substantial portion of respondents are native Spanish speakers with Catalan origins, whereas the group of native Catalan speakers with non-Catalan origins is negligible.

The results indicate that the small negative effect of exposure to the reform on the likelihood of identifying as "only Catalan" is entirely driven by individuals with a non-Catalan background. Specifically, this effect is zero for native Catalan speakers and for those with both parents born in Catalonia. However, each additional year of exposure to bilingual education in school reduces the likelihood of a strong attachment to Catalan identity among native Spanish speakers (-0.9 percentage points), individuals who report both Catalan and Spanish as native languages (-1 percentage points), and those with one or both parents born outside Catalonia (-1.3 and -0.7 percentage points, respectively). These coefficients reveal sizeable effects of exposure to the reform. In particular, given that 15% of the native Spanish speakers unaffected by the reform feel "only Catalan", this suggests that full exposure to bilingual education reduces that likelihood by 9 percentage points, 60% of the initial level. Similarly, the propensity to adopt Catalan as self-identification language is negatively affected by the language-in-education reform for native Spanish speakers and for individuals with one parent born outside Catalonia. However, we detect a positive coefficient for native Catalan speakers. Finally, the reform has reduced the propensity of native Spanish speakers and individuals with one parent born outside Catalonia to support independence.

Thus, the improvement in Catalan proficiency among native Spanish speakers resulting from language exposure at school induced by the reform does not appear to be associated with a higher willingness to adopt the Catalan identity, contrary to the predictions of the economic theories of identity formation.[12] Moreover, our results on heterogeneous effects

[12] At various moments of the reform different researchers have documented a meaningful change in the pupils' attitudes towards languages (Woolard and Ghang, 1990; and Newman and Trenchs-Parera, 2015): Catalan was much less identified as the language of a particular group, and increasingly perceived as "the

suggest that, possibly in response to the policy shift, an alternative mechanism was initiated to counteract the influence of the school and preserve the Spanish identity among native Spanish speakers and individuals with non-Catalan parental origins. This mechanism could have resulted in a backlash like that reported by Fouka (2020) and Cinnirella and Schueler (2018).

The rest of the paper is organized as follows. In section 2 we present the institutional background. Section 3 contains the description of the data and presents some descriptive statistics. In section 4 we present the identification strategy of our estimation, and in section 5 we describe all the results, including our robustness checks and falsification analysis, as well as the replication of the findings by CF-M and the analysis of heterogeneous effects.

2. Institutional Background

Catalonia is a bilingual Spanish region where Catalan and Spanish (Castilian), which are both Romance languages, have coexisted for several centuries. Historically, their relative roles in society have been very sensitive to political and institutional changes. At the beginning of the XVIII century, explicit policies that aimed at the homogenization of the Spanish territories were implemented. As a result, the Catalan language was gradually excluded from most public spheres. In the mid-XIX century, the Spanish Parliament passed an ambitious law (Moyano Law, 1857) whose goal was the creation of a compulsory, public school system with Spanish as the only language. In practice, the enrollment ratio stayed low for a long time. As a result, the popular use of Catalan in Catalonia remained high.

In the second half of the XIX century, a cultural and linguistic movement emerged (called *La Renaixença*), whose goal was the promotion of the use of Catalan in literature and, more generally, in any form of cultural expression. Soon, the movement adopted a more political attitude by supporting the use of Catalan in education and other social activities. The first manifestation of modern, Catalan-oriented political activity was the creation, at the beginning of the XX century, of *La Lliga Regionalista*, a right-wing political party, with very modest goals in terms of decentralization, but that represented a clear expression of the Catalan identity, mixed with conservative values.

During the Second Spanish Republic (1931-1936) a new party (*Esquerra Republicana de Catalunya*, ERC) took over the representation of the Catalan identity, this time in combination with republican and progressive values. ERC obtained a landslide victory in the 1932 elections to the Catalan parliament, and took over the regional government in

language of everyone". Our results seem consistent with theirs. Native Spanish speakers can learn Catalan without renouncing to their inherited identity.

coalition with small socialist parties. The Catalan language became official in the region and was also introduced as language of instruction in the public school system. The regional government was suspended in 1934, and only restored a few months before the civil war (1936-1939).

Franco's victory implied a major setback for the Catalan language. Under his dictatorial regime (1939-1975), Spanish was the only official language and the only language used in education. Any infringement was heavily punished. Thus, Catalan was banned from public life and its use restricted to the private sphere. Nevertheless, the Catalan language was transmitted (mostly orally) across generations by most families of Catalan origin.

A strong sense of Catalan identity survived throughout these years, sustained mostly by civic groups as well as clandestine political parties and labor unions. For instance, the main publication of the largest political party of the anti-Franco movement (*Partit Socialista Unificat de Catalunya*, the Catalan communist party), called *Treball*, was written entirely in Catalan. Similarly, the main underground labor union in Catalonia was named *Comissió Obrera Nacional de Catalunya*, implicitly claiming Catalonia as a nation.

The use of Spanish in Catalonia was strongly reinforced by massive migration flows from other Spanish regions, especially in the 1950s and 1960s. At the beginning of the 1980s Catalan was the native language of about one half of the population, who were also fully competent in Spanish. In contrast, most of the native Spanish speakers (40% of the population were born outside the region) were monolingual or only passively bilingual (Siguan, 1991).

The Spanish Constitution of 1978 designed a two-tier government, in which the central government shares some of the political power with regional governments. Right after the constitution of the Catalan government, the regional parliament passed in 1983 the *Llei de Normalització Lingüística* (Language Normalization Act, LNA). The aim of the LNA was to make all pupils fully competent in both languages at the end of compulsory education. It also defined an integrative education model, in which children were not separated on the basis of the language spoken at home. Hence, the educational system underwent a gradual transition from a solely Spanish-based system to one where Catalan emerged as the primary language of instruction.[13] The LNA was modified in 1998 (*Llei de Política Lingüística,* Law of Language Policy, LLP). None of the main goals of the LNA were affected, and some of the mechanisms reinforced.

On top of the language-in-education reform, the new legislation also sought to promote the knowledge and use of Catalan using a variety of means, including a Catalan-only TV

[13] At least in compulsory education. Spanish was still dominant in post-compulsory secondary education. See Cappellari and Di Paolo (2018) for more details about the introduction of Catalan at school with the LNA reform.

channel (*Televisió de Catalunya*, TV3) and a radio station, several campaigns, and language proficiency requirements for public sector jobs.

The LNA received the widest support in the Catalan Parliament (one abstention, zero negative votes). For the first two decades, the regional government (and hence the implementation of the law) was in the hands of a Catalan-oriented coalition (*Convergència i Unió*), whose electoral support never rose significantly above 50%. After fifteen years of practical application of the LNA, the *Llei de Política Lingüística* (Language Policy Law, LLP) was approved in the Catalan Parliament in 1998 by a wide majority.[14]

For the first two decades, the linguistic policies of the regional government enjoyed a broad social support in Catalonia (in sharp contrast with the reactions in the rest of Spain), and was only challenged by small, very active groups with little impact on local politics. In the XXI century, after the creation of a new political party, *Ciutadans, Partido de la Ciudadania*, language policy became much more controversial inside Catalonia, including its parliament.

Dissatisfaction with the financing of the regional government, and concern about the fragile legal status of language policies, led a bunch of political parties (including Catalan-oriented parties and the Catalan branch of the Spanish Socialist Party) to write a new charter for the region (*Estatut d'Autònomia*). The new draft was approved by the Catalan Parliament in 2005, but was significantly modified by the Spanish Parliament. Nevertheless, it was ratified in a referendum in 2006. In a decision that was unprecedented, the Spanish Constitutional Court reviewed the approved (and sanctioned in a referendum) draft and, in June 2010, declared a number of articles unconstitutional and reinterpreted several others. As a result, the level of decentralization established in the new charter was severely curtailed. In July 2010, about one million people demonstrated in Barcelona against such a decision. From 2012 to 2017, massive demonstrations took place every year on September 11th (the Catalan national day). Support for independence increased throughout this period to almost 50% from relatively low historical levels, 15-20% (*Centre d'Estudis d'Opinió*). Casas et al. (2024) showed that the ruling of the constitutional court was a turning point for the independence movement: it increased support for independence by 5 percentage points and a 10% decrease in trust in courts and satisfaction with democracy.

In October of 2017, the regional government held a referendum on the independence of Catalonia, which was declared illegal by Spanish authorities. As a result, the main leaders

[14] Only two small parties voted against the 1998 law, and they did it for opposite reasons. One of them (PP, a Spanish-oriented party, with 17 seats) found the law discriminatory against the Spanish language, and the other (ERC, a Catalan-oriented party, 13 seats) believed it was too shy to rescue the Catalan language from its extinction path. The total number of seats in the Catalan Parliament was 135. Thus, the law was approved by more than two thirds.

of the independence movement and members of the Catalan government were either imprisoned or exiled, and the regional government's powers were temporarily suspended.

3. Data and Descriptive Statistics

The main dataset for the empirical analysis comes from the Barometer of Political Opinion (BPO) conducted by the Center of Opinion Studies (COS) of the regional Government. We also test the reliability of our findings by using several alternative data sources.[15]

The BPO is a periodic survey that began in 2004. It uses a multistage stratified sampling procedure to select random samples from the population of Catalonia aged 18 and older, all of whom hold Spanish nationality. The primary sampling units are municipalities, which are randomly selected based on province and city size. The secondary sampling units are census tracts within the selected municipalities. Within these tracts, individuals are chosen using cross quotas based on gender, age group, and place of birth, in accordance with official data from the Population Census (*Padró Municipal d'Habitants*).

For this analysis, we use data from the 2014-2020 waves, which were collected using the Computer Assisted Personal Interview (CAPI) method. Earlier waves were excluded for two main reasons: first, from 2004 to 2013, the survey employed a different method (Computer Assisted Telephone Interview, or CATI); second, prior to 2012, the survey did not collect data on native and self-identification languages, which are critical to our analysis. Multiple waves were conducted within certain years: two waves in 2014, and three waves between 2015 and 2020. Depending on the specific wave, each typically includes between 1,500 and 2,000 observations.

In the BPO survey there are two questions that reflect the respondents' national and linguistic identities. Proxies for national identity can be constructed from the responses to the question about the feeling of belonging that, as mentioned in the introduction, are coded in an ordinal scale defined over five options. The definition of our preferred indicator of identity is based on a self-reported measure of the feeling of belonging. As described in the introduction, our main outcome is an indicator that takes the value 1 if the individual feels only Catalan, and 0 otherwise — which we label as *National Identity*. We also consider as alternative outcome a dummy that is 1 if the individual feels more Catalan than Spanish or only Catalan (*Weak National Identity*) is provided in the introduction. These definitions differ from those used in CF-M: the entire 1-5 scale treated

[15] The additional datasets are described in detail in Appendix A, together with basic descriptive statistics.

as continuous variable (*Identity*) and an indicator that includes neutral identity feelings (*IdentityD*), which is unlikely to capture national identity.

We also consider a question about the language of self-identification to measure linguistic identity. In particular, respondents must choose one of the four options: "Catalan", "Spanish", "Catalan and Spanish" or "others".[16] We primarily focus on the variable *Linguistic Identity*, that takes value 1 if the respondent chooses "Catalan", and zero otherwise. Once again, we emphasize a distinctive indicator of the Catalan identity. In Appendix B we also discuss, as a robustness check, the variable *Weak Linguistic Identity*, that takes value 1 if the respondent selects "Catalan" or "Catalan and Spanish", and zero otherwise.

The survey also contains a question about the political organization of the state. In particular, it asks whether Catalonia should be: (i) a "Spanish Region", (ii) an "Autonomous Community" (the current situation), (iii) a "Federal State within Spain", and (iv) an "Independent State". Our leading indicator of political preferences is *Independence*, a variable that takes the value 1 if the respondent selects (iv), and zero otherwise. As a robustness check, in Appendix B we also consider the variable *Sovereignty*, that takes the value 1 if (iii) or (iv) are picked, and zero otherwise. This variable can be regarded as a weaker version of *Independence*, yet still reflecting a clear endorsement for a higher degree of sovereignty for Catalonia.

Summarizing, the paper focuses on three key outcomes: *National Identity*, *Linguistic Identity*, and *Independence*. The first two variables aim at capturing different aspects of the respondents' identity, while the third reflects their political stance on a well-defined, contentious issue. As a robustness check, we also examine *Weak National Identity*, *Weak Linguistic Identity*, and *Sovereignty*. Additionally, to better understand the differences between our results and those of CF-M, we also incorporate their two definitions of national identity: *Identity* and *IdentityD*.

In addition to the outcome variables, the survey includes several sociodemographic characteristics that serve as control variables. These include sex, age, native language (first language spoken at home during the childhood, classified in the same way as self-identification language), place of birth (Spanish regions for those born in the rest of Spain and country for foreign born individuals), place of birth of both parents, current county

[16] More specifically, the "others" category includes different combination of languages, but also the "Aranès" language, a variety of Occitan, which is spoken in the Aran Valley, a territory of Catalonia located in the North of the Pyrenees, and more accessible from Occitania than from Catalonia. Due to the very limited number of individuals that satisfy our sample restrictions and report other combination of languages or "Aranès" as self-identification languages, we discarded these observations for the empirical analysis.

of residence[17] and municipality size (categorized by population), educational attainment, cohabitation status, and the number of children. Additionally, the survey asks about the media sources (TV, newspapers, and radio) through which individuals access political news, variables that are further explored in the empirical analysis.

The main estimation sample consists of individuals born in Catalonia, as well as those born in other Spanish regions who migrated at the age of six or younger. Thus, we retain only Spaniards who completed their entire education in Catalonia, either before or after the introduction of the language-in-education reform. In addition, we exclude individuals born before 1950 to avoid cohorts with a relatively low survival rate and individuals who lived a significant portion of their lives under Franco's dictatorship, which could influence their willingness to openly discuss sensitive issues related to national identity. We also exclude those born after 1990 as these may still be students at the time of the survey.

With these criteria, the birth cohorts included in the empirical analysis are mostly centered around the key year of the LNA's reform as far as compulsory education is concerned. More specifically, the first individuals partially exposed to the reform were born in 1970, so we retain individuals born within a 20 years window before and after this reform. Finally, we remove 194 observations of individuals who meet the above criteria, but whose native language is neither Catalan nor Spanish, as well as observations lacking data on linguistic identity and parents' place of birth. After applying these restrictions, the final pooled sample consists of 18,081 observations.[18]

For the falsification analysis, we use two different placebo samples. First, we consider individuals born between 1940 and 1969. In this case, we retain all individuals who were schooled in Catalonia, but never exposed to bilingual education during compulsory schooling. Second, we consider individuals born between 1950 and 1990 outside Catalonia, either in other regions of Spain or abroad, whose mother tongue is either Spanish or Catalan, and who migrated to Catalonia at age 16 or older. Thus, individuals in this second placebo sample belong to the same birth cohorts as those in our baseline sample, but they were too old when migrated to Catalonia to have been exposed to the language-in-education reform during compulsory schooling.

[17] Based on the information about the county of residence, we impute the share of residents in the county who can speak and write in Catalan, obtained from the Census of 2011.

[18] There are, however, few additional observations that are discarded in some regressions due to missing values in the outcomes (feeling of belonging and preferences about the political organization of the state). Descriptive statistics for these variables and the corresponding regressions are obtained only using the available observations.

Table 1: Descriptive Statistics

Variable	Mean	Std. dev.
Outcome Variables		
Feeling of belonging*		
only Spanish	0.021	0.144
more Spanish than Catalan	0.029	0.167
equal Spanish than Catalan	0.324	0.468
more Catalan than Spanish	0.268	0.443
only Catalan	0.359	0.480
Self-Identification language**		
Catalan	0.571	0.495
Spanish	0.290	0.454
Catalan & Spanish	0.139	0.346
Political Organization of the State***		
Catalonia should be a Spanish Region	0.039	0.194
Catalonia should be an Autonomous Community	0.205	0.404
Catalonia should be a Federal State	0.240	0.427
Catalonia should be an Independent State	0.516	0.500
Explanatory Variables		
years of bilingual compulsory education	3.858	4.089
male	0.496	0.500
age	46.52	11.25
mother tongue = Catalan	0.493	0.500
mother tongue = Spanish	0.440	0.496
mother tongue = Catalan & Spanish	0.068	0.251
both parents born in Catalonia	0.456	0.498
one parent born in Catalonia	0.226	0.418
both parents born outside Catalonia	0.318	0.466
compulsory education or less	0.274	0.446
post-compulsory secondary education	0.466	0.499
tertiary education	0.259	0.438
living with a partner (married or couple)	0.673	0.469
number of children	1.275	1.059
Catalan TV for political information	0.469	0.499
Catalan radio for political information	0.291	0.454
Catalan newspapers for political information	0.114	0.318
living in the county of Barcelona (*Barcelonès*)	0.225	0.418
living in Barcelona city	0.165	0.371
% of individuals who speak and write in Catalan	0.743	0.058

Number of observations: 18081; *17516 valid observations; ** 18081 valid observations; ** 17070 valid observations. % of individuals who speak and write in Catalan in the county (comarca) of residence, source: 2011 Census.

The descriptive statistics for the main estimation sample are presented in Table 1.[19] Regarding the outcome variables, since most individuals born outside Catalonia have been excluded, the sample is skewed towards respondents with a Catalan identity. Specifically, 36% of individuals in the main sample identify as solely Catalan, 27% feel more Catalan than Spanish and 32% consider themselves equally Spanish and Catalan. Only a tiny fraction identifies as either more Spanish than Catalan or exclusively Spanish. Regarding linguistic identity, 57% of the respondents report Catalan as their language of self-identification, while 14% select both Catalan and Spanish. The remaining 29% identify Spanish as their language of self-identification.

The final dependent variable reflects the preferences regarding the organization of the state. In our sample, 24% of respondents favor a federal state and more than twice as many (52%) prefer an independent state. Only 4% would like Catalonia to be a Spanish region, while 20% support the current arrangement, in which Catalonia is an autonomous community within Spain.

As for the control variables, the estimation sample is quite balanced in terms of gender, with an average age around 46. The proportion of individuals whose native language is Catalan slightly exceeds that of native Spanish speakers (49% and 44%, respectively). Finally, more than half of individuals in the sample have at least one parent born outside Catalonia, while 46% have both parents born in Catalonia.

To visualize the evolution of the main outcomes across birth cohorts, Figure 1 displays binned scatter plots of their averages by year of birth, along with a local linear smoothing plot. Additionally, we account for a discontinuity at 1970, which marks the first cohort exposed to bilingual education during compulsory schooling.

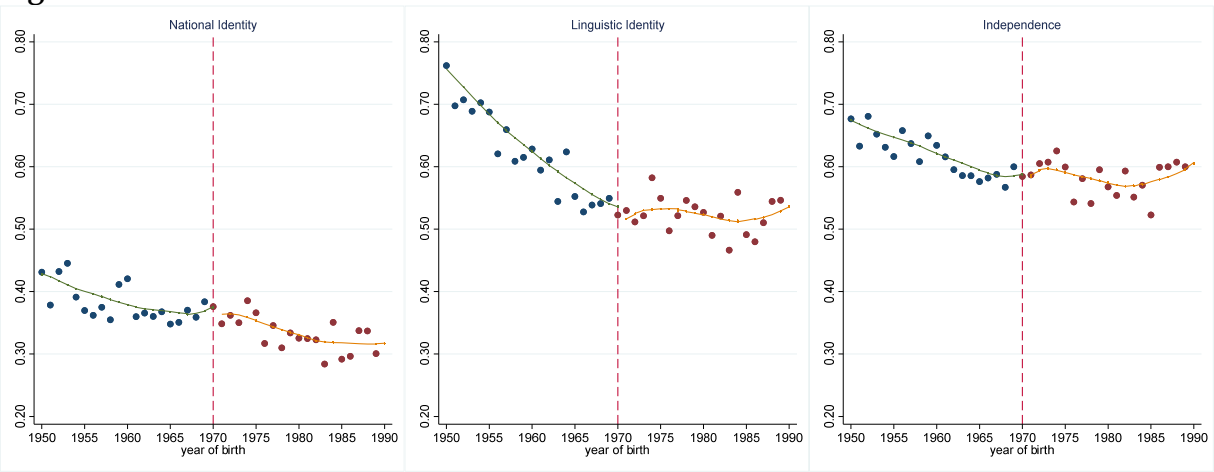
The evolution across the cohorts is not obvious. A noticeable downward trend is evident in all three variables for cohorts not exposed to the language-in-education reform (i.e. those born before the 1970s), with the trend being particularly pronounced for *Linguistic Identity*. A negative trend in *National Identity* is also observed among the earliest cohorts affected by the LNA reform, and appears somewhat stronger than in cohorts just below the cutoff (born in 1970), while the evolution of this variable is more stable in the last cohorts of our dataset. However, the negative trend in *Linguistic Identity* and *Independence* dissipates among those exposed to the LNA reform during compulsory education, with these outcomes fluctuating around a constant value for younger cohorts (1970-1990).

Nevertheless, is it difficult to draw definitive conclusion about the effect of the LNA reform on identity formation based on these plots alone. The variation across cohorts in repeated cross-sections can be influenced by factors such as age differences, variation in respondents' language background, or even spurious cohort trends. For this reason, in

[19] Descriptive statistics for the placebo samples are shown in Tables B1 and B2 of Appendix B.

the next section, we present the empirical strategy we use to address these potential challenges and obtain causal estimates of the impact of the language-in-education reform on identity formation.

Figure 1: evolution of the outcomes across the cohorts



4. Empirical Strategy

To assess the impact of bilingual education on identity formation, and following the approach originally proposed by CF-M, we exploit the variation in exposure to bilingual education across birth cohorts, induced by the implementation of the LNA reform in 1983. To mitigate potential biases arising from endogenous educational choices, we focus exclusively on potential exposure to bilingual education during compulsory schooling.[20]

Primary education in Spain starts at age 6. For individuals born before 1982, the duration of compulsory education, which includes both primary and lower secondary education, was 8 years (i.e. until age 14). This was the framework established by the General Law of Education (LGE) of 1970. Consequently, individuals born before 1970 were never exposed to Catalan during compulsory schooling.

The language-in-education component of the LNA reform was sharply introduced across all grades in primary and lower secondary education in 1983. For cohorts already attending school in the academic year of the reform, the extent of exposure varied depending on how many years remained in their compulsory education. This means that individuals born between 1970 and 1976 received part of compulsory education in Spanish, and part in both Catalan and Spanish, resulting in partial exposure to the new language policy. In contrast, individuals born after 1976 were entirely schooled under the

[20] A variant of this approach has been used by Caminal and Di Paolo (2019) and by Caminal et al. (2021) to estimate the causal effect of oral fluency in Catalan among native Spanish speakers on partnership formation and intergenerational language transmission, respectively.

bilingual system and received 8 years of compulsory education in both Catalan and Spanish.

A subsequent reform in 1990 (Law of General Ordering of the Educational System, LOGSE) extended compulsory education from 8 to 10 years, leading to longer exposure to bilingual schooling for cohorts born in 1983 and onwards. To capture this differential intensity of exposure to the reform, we construct individual i's number of (potential) years of bilingual compulsory education, E, as a stepwise function of the birth year[21] ($\tau(i)$):

$$E_{\tau(i)} = \begin{cases} 10 & if \quad \tau(i) \geq 1983 \\ 8 & if \quad 1977 \leq \tau(i) < 1983 \\ \tau(i) - 1969 & if \quad 1970 \leq \tau(i) < 1977 \\ 0 & if \quad \tau(i) < 1970 \end{cases}. \tag{1}$$

The empirical analysis seeks to estimate the impact of exposure to Catalan during compulsory education ($E_{\tau(i)}$) by estimating the following equation by OLS:

$$Y_i = \alpha + \beta E_{\tau(i)} + f(age_i) + \delta' X_i + \theta_{yw} + \varepsilon_i. \tag{2}$$

where Y_i represents each of the different outcome variables: national and linguistic identity, and support for independence. We include as controls a second-order age polynomial ($f(age_i)$), to control for age-differences in the outcomes[22], a vector of individual-level control variables (X_i), and year-wave fixed effects (θ_{yw}), which account for potential trends in the outcomes over time. The coefficient of interest is β, which captures the impact of each additional year of compulsory exposure to bilingual education.

The set of controls (X_i) initially includes only age and gender. To partial-out the correlation between individual's current identity and their background and to consider possible intergenerational inheritance of identity, we subsequently add mother tongue and parents' place of birth (origin) to the vector of individual characteristics. This

[21] Birth year is imputed based on age, the only variable available in the BPO database and other relevant data sources, including the CIS survey used by CF-M. This method introduces some ambiguity in determining the exact year of birth, as individuals may have been born in the year prior to the one inferred from their age at the time of the survey. Consequently, this could lead to measurement error in the exposure variable (except for individuals born before 1969 or after 1983). To address this issue, we employed an alternative probabilistic imputation method that incorporates the month of the interview to refine the year of birth estimate. We can then recalculate the exposure variable accordingly. Notably, this alternative method produced results that are consistent with those obtained using the initial age-based imputation approach. These additional results are available upon request.

[22] Because we control for wave dummies throughout, the age polynomial captures a combination of age and cohort effects.

represents our preferred specification in terms of plausibly exogenous individual-level controls. However, we also present a third additional specification in which we also control for completed education, cohabitation status, and the number of children.

4.1. Robustness and Falsification Checks

We perform several checks regarding the main variable of interest. The measure of exposure to bilingual education during compulsory schooling, $E_{\tau(i)}$, is not a choice variable, as it depends exclusively on year of birth and on the duration of compulsory schooling. Therefore, it is not driven by unobserved individual characteristics that could affect identity. Nevertheless, there are different issues that could challenge the interpretation of the β coefficient as the causal effect of potential exposure to bilingualism at school on identity formation in an Intention-to-Treat framework.[23]

The first concern is the specification of the relationship between age and the outcome(s). The availability of repeated cross-sections of data implies that the sample contains individuals born in the same year who have different ages when they participated in the survey. On the contrary, the regressions reported in CF-M are based on a single cross-section, implying that the coefficients of the age polynomial are only separately identified from the exposure coefficient based on strong functional form assumptions. However, the choice of the functional form $f(age_i)$ remains somewhat arbitrary. To address this, we assess the sensitivity of the results by using higher order age polynomials, and age dummies, in order to rule out the possibility that the coefficient of interest is contaminated by unaccounted-for age differences.

Related to the previous point, our baseline specification includes year-wave fixed effects to capture time trends in the identity variables. Since age and year of birth are collinear within a given wave, we also present results from alternative specifications of time effects. These include quadratic and linear time trends across the different waves of the BPO survey, as well as a specification using year dummies instead of wave dummies, given the availability of multiple waves per year, and the grouping of years.

The most important threat for the identification strategy is the potential presence of spurious cohort trends in the outcomes. This is because $E_{\tau(i)}$ might capture other cohort-specific confounders that might affect individual's identity through channels other than language exposure during compulsory education. To address this concern, we conduct two types of falsification checks, using placebo reforms for groups of individuals who were never actually exposed to the LNA reform during compulsory education.

[23] Because the exposure variable ($E_{\tau(i)}$) is defined according to birth cohort, we cluster the standard errors at the year of birth level. Moreover, because we use different but interrelated measure of identity, we also verify the robustness of the results by correcting statistical inference for multiple hypothesis testing using the Romano-Wolf approach.

Specifically, our first falsification exercise replicates the approach used by CF-M. We focus on individuals born in Catalonia and those born in the rest of Spain who migrated at age 6 or earlier, similar to our main sample, but in the period from 1940 to 1969. This "older" cohort grew up in Catalonia but was never exposed to the LNA reform during compulsory education. To test for spurious effects, we construct a set of placebo exposure measures by considering hypothetical reforms that could have been implemented before the actual 1983 reform (specifically, between 1970 and 1963). We then estimate equation (2) using this placebo exposure measures rather than the true measure, while focusing on individuals from older cohorts who were never exposed to the reform. If the coefficients from placebo exposure measures were systematically and significantly different from zero, it would indicate the presence of pre-existing spurious cohort trends in the outcomes.

In the second falsification exercise, we analyze an additional sample of individuals born in other Spanish regions or outside Spain who belong to the same birth cohorts as our baseline sample (1950–1990) but migrated to Catalonia at age 16 or later, and thus were never exposed to Catalan during compulsory schooling. For this placebo sample, we impute the exposure measure "as if" they had been schooled in Catalonia and then estimate the regression specified in equation (2). If the coefficient for the placebo exposure measure was significant, it would suggest that the results from the baseline sample are driven by contemporaneous spurious cohort effects, rather than reflecting the causal effect of the LNA reform.

We also provide evidence regarding the potential role of confounders. An important element to consider is the role of the media. As mentioned in the institutional background section, the LNA reform led to the creation of a Catalan-only television channel (TV3) and promoted the establishment of radio channels that exclusively use Catalan as their language of communication. These media outlets represented additional tools to further promote the language. Therefore, one could argue that the exposure variable may not only capture the effect of language exposure in education but may also reflect the growing presence of Catalan-language media. To assess whether media exposure could act as a potential confounder, we exploit the information about the TV channels, newspapers and radio stations that individuals typically use to follow political news. The survey includes information on current media preferences, which we use to create indicators for Catalan-language TV, newspapers and radio. However, we acknowledge that these variables might themselves be influenced by language exposure at school. This means that including them as additional controls in the regression, as done by CF-M for TV consumption, could introduce a 'bad control' problem. Therefore, we take an alternative approach: we regress each of the three Catalan-media use indicators on the exposure variable and the same set of controls used in our preferred specification. We thus check whether the results regarding the exposure coefficient are consistent with the possibility of media acting as a confounder.

Another potential confounding factor is the changing sociolinguistic composition of the place of residence. Catalonia experienced a significant inflow of internal migrants from other Spanish regions during the 1950s and 1960s, many of whom settled in specific areas of the city of Barcelona and surrounding municipalities within the Metropolitan Area. The geographical concentration of these internal migrants remained relatively stable for several decades, although it has decreased in recent years—partially due to the rise of international migration at the beginning of the 21st century (Garcia-López et al., 2021). This means that the exposure variable, which primarily relies on cohort variation, could partially capture trends in sociolinguistic composition at the local level—an aspect that may be linked to identity formation. Ideally, to analyze the relevance of this contextual factor, we would need information on residential history at the very local level. Unfortunately, such data is not available in any of the datasets used for the empirical analysis. However, the BPO dataset does contain information about the county of current residence and city size. Therefore, we construct three variables related to the sociolinguistic composition of the broader area of residence: 1) the share of individuals who can speak and write in Catalan in the county, 2) an indicator for living in the county of Barcelona, and 3) and indicator for residing in the city of Barcelona. Following a similar approach to the one used for analyzing the role of Catalan media, we regress these variables on the exposure measure and our preferred set of controls, and check whether the results suggest that the main findings could be driven by geographical variation in sociolinguistic composition.

An additional challenge of the empirical strategy is the variation in the actual amount of exposure to Catalan at school and changes in teachers' composition. As carefully explained by Cappellari and Di Paolo (2018), the intensity of the use of Catalan as medium of instruction during the early decades of the reform was not uniform across schools. This variation was largely influenced by teachers' proficiency in Catalan and the linguistic background of students. The Catalan Government introduced measures to increase teachers' skills in Catalan and encourage teacher turnover, particularly in the early 1990s, along with language immersion programs designed to increase the presence of Catalan in schools where most students came from Spanish-speaking backgrounds. Unfortunately, we are not able to approximate this variation in exposure to Catalan across schools and over time. Nevertheless, we can observe that exposure to Catalan at school has increased for all cohorts affected by the reform. For those who completed compulsory education before the reform, exposure was virtually nonexistent, especially among individuals educated during the Franco Regime, as the use of Catalan in education was forbidden and anecdotal.

Finally, we further examine the stability of our results by replicating the estimation from our preferred specification using alternative datasets instead of our main source (the BPO survey for the period 2014-2020). Specifically, we re-estimate the main model using the available information on outcomes and control variables from each of the three alternative databases: The Survey of the Political Environment (SPE) and Òmnibus

(OGC), both conducted by the COS, and the Survey of the Language Use of the Population (SLUP), conducted by the Catalan Statistical Institute. The goal is to ensure that the main results are not driven by the particular data source used in the primary analysis.

4.2. Comparison with CF-M

After presenting compelling evidence regarding the robustness of our findings, we compare our results with those reported by CF-M. We proceed as follows. First, we replicate the analysis using the same dataset as CF-M, a single cross-section of data from the CIS survey of 2001. We consider the same outcomes used by CF-M (*IdentityD*, *Identity*), as well as our preferred definitions of identity (*National Identity*) and our weaker definition of identity (*Weak National Identity*). We then estimate the corresponding regressions across six different specifications using: 1) the same sample[24] and baseline specification as CF-M that includes only the exposure variable and a fourth-order age polynomial, 2) the CF-M sample and their full set of control variables, 3) the CF-M sample and controls, but with a second-order age polynomial instead of a fourth-order polynomial, 4) the CF-M sample and our set of control variables, 5) a sample with the same conditions as our main estimates and the CF-M controls, and 6) our sample and control variables, corresponding to our preferred specification.

Second, we use the BPO dataset and the same four identity outcomes as before, estimating the model across the same specifications as in the previous exercise: 1) the CF-M sample and a fourth-order age polynomial, the CF-M sample and control variables, 3) the CF-M sample and covariates, but controlling for a second-order age polynomial, 4) the CF-M sample with our controls, 5) our sample with CF-M controls and 6) our sample and control variables (which corresponds to our preferred specification). Because the BOP survey includes different waves, we control for wave dummies throughout.

The goal of these exercises is to investigate how the dataset, the specification of the age polynomial, definitions of identity, choice of controls, and sample restrictions contribute to explaining the differences between our results and those reported by CF-M.

4.3. Heterogeneous Effects

We conclude the analysis by presenting additional estimates that account for heterogeneous effects of exposure to bilingual education. For each of the three main outcomes, we estimate the regression model to allow for heterogeneous effects of the

[24] The CF-M sample corresponds to individuals born between 1908 and 1983 with valid information for the identity feelings variable, excluding those who migrated to Catalonia being 18 years old or older. Their estimations also include territorial weights for provinces (Barcelona = 1.7844, Girona = 0.452, Lleida = 0.3072, Tarragona = 0.4838). We would like to express our gratitude to the authors for providing us the do-file that replicate their results.

exposure variable, based on three exogenous and predetermined covariates: gender, mother tongue, and parental origins. The aim of the analysis is to determine whether the introduction of Catalan as the language of instruction has a differential effect on our identity measures across different groups.

5. Results

5.1. Baseline results

Selected estimates from equation (2) are presented in Table 2. The results focus on the main outcome variables described earlier: *National Identity*, *Linguistic Identity* and *Independence*. For all regressions, we cluster the standard errors by year of birth, which is the level of variation of the exposure variable.

Table 2: Exposure to Bilingual Education and Identity

dependent variable:	(1) National Identity	(2) Linguistic Identity	(3) Independence
Panel A			
years of bilingual compulsory education	-0.002	0.004	0.001
	(0.002)	(0.003)	(0.002)
adjusted R-squared	0.010	0.036	0.013
Panel B			
years of bilingual compulsory education	-0.005**	-0.001	-0.003
	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.219	0.610	0.222
Panel C			
years of bilingual compulsory education	-0.005**	-0.001	-0.003*
	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.226	0.612	0.229
mean of the dependent variable	0.359	0.571	0.515
number of observations	17516	18081	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Panel A: controls variables = wave dummies, age and age squared, gender. Panel B: control variables = wave dummies, age and age squared, gender, mother tongue, parents' origins. Panel C: control variables = wave dummies, age and age squared, gender, mother tongue, parents' origins, completed education, living with the partner, number of children.

The coefficients of exposure to bilingual education displayed in Panel A are obtained by controlling for wave dummies, gender and a second-order age polynomial. In this

specification, the coefficients for the exposure variable are very small and not statistically significant across any of the three outcomes. To partial-out the linguistic background of the individual, Panel B presents results from our preferred specification, which adds as control variables indicators for native language (only Catalan, Catalan & Spanish and only Spanish) and parental origins (both parents born in Catalonia, one parent born in Catalonia and both parents born outside Catalonia). As for the previous specification, the coefficient for exposure remains virtually zero for linguistic identity and support for independence. However, the estimate for *National Identity* (first column) remains quantitatively small (-0.005, s.e. 0.002), but is slightly higher than for the first specification and turns out to be statistically significant.[25] Specifically, the results suggest that for each additional year of potential exposure to bilingual education during compulsory education, decreases the likelihood of feeling only Catalan by 0.5 percentage points. The proportion of individuals with 0 years of exposure who feel only Catalan is 38%. This implies that, on average, an individual exposed to bilingual education for the full 10 years of compulsory schooling is 5 percentage points less likely to feel only Catalan. Thus, full exposure to bilingualism at school reduces the likelihood of identifying solely as Catalan by 13%. In Panel C we present additional estimates that include additional controls for completed education, cohabiting status, and the number of children. The inclusion of these choice variables does not modify the main conclusions derived from our preferred specification. Finally, Table B4 in Appendix B reports estimates using weaker identity measures as outcomes. None of these estimates are statistically or economically significant.

5.2. Robustness and Falsification Checks

We then assess the robustness of the results by examining the specification of the age polynomial and time/wave effects for each of the three outcomes. As alternative specifications for the age effect, we include third- and fourth-order polynomials, as well as age dummies, which is the most flexible approach for capturing age differences (see Tables 3a, 3b, and 3c, respectively). The results for linguistic identity and support for independence remain unaffected by the functional form of the age effect. For *National Identity*, the coefficient of exposure is slightly higher when using a third-order age polynomial or age dummies, but is less precisely estimated compared to our preferred specification (second-order age polynomial). Regarding time effects, we explore alternative models with varying specifications for time controls. The results, presented in Table B5 of Appendix B alongside the baseline estimates (column 1), show that the coefficient of exposure remains virtually unchanged across these specifications. Specifically, the coefficient is consistent when using year dummies instead of wave

[25] In Table B3 of Appendix B we compare the p-value of the estimates obtained from the baseline specification with the p-value corrected for multiple hypothesis testing with the Romano-Wolf technique. The statistical significance of the estimates is generally unaffected by the existing correlation between the three interrelated outcomes that we consider.

dummies (column 2), dummies for grouped years (column 3), a linear yearly trend (column 4), and a quadratic trend (column 5).

Table 3a: Sensitivity to Age Polynomial - dependent variable: National Identity

	(1)	(2)	(3)	(4)
years of bilingual compulsory education	-0.005**	-0.008***	-0.005*	-0.009**
	(0.002)	(0.003)	(0.003)	(0.004)
age	0.007*	0.031	-0.131*	
	(0.004)	(0.019)	(0.075)	
age2	-0.000**	-0.001	0.005*	
	(0.000)	(0.000)	(0.003)	
age3		0.000	-0.000*	
		(0.000)	(0.000)	
age4			0.000*	
			(0.000)	
age fixed effects	no	no	no	yes
adjusted R-squared	0.219	0.219	0.219	0.219
number of observations	17516	17516	17516	17516

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, gender, mother tongue, parents' origins.

Table 3b: Sensitivity to Age Polynomial - dependent variable: Linguistic Identity

	(1)	(2)	(3)	(4)
years of bilingual compulsory education	-0.001	-0.002	0.000	-0.004
	(0.002)	(0.002)	(0.002)	(0.003)
age	0.001	0.007	-0.088	
	(0.002)	(0.013)	(0.054)	
age2	0.000	-0.000	0.003	
	(0.000)	(0.000)	(0.002)	
age3		0.000	-0.000	
		(0.000)	(0.000)	
age4			0.000	
			(0.000)	
age fixed effects	no	no	no	yes
adjusted R-squared	0.610	0.610	0.610	0.610
number of observations	18081	18081	18081	18081

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, gender, mother tongue, parents' origins.

Table 3c: Sensitivity to Age Polynomial - dependent variable: Independence

	(1)	(2)	(3)	(4)
years of bilingual compulsory education	-0.003	-0.001	0.001	0.002
	(0.002)	(0.003)	(0.003)	(0.003)
age	0.005*	-0.011	-0.130	
	(0.003)	(0.017)	(0.089)	
age2	-0.000*	0.000	0.004	
	(0.000)	(0.000)	(0.003)	
age3		-0.000	-0.000	
		(0.000)	(0.000)	
age4			0.000	
			(0.000)	
age fixed effects	no	no	no	yes
adjusted R-squared	0.222	0.222	0.222	0.221
number of observations	17070	17070	17070	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, gender, mother tongue, parents' origins.

To discard the possibility that the estimates of exposure obtained from our preferred specification are driven by spurious cohort trends, we show the results from two falsification exercises discussed earlier. In Tables 4a, 4b, and 4c, we present the coefficients of placebo exposure variables for each outcome, based on a hypothetical reform implemented 13–20 years prior to the 1983 LNA reform. These results focus on the older cohort of never-treated individuals (born between 1940 and 1969). As shown, the coefficients for the placebo exposure variables are generally not significant and very close to zero. This falsification exercise, therefore, does not reveal any systematic patterns that might indicate the presence of spurious pre-existing trends across cohorts that could affect the exposure variable in the baseline estimates.

Table 4a: Falsification (1) - dependent variable: National Identity

		placebo reform X years before 1983							
	baseline	13	14	15	16	17	18	19	20
years of bilingual compulsory education	-0.005**								
	(0.002)								
fake exposure (placebo reform X years before 1983)		-0.003							
		(0.004)							
fake exposure (placebo reform X years before 1983)			-0.003						
			(0.004)						
fake exposure (placebo reform X years before 1983)				-0.003					
				(0.004)					
fake exposure (placebo reform X years before 1983)					-0.001				
					(0.005)				
fake exposure (placebo reform X years before 1983)						-0.000			
						(0.004)			
fake exposure (placebo reform X years before 1983)							-0.001		
							(0.004)		
fake exposure (placebo reform X years before 1983)								0.000	
								(0.003)	
fake exposure (placebo reform X years before 1983)									0.001
									(0.003)
R-squared	0.220	0.168	0.168	0.168	0.168	0.168	0.168	0.168	0.168
number of observations	17516	10322	10322	10322	10322	10322	10322	10322	10322

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins. Placebo reform for never-treated cohorts (1940-1969).

Table 4b: Falsification (1) - dependent variable: Linguistic Identity

		placebo reform X years before 1983							
	baseline	13	14	15	16	17	18	19	20
years of bilingual compulsory education	-0.001								
	(0.002)								
fake exposure (placebo reform X years before 1983)		-0.005*							
		(0.003)							
fake exposure (placebo reform X years before 1983)			-0.003						
			(0.002)						
fake exposure (placebo reform X years before 1983)				-0.002					
				(0.002)					
fake exposure (placebo reform X years before 1983)					-0.000				
					(0.002)				
fake exposure (placebo reform X years before 1983)						-0.000			
						(0.002)			
fake exposure (placebo reform X years before 1983)							0.000		
							(0.003)		
fake exposure (placebo reform X years before 1983)								0.001	
								(0.003)	
fake exposure (placebo reform X years before 1983)									0.001
									(0.003)
R-squared	0.610	0.581	0.581	0.581	0.581	0.581	0.581	0.581	0.581
numbef of observations	18081	10562	10562	10562	10562	10562	10562	10562	10562

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins. Placebo reform for never-treated cohorts (1940-1969).

Table 4c: Falsification (1) - dependent variable: Independence

		placebo reform X years before 1983							
	baseline	13	14	15	16	17	18	19	20
years of bilingual compulsory education	-0.003								
	(0.002)								
fake exposure (placebo reform X years before 1983)		0.004							
		(0.003)							
fake exposure (placebo reform X years before 1983)			0.005						
			(0.003)						
fake exposure (placebo reform X years before 1983)				0.005*					
				(0.003)					
fake exposure (placebo reform X years before 1983)					0.004				
					(0.003)				
fake exposure (placebo reform X years before 1983)						0.003			
						(0.003)			
fake exposure (placebo reform X years before 1983)							0.002		
							(0.003)		
fake exposure (placebo reform X years before 1983)								0.001	
								(0.003)	
fake exposure (placebo reform X years before 1983)									-0.000
									(0.003)
R-squared	0.223	0.184	0.184	0.184	0.184	0.184	0.184	0.184	0.184
numbef of observations	17070	10047	10047	10047	10047	10047	10047	10047	10047

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins. Placebo reform for never-treated cohorts (1940-1969).

Additionally, Table 5 presents the results of an alternative falsification exercise. This alternative approach focuses on individuals born within the same cohorts as our estimation sample (1950–1990), but who never received Catalan language instruction in school due to migrating to Catalonia after completing compulsory education. In this scenario, the coefficients of placebo exposure are virtually zero and statistically not significant. These findings, like those of the first falsification exercise, do not indicate any systematic pattern consistent with the presence of contemporaneous spurious trends that could influence the primary results.

Table 5: Falsification (2)

Dependent variable:	National Identity		Linguistic Identity		Independence	
	baseline	placebo	baseline	placebo	baseline	placebo
years of bilingual compulsory education	-0.005**	-0.002	-0.001	0.002	-0.003	-0.005
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.005)
R-squared	0.220	0.066	0.610	0.607	0.223	0.060
number of observations	17516	2486	18081	2790	17070	2599

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins. Placebo reform for never-treated individuals born in the rest of Spain, migrated after the end of compulsory education (same cohorts than the baseline).

To provide suggestive evidence regarding the influence of Catalan media and the sociolinguistic composition of the place of residence, we present regression estimates of the exposure variable. These regressions use the following outcomes: (a) current use of Catalan TV, radio, and newspapers (Table 6), and (b) various proxies for the presence of

Catalan speakers in the place of residence (Table 7). Regarding the role of Catalan media, the coefficient for years of exposure to bilingual education during compulsory schooling exhibits a marginally significant relationship with the propensity to watch Catalan TV for information on current political events. However, the negative sign of this coefficient is not consistent with a confounding role. This is because younger cohorts with greater exposure to Catalan TV, would be expected to exhibit a more intense Catalan identity. For other media outlets, the coefficient of exposure is virtually zero and not statistically significant.

Table 6: Catalan Media as Potential Confounders

Dependent variable:	TV	Radio	Newspapers
years of bilingual compulsory education	-0.005*	0.001	-0.003
	(0.003)	(0.002)	(0.002)
R-squared	0.147	0.073	0.039
number of observations	18113	18113	18113

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

Similar evidence emerges for the three indicators of local sociolinguistic composition: residing in Barcelona or its metropolitan area (county of Barcelona) and the proportion of residents in the county who can speak and write in Catalan. As shown in Table 7, the years of exposure to bilingual education during compulsory education do not significantly affect any of these outcomes, and the coefficients are, in any case, close to null.

Table 7: Residential Location as Potential Confounders

Dependent variable:	County of Barcelona	Barcelona	% speak and write in Catalan in the county
years of bilingual compulsory education	-0.003	-0.003*	0.000
	(0.002)	(0.001)	(0.000)
R-squared	0.036	0.020	0.097
number of observations	18113	18113	18113

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

To further validate the robustness of our results, we re-estimated equation (2) using alternative datasets. The SPE dataset allows for replication of results across all outcomes examined in this paper, using the same control variables as in our baseline analysis (see Table B6 of Appendix B). For *National Identity*, the results using the SPE dataset closely resemble those obtained with the BPO data. Specifically, the estimated exposure

coefficient is -0.007 (s.e. 0.004), in line with the one obtained using BPO data (-0.005, s.e. 0.002), although only marginally significant in this case. These findings from the SPE dataset consistently demonstrate that exposure to Catalan during compulsory schooling is not significantly associated with support for independence. However, we observe a negative and significant exposure coefficient for Linguistic Identity in the SPE analysis, with a positive albeit small estimate (0.008, s.e. 0.003).

Results from the Òmnibus dataset (Table B6 of Appendix B), which exclude parental origin controls, are very similar to those derived from the main dataset. The exposure coefficient for the *National Identity* is equal to -0.007 (marginally significant), while the coefficient for Linguistic Identity is virtually zero.

Finally, using the SLUP database, which only permits analysis of Linguistic Identity as the outcome, yields and exposure coefficient to bilingual education that is very close to zero and not significant (Table B6 of Appendix B).

In summary, the core findings presented in this paper remain consistent across these alternative data sources.

The evidence reported so far indicates that the introduction of bilingual education in Catalan schools barely affected identity formation and political preferences. The results suggest that any aggregate effect, if present, is likely negative and quantitatively small, primarily manifested as a slight decrease in strong feelings of Catalan identity (i.e., identifying as "only Catalan").

5.3. Comparison with CF-M: results

These findings stand in stark contrast to those of CF-M, who asserted that "respondents who have been exposed for a longer time period to teaching in Catalan have stronger Catalan feelings". To investigate this discrepancy, we replicated their analysis using the same data as in their paper (CIS data), but employing alternative outcome definitions and model specifications. We also conducted the counterpart supplementary analysis: using our data (BPO data), while adopting a methodology similar to CF-M in terms of outcome definition, sample selection, and control variables.

Table 8, presents the results from the CIS data analysis. The first two panels present the exposure coefficients for each of the identity indicators (*National Identity*, *Weak National Identity*, *Identity* and *IdentityD*, all specified in the introduction), using the same estimation sample and control variables as in CF-M. The first panel contains the estimates of the exposure coefficient obtained controlling only for the 4^{th} order age polynomial, while the second panel includes the full set of control variables used by CF-M. The estimates of the first column are obtained, respectively, using the entire 1-5 ordinal scale as outcome (Identity) and the dummy that is equal to 1 if the individual feels equally

Catalan than Spanish, more Catalan than Spanish or only Catalan (*IdentityD*). These correspond to the results presented by CF-M in their paper. However, the exposure coefficient diminishes considerably in magnitude and loses statistical significance when adopting a more transparent definition of national identity. These results demonstrate that the findings of CF-M are sensitive to the outcome definition.

Table 8: Replication of CF-M with CIS Data

Dependent variable:	(1) Identity	(2) IdentityD	(3) Weak National Identity	(4) National Identity
CF-M sample, only 4th order age polynomial				
years of bilingual compulsory education	0.075***	0.028***	0.017	0.014
	(0.023)	(0.007)	(0.011)	(0.012)
R-squared	0.011	0.017	0.012	0.002
number of observations	2309	2309	2309	2309
CF-M sample, CF-M controls				
years of bilingual compulsory education	0.058**	0.024***	0.010	0.012
	(0.025)	(0.007)	(0.012)	(0.011)
R-squared	0.318	0.222	0.271	0.137
number of observations	2309	2309	2309	2309
CF-M sample, CF-M controls with 2nd order age polynomial				
years of bilingual compulsory education	-0.003	-0.002	-0.003	0.001
	(0.019)	(0.006)	(0.008)	(0.006)
R-squared	0.314	0.214	0.270	0.135
number of observations	2309	2309	2309	2309
CF-M sample, C-D-F controls				
years of bilingual compulsory education	-0.008	-0.004	-0.005	0.001
	(0.016)	(0.006)	(0.007)	(0.006)
R-squared	0.325	0.144	0.308	0.134
number of observations	2278	2278	2278	2278
C-D-F sample, CF-M controls				
years of bilingual compulsory education	0.037	0.010	0.018	0.009
	(0.041)	(0.009)	(0.026)	(0.019)
R-squared	0.287	0.136	0.286	0.185
number of observations	1400	1400	1400	1400
C-D-F sample, C-D-F controls				
years of bilingual compulsory education	0.069***	0.030***	0.010	0.014
	(0.025)	(0.008)	(0.013)	(0.015)
R-squared	0.313	0.094	0.319	0.170
number of observations	1386	1386	1386	1386

OLS estimations with territorial weights for provinces (Barcelona = 1.7844, Girona = 0.452, Lleida = 0.3072, Tarragona = 0.4838). *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. C-D-F controls variables = wave dummies, age and age squared, gender, mother tongue, parents' origins. C-F-M control variable = fourth-order age polynomial, gender, dummies for i) born outside Catalonia, ii) born in Catalonia with one parent born outside Catalonia and iii) born in Catalonia with one parents born outside Catalonia, municipality size and province of residence.

Furthermore, as reported in the third panel, the exposure coefficient becomes virtually zero when using a second-order age polynomial, even when using the outcome definitions adopted by CF-M. It is worth noting that CF-M only reported results using third- and fifth-order age polynomials as robustness checks. That is, we obtain results similar to our baseline specification (very small, negative, and not statistically significant effects), even with the same data and model employed by CF-M, with the exception of a second-order age polynomial, instead of a third, fourth or fifth, as in their paper.

Similar evidence is obtained from a model that includes the same control variables as our preferred specification in the baseline analysis. Overall, this replication exercise indicates that the conclusions drawn by CF-M regarding the effects of the LNA reform may be unreliable. This is mostly because of their ambiguous definition of national identity, but also due to the specification of the age polynomial in combination with the use of a single cross-section of data.

In contrast to the instability observed in the CIS data analysis, the results obtained from the multiple cross sections of the BPO database exhibit considerable stability, as evident in Table 9. Specifically, we consistently observe very similar coefficients for the exposure variable in regressions where *National Identity* is the outcome, regardless of the chosen sample or control variables. Conversely, other outcome definitions yield virtually null coefficients.

Table 9: Replication of CF-M with BOP data

Dependent variable:	(1) Identity	(2) IdentityD	(3) Weak National Identity	(4) National Identity
CF-M sample, only 4th order age polynomial				
years of bilingual compulsory education	-0.006	0.002	-0.001	-0.008***
	(0.006)	(0.002)	(0.003)	(0.003)
R-squared	0.009	0.003	0.009	0.006
number of observations	23451	23451	23451	23451
CF-M sample, CF-M controls				
years of bilingual compulsory education	-0.003	0.003	0.000	-0.007***
	(0.005)	(0.002)	(0.002)	(0.003)
R-squared	0.168	0.060	0.149	0.093
number of observations	23451	23451	23451	23451
CF-M sample, CF-M controls with 2nd order age polynomial				
years of bilingual compulsory education	-0.009*	0.001	-0.002	-0.008***
	(0.005)	(0.002)	(0.002)	(0.002)
R-squared	0.168	0.060	0.148	0.093
number of observations	23451	23451	23451	23451
CF-M sample, C-D-F controls				
years of bilingual compulsory education	-0.010**	0.001	-0.003	-0.008***
	(0.004)	(0.001)	(0.002)	(0.002)
R-squared	0.318	0.064	0.297	0.222
number of observations	23451	23451	23451	23451
C-D-F sample, CF-M controls				
years of bilingual compulsory education	0.001	0.003	0.002	-0.005*
	(0.006)	(0.002)	(0.003)	(0.003)
R-squared	0.099	0.028	0.088	0.063
number of observations	17516	17516	17516	17516
C-D-F sample, C-D-F controls				
years of bilingual compulsory education	-0.003	0.002	0.000	-0.005**
	(0.004)	(0.001)	(0.002)	(0.002)
R-squared	0.303	0.049	0.280	0.220
number of observations	17516	17516	17516	17516

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. C-D-F controls variables = wave dummies, age and age squared, gender, mother tongue, parents' origins. C-F-M control variable = fourth-order age polynomial, gender, dummies for i) born outside Catalonia, ii) born in Catalonia with one parent born outside Catalonia and iii) born in Catalonia with one parents born outside Catalonia, municipality size and province of residence.

5.4 Analysis of Heterogeneous Effects

To conclude our analysis, we examine heterogeneous effects of exposure to bilingual education during compulsory education, by interacting years of exposure with exogenous covariates. For ease of interpretation, we report the marginal effects alongside

the p-value of the F-test for the equality of exposure coefficients across groups. We begin by analyzing potential gender-based heterogeneity, with the results presented in Table 10. Our findings do not reveal any significant differences in the effects of exposure to bilingual education between males and females. However, we observe a small negative effect of exposure on support for the independence of Catalonia among women, whereas the corresponding coefficient for males is quantitatively very small and not statistically different from zero.

Table 10: Heterogeneous Effects by Gender

Dependent variable:	(1) National Identity	(2) Linguistic Identity	(3) Independence
male	0.027**	-0.003	-0.016*
	(0.010)	(0.007)	(0.009)
years of bilingual compulsory education X female	-0.005**	-0.001	-0.004**
	(0.003)	(0.002)	(0.002)
years of bilingual compulsory education X male	-0.006**	-0.001	-0.002
	(0.003)	(0.002)	(0.002)
F-test for coefficients equality (years of bilingual compulsory education), p-value	0.901	0.806	0.064
R-squared	0.220	0.610	0.223
number of observations	17516	18081	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

Table 11 presents the heterogeneous effects of exposure to bilingual education during compulsory schooling across different linguistic communities defined according to their native language. Notably, the negative effect of exposure to the LNA reform on *National Identity*, observed previously, appears to be primarily driven by individuals with a Spanish or mixed language background (approximately -0.01 percentage points for each additional year of exposure). In contrast, this effect is negligible for native Catalan speakers.

Regarding Linguistic Identity, exposure to bilingual education exhibits an intriguing pattern: a positive effect for native Catalan Speakers (0.006, s.e. 0.001) and negative of equal magnitude for native Spanish speakers (-0.006, s.e. 0.002). That is, native Catalan speakers who received compulsory education under the bilingual system tend to be more attached to their mother tongue, while the opposite pattern is detected for native Spanish speakers.

Furthermore, years of compulsory education under the bilingual system have a small negative effect on support for Catalan independence among native Spanish speakers (-0.008, s.e. 0.002), while a smaller, marginally significant, positive effect for native Catalan speakers (0.004, s.e. 0.002).

It seems worth mentioning that, in contrast to the findings related to exposure to the LNA reform, the base effect of native language is substantial. Individuals with a Spanish or mixed language background exhibit significantly lower levels of: (a) *National identity* (30

and 17 points less likely to feel "only Catalan" compared to native Catalan speakers), (b) *Linguistic Identity* (62 and 45 percentage points less likely to adopt only Catalan as their self-identification language for those with Spanish-only and both languages as their native language, respectively). For these two linguistic groups, the reduction in the probability of supporting Catalan independence, relative to native Catalan speakers, is virtually identical to the reduction observed for *National Identity*.

Table 11: Heterogeneous Effects by Native Language

Dependent variable:	(1) National Identity	(2) Linguistic Identity	(3) Independence
mother tongue = Catalan	reference category		
mother tongue = Spanish	-0.302***	-0.620***	-0.301***
	(0.014)	(0.016)	(0.016)
mother tongue = Spanish & Catalan	-0.174***	-0.453***	-0.196***
	(0.019)	(0.023)	(0.023)
years of bilingual compulsory education X mother tongue = Catalan	-0.000	0.006***	0.004*
	(0.003)	(0.001)	(0.002)
years of bilingual compulsory education X mother tongue = Spanish	-0.009***	-0.006***	-0.008***
	(0.003)	(0.002)	(0.002)
years of bilingual compulsory education X mother tongue = Spanish & Catalan	-0.010***	-0.001	-0.001
	(0.003)	(0.003)	(0.004)
F-test for coefficients equality (years of bilingual compulsory education), p-value	0.000	0.000	0.000
R-squared	0.222	0.613	0.225
number of observations	17516	18081	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

Our findings on the impact of parental origins reveal heterogeneous effects of the language-in-education reform, as reported in Table 12. For individuals with both parents born in Catalonia, the reform had no significant impact on national and linguistic identity, nor on support for independence. However, the reform negatively affected these outcomes for individuals with at least one parent born outside Catalonia. Furthermore, a base effect of parental origins is evident, with individuals having both parents born outside Catalonia exhibiting significantly lower levels across all three outcomes (approximately 15 percentage points lower) compared to those with both parents born in Catalonia.

Table 12: Heterogeneous Effects by Parental Origins

Dependent variable:	(1) National Identity	(2) Linguistic Identity	(3) Independence
both parents born in Catalonia	reference category		
one parent born outside Catalonia	-0.061***	-0.005	-0.038**
	(0.014)	(0.011)	(0.017)
both parents born outside Catalonia	-0.151***	-0.148***	-0.155***
	(0.014)	(0.016)	(0.020)
years of bilingual compulsory education X both parents born in Catalonia	-0.002	0.003	0.002
	(0.003)	(0.002)	(0.002)
years of bilingual compulsory education X one parent born outside Catalonia	-0.011***	-0.006***	-0.007***
	(0.003)	(0.002)	(0.003)
years of bilingual compulsory education X both parents born outside Catalonia	-0.005**	-0.001	-0.004*
	(0.002)	(0.002)	(0.002)
F-test for coefficients equality (years of bilingual compulsory education), p-value	0.000	0.000	0.000
R-squared	0.221	0.611	0.224
number of observations	17516	18081	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

In summary the analysis of heterogeneous effects reveals a moderate, yet statistically significant, backlash effect of the introduction of Catalan as medium of instruction. This finding aligns with research by Fouka (2020), which demonstrated that the prohibition of German in US Schools during the Americanization Movement had unintended consequences of strengthening German-American cultural identity.

6. Conclusions

In this paper, we study the impact of language-in-education policies on identity. In particular, we provide new evidence on the effects of a reform introduced in the 1980's that made Catalan an additional language of instruction in Catalan schools. While this topic was previously analysed by Clots-Figueras and Masella (2013), our findings differ from theirs. We observe that the aggregate effect of the reform on various indicators of Catalan identity is small, and sometimes negative. In fact, exposure to the reform appears to have discouraged the adoption of a Catalan identity, particularly among individuals from Spanish-speaking families and those with parents born outside Catalonia.

The absence of a positive effect on the adoption of a Catalan identity can be viewed as a natural and expected outcome, given the political and institutional context. The conditions surrounding this reform differ significantly from those during the formation of most Western nation-states. Specifically, the reform took place within the Spanish institutional framework. Furthermore, the Spanish-speaking community in Catalonia has never been a minority. While the reform explicitly aimed to expand Catalan language skills, especially among Spanish-speaking students, the implicit goal of fostering a Catalan identity likely encountered significant challenges. In fact, our evidence suggests

that the reform might have triggered a backlash within the Spanish-speaking community.

It is important to emphasize that, beyond to role of the education system, the primary predictor of an individual's identity is family background. Both the native language and geographic origins of parents have a profound impact on identity indicators. Given the relatively minor role of the education system, the evolution of the identity distribution in Catalan society can likely be predicted by examining the demographic trends of various social groups.

References

Akerlof, G. and R. Kranton (2010), *Identity Economics*. Princeton University Press.

Akerlof, G and R. Kranton (2000), Economics and Identity. *The Quarterly Journal of Economics* 115 (3), 715-753.

Alesina, A. P. Giuliano, and B. Reich (2021), Nation-building and Education. *The Economic Journal* 131 (August), 2273-2303.

Alesina, A., B. Reich, and A. Riboni (2020), Nation-building, Nationalism, and Wars. *Journal of Economic Growth* 25 (4), 381-430.

Alesina, A., E. Spolaore, and R. Wacziarg (2000), Economic Integration and Political Desintegration. *American Economic Review* 90 (5), 1276-96.

Aspachs-Bracons, O., I. Clots-Figueras, and P. Masella (2008), Compulsory Language, Educational Policies, and Identity Formation. *Journal of the European Economic Association*, Papers and Proceedings, vol. 6, 433-444.

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo (2018), Nation-Building through Compulsory Schooling during the Age of Mass Migration. *Economic Journal* 129 (167), 69-102.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong (2019), Unity in Diversity? How Intergroup Contact Can Foster Nation Building. *American Economic Review* 109 (11), 3987-4025.

Blanc, G. and M. Kubo (2021), Schools, Language and Nations: Evidence from a Natural Experiment in France, mimeo.

Caminal, R., & Di Paolo, A. (2019), Your language or mine? The noncommunicative benefits of language skills. *Economic Inquiry* 57(1), 726-750.

Caminal, R., L. Cappellari, A. Di Paolo (2021), Language-in-education, language skills and the intergenerational transmission of language in a bilingual society, *Labour Economics*, vol. 70, num. 101975.

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

Carvalho, J.P., M. Koyama, and C. Williams (2024), Resisting Education. *Journal of the European Economic Association* 22 (6), 2549-2597.

Casas, A., Curci, F., and De Moragas, A. I. (2024), Judicial Decisions, Backlash and Secessionism: The Spanish Constitutional Court and Catalonia. *The Economic Journal*, 134 (664): 3202–3231.

Cinnirella, F. and R. Schueler (2018), Nation building: The Role of central spending in education. *Explorations in Economic History* 67, 18-39.

Clots-Figueras, I., and P. Masella (2013), Education, Language and Identity. *The Economic Journal* 123 (August), F332-F353.

Charness, G. and Y. Chen (2020), Social Identity, Group Behavior and Teams. *Annual Review of Economics* 12: 691-713.

Dehdari. S. and K. Gehring (2022), The Origins of Common Identity: Evidence from Alsace-Lorraine. *American Economic Journal: Applied Economics* 14 (1), 261-292.

Fishman, J. (1991), Reversing Language Shift: Theoretical and empirical foundations of assistance to threatened langauges. *Multilingual Matters*, 76.

Fouka, V. (2020). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *Review of Economic Studies* 87, 204-239.

Ginsburgh, V. and S. Weber (2020), The Economics of Language. *Journal of Economic Literature* 58 (2), 348-404.

Hauk, E. and J. Ortega (2019), Schooling, nation building and industrialization. *Journal of Theoretical Politics* 33(1), 56-94.

Garcia-López, M. À., Nicolini, R., & Roig Sabaté, J. L. (2021), Urban spatial structure in Barcelona (1902–2011): Immigration, spatial segregation and new centrality governance. *Applied Spatial Analysis and Policy*, *14*(3), 591-629.

Laitin, D. and R. Ramachandrian (2023), The Historical Sources of Language Policy, mimeo.

Newman, M. and M. Trenchs-Parera (2015), Language Policies, Ideologies and Attitudes in Catalonia. Part 1: Reversing Language Shift n the Twentieth Century. *Language and Linguistics Compass* 9/7, 285-294.

Ricento, T. (2003), The Discursive Construction of Americanism. *Discourse and Society* 14 (5), 611-637.

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

Soysal, Y. N., and D. Strang (1989), Construction of the first mass education systems in nineteenth-century Europe. *Sociology of education* 277-288.

Shayo, M. (2020), Social Identity and Economic Policy. *Annual Review of Economics* 12: 355-89.

Woolard, K. and T. Gahng (1990), Language Policies and Attitudes in Autonomous Catalonia. *Language in Society*, Vol. 19, No. 3, pp. 311-330

You, Y. (2018), Unified Language, Labor and Ideology, mimeo.

Appendix A: Additional Data Sources

Survey of the Political Environment, SPE (*Enquesta sobre Context Polític*)

This is a yearly survey covering the period 2014-2021, which was conducted by the Public Opinion Center (Centre d'Estudis d'opinió (CEO) of the regional government. It is a random sample of the population with Spanish nationality residing in Catalonia, aged 18 or more. The survey was collected using a multistage stratified procedure. Primary sample units are municipalities, randomly selected within province and city size, and the secondary sample units consists of census tracts of selected municipalities. Individuals within the secondary sample units are selected by applying crossed quotas defined according to gender, age, and place of birth, based on official figures from the Population Census (*Padró Municipal d'Habitants*) of the year before each survey. Interviews were carried out through the Computer Assisted Telephone Interview (CATI) methodology. The SPE survey contains most of the variables that we retained from the BPO database (our main data source) and all the outcome variables, which are defined in the same way. The sample we used for the empirical analysis is obtained by applying the same selection criteria than for the main dataset: individuals born in Catalonia and individuals born in other Spanish regions who migrated before age 6, born between 1950 and 1990. Table A1 contains descriptive statistics of selected variables.

Table A1: Descriptive Statistics of the SPE survey

Variable	Mean	Std. dev.
Outcome Variables		
Feeling of belonging*		
only Spanish	0.021	0.145
more Spanish than Catalan	0.023	0.151
equal Spanish than Catalan	0.366	0.482
more Catalan than Spanish	0.274	0.446
only Catalan	0.314	0.464
Self-Identification language**		
Catalan	0.515	0.500
Spanish	0.262	0.440
Catalan & Spanish	0.223	0.416
Political Organization of the State***		
Catalonia should be a Spanish Region	0.040	0.196
Catalonia should be an Autonomous Community	0.227	0.419
Catalonia should be a Federal State	0.285	0.451
Catalonia should be an Independent State	0.448	0.497
Explanatory Variables		
years of bilingual compulsory education	3.571	4.004
male	0.498	0.500
age	50.68	11.74
mother tongue = Catalan	0.449	0.497
mother tongue = Spanish	0.348	0.476
mother tongue = Catalan & Spanish	0.203	0.402
both parents born in Catalonia	0.414	0.493
one parent born in Catalonia	0.224	0.417
both parents born outside Catalonia	0.362	0.481

Number of observations: 4427; *4285 valid observations; ** 4417 valid observations; *** 4231 valid observations.

Òmnibus of the Generalitat de Catalunya (OGC)

The survey is a periodic survey that started in 2007 and is administrated by the the Public Opinion Center (*Centre d'Estudis d'Opinió*, CEO) of the regional government. It is a random survey of individuals aged 18 or more residing in Catalonia, obtained through a multistage stratified procedure. The primary strata are municipalities, randomly selected within provinces according to their size, and the secondary strata are individuals, who are selected by applying crossed quotas defined according to gender, age group and place of birth, based on based on official figures from the Population Census (*Padró Municipal*

d'Habitants) of the year before each survey. Interviews were carried out through the Computer Assisted Telephone Interview (CATI) methodology. The OGC survey contains fewer relevant variables for our purposes than other surveys. Specifically, regarding outcomes variable, it only includes information about the "feeling of belonging" variable (1-5 scale, defined in the same way than in other data sources) and self-identification language. As for control variables, the OGC survey contains information about basic sociodemographic characteristics and mother tongue, but there is no information about parents' place of birth. Moreover, the question about the year of arrival to Catalonia has been removed since the first wave of 2022. For this reason, we retained data covering the period 2014-2021[26], applying the same filters that we used for the baseline survey (individuals born in Catalonia and those born in other Spanish regions who migrated before age 6, between 1950 and 1990). Descriptive statistics of the relevant variables are reported in Table A2.

Table A2: Descriptive Statistics of the OGC survey

Variable	Mean	Std. dev.
Outcome Variables		
Feeling of belonging*		
only Spanish	0.023	0.151
more Spanish than Catalan	0.025	0.157
equal Spanish than Catalan	0.364	0.481
more Catalan than Spanish	0.284	0.451
only Catalan	0.303	0.460
Self-Identification language**		
Catalan	0.553	0.497
Spanish	0.290	0.454
Catalan & Spanish	0.157	0.364
Explanatory Variables		
years of bilingual compulsory education	3.670	4.103
male	0.500	0.500
age	45.84	11.29
mother tongue = Catalan	0.475	0.499
mother tongue = Spanish	0.455	0.498
mother tongue = Catalan & Spanish	0.070	0.256

Number of observations: 7982; *7660 valid observations; ** 7954 valid observations.

[26] The OGC survey contains three waves for 2014, two waves for 2015, 2016 and 2017 and a single wave for 2018-2021.

Survey of the Language Use of the Population (SLUP)

This survey (*Enquesta d'Usos Lingüístics de la Població de Catalunya*) is one of the official statistics carried out by the Catalan Statistical Institute (IDESCAT) every five years since 2008. For the first wave of 2008, the sample was obtained through a multistage stratified procedure in which municipalities (and districts within the city of Barcelona) represented the first strata, randomly selected according to the number of inhabitants, and the second strata were individuals aged 15 or more, randomly selected after applying quotas defined according to gender, age and place of birth (based on the Population Census of the survey's year). For the waves of 2013 and 2018, the selection of individuals was carried out in the same way but, for the first stage, the sampling unit were randomly selected municipalities when the number of inhabitants was below 50,000 and census tracts for larges municipalities. The survey was carried out with several methods (CAWI, CATI and CAPI) depending on each specific case. The survey contains a large set of variables, mostly related to language. Here we retain sociodemographic information (gender, year of birth, place of birth of individuals and their parents and year of arrival to Catalonia for those born abroad), mother tongue and self-identification language. This last variable represents the only identity proxy that is available in the SLUP dataset. We apply the same sample restrictions than for other data sources (individuals born between 1950 and 1990, in Catalonia or in other Spanish regions, who migrated before age 6). Table A3 reports descriptive statistics of relevant variable for the estimation sample.

Table A3: Descriptive Statistics of the SLUP survey

Variable	Mean	Std. dev.
Outcome Variables		
Self-Identification language		
Catalan	0.579	0.494
Spanish	0.328	0.469
Catalan & Spanish	0.094	0.291
Explanatory Variables		
years of bilingual compulsory education	3.637	4.041
male	0.492	0.500
age	43.096	11.519
mother tongue = Catalan	0.506	0.500
mother tongue = Spanish	0.452	0.498
mother tongue = Catalan & Spanish	0.042	0.202
both parents born in Catalonia	0.448	0.497
one parent born in Catalonia	0.223	0.416
both parents born outside Catalonia	0.328	0.470

Number of observations: 10189.

The Survey on the Social and Political Situation of Catalonia (CIS)

This survey (*Encuesta sobre la Situación Social y Política de Cataluña,* SSPC) was carried out by the *Centro de Investigaciones Sociológicas* (CIS) in 2001. This is the same cross-sectional dataset used in CF-M. It contains a random sample of the population of the population aged 18 or more residing in Catalonia, obtained through a multistage stratified sampling procedure. Primary sample units are municipalities, selected according to their size, secondary sample units are census tracts (randomly selected within each municipality). Finally, individuals are randomly selected within each selected census tract, by applying crossed quotas according to gender and age. The survey was conducted with personal interviews within the households. With the aim of replicating the main findings reported in the paper by CF-M, we retain information about gender, age, place of birth, year of arrival to Catalonia (in categories), parents' place of birth, province of residence, municipality size and the variable capturing the feeling of belonging (1-5 scale, defined as in other datasets). Additionally, we incorporate information about individuals' mother tongue, which was not considered by CF-M. We apply the same sample selection as in CF-M, that is, we retain all birth cohorts and remove individuals not born in Catalonia who migrated after age 18, with valid information about the main variables, ending up with a final sample of 2309 observations. However, in the replication exercise we also consider other sample restrictions that mimic the criteria we use for our main estimation sample. Table A4 of the Appendix contains basic descriptive statistics of relevant variables, which virtually coincide with those reported by CF-M.

Table A4: Descriptive Statistics of the CIS Survey

	CFM Sample		CDF Sample	
Outcome Variables	Mean	Std. dev.	Mean	Std. dev.
Feeling of belonging				
only Spanish	0.080	0.272	0.059	0.235
more Spanish than Catalan	0.047	0.212	0.051	0.219
equal Spanish than Catalan	0.320	0.467	0.324	0.468
more Catalan than Spanish	0.324	0.468	0.321	0.467
only Catalan	0.228	0.420	0.246	0.431
Explanatory Variables				
years of bilingual compulsory education	1.754	3.129	2.807	3.556
female	0.483	0.500	0.496	0.500
age	44.99	18.09	33.29	9.458
mother tongue = Catalan	0.360	0.480	0.389	0.488
mother tongue = Spanish	0.582	0.493	0.540	0.499
mother tongue = Catalan & Spanish	0.059	0.235	0.071	0.256
both parents born in Catalonia	0.536	0.499	0.510	0.500
one parent born in Catalonia	0.138	0.345	0.173	0.378
both parents born outside Catalonia	0.327	0.469	0.317	0.466
born outside Catalonia	0.159	0.366	0.078	0.268
born in Catalonia, parents born outside Catalonia	0.179	0.384	0.243	0.429
born in Catalonia, one parent born in Catalonia	0.131	0.337	0.170	0.376
born in Catalonia, both parents born in Catalonia	0.531	0.499	0.509	0.500
province = Barcelona	0.415	0.493	0.426	0.495
province = Girona	0.196	0.397	0.185	0.388
province = Lleida	0.204	0.403	0.201	0.401
province = Tarragona	0.185	0.388	0.187	0.390
municipality size < 2000 inhabitants	0.146	0.353	0.136	0.343
municipality size = 2001-10000 inhabitants	0.210	0.407	0.200	0.400
municipality size = 10001-50000 inhabitants	0.225	0.418	0.231	0.421
municipality size =50001-100000 inhabitants	0.107	0.309	0.104	0.306
municipality size = 100001-400000 inhabitants	0.176	0.381	0.191	0.393
municipality size > 1000000 inhabitants	0.137	0.344	0.138	0.345

Number of observations: CFM Sample = 2309 (mother tongue, 2278 valid observations); CDF sample = 1400 (mother tongue, 1386 valid observations).

Appendix B: Additional Results

Table B1: Descriptive Statistics for Falsification Analysis (1)

Variable	Mean	Std. dev.
Outcome Variables		
Feeling of belonging*		
only Spanish	0.020	0.140
more Spanish than Catalan	0.022	0.148
equal Spanish than Catalan	0.285	0.451
more Catalan than Spanish	0.282	0.450
only Catalan	0.391	0.488
Self-Identification language**		
Catalan	0.666	0.472
Spanish	0.223	0.416
Catalan & Spanish	0.111	0.314
Political Organization of the State***		
Catalonia should be a Spanish Region	0.036	0.185
Catalonia should be an Autonomous Community	0.183	0.386
Catalonia should be a Federal State	0.237	0.426
Catalonia should be an Independent State	0.544	0.498
Explanatory Variables		
male	0.473	0.499
age	60.05	8.395
mother tongue = Catalan	0.576	0.494
mother tongue = Spanish	0.375	0.484
mother tongue = Catalan & Spanish	0.049	0.216
both parents born in Catalonia	0.521	0.500
one parent born in Catalonia	0.174	0.379
both parents born outside Catalonia	0.305	0.461

Number of observations: 10578; *10322 valid observations; ** 10562 valid observations; *** 10047 valid observations.

Table B2: Descriptive Statistics for Falsification Analysis (2)

Variable	Mean	Std. dev.
Outcome Variables		
Feeling of belonging*		
only Spanish	0.222	0.416
more Spanish than Catalan	0.130	0.337
equal Spanish than Catalan	0.500	0.500
more Catalan than Spanish	0.099	0.298
only Catalan	0.049	0.217
Self-Identification language**		
Catalan	0.060	0.238
Spanish	0.886	0.317
Catalan & Spanish	0.053	0.225
Political Organization of the State***		
Catalonia should be a Spanish Region	0.100	0.300
Catalonia should be an Autonomous Community	0.470	0.499
Catalonia should be a Federal State	0.290	0.454
Catalonia should be an Independent State	0.140	0.347
Explanatory Variables		
years of (placebo) bilingual compulsory education	2.891	3.893
male	0.443	0.497
age	50.68	11.74
mother tongue = Catalan	0.048	0.215
mother tongue = Spanish	0.939	0.239
mother tongue = Catalan & Spanish	0.012	0.111
both parents born in Catalonia	0.006	0.080
one parent born in Catalonia	0.033	0.178
both parents born outside Catalonia	0.961	0.194

Number of observations: 2809; *2486 valid observations; ** 2790 valid observations; *** 2599 valid observations.

Table B3: p-value Correction for Multiple Hypothesis Testing

	baseline p-value	resample p-value	Romano-Wolf p-value
National Identity	0.0287	0.0123	0.0257
Linguistic Identity	0.5038	0.4135	0.4135
Independence	0.1219	0.0603	0.1076

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, age and age squared, gender, mother tongue, parental origins.

Table B4: Weaker Identity Measures

Dependent variable:	(1) Weak National Identity	(2) Weak Linguistic Identity	(3) Sovereignty
Panel A			
years of bilingual compulsory education	0.004*	0.004	0.001
	(0.002)	(0.003)	(0.002)
adjusted R-squared	0.012	0.029	0.014
Panel B			
years of bilingual compulsory education	0.000	-0.002	-0.001
	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.279	0.467	0.145
Panel C			
years of bilingual compulsory education	0.000	-0.001	-0.000
	(0.002)	(0.001)	(0.002)
adjusted R-squared	0.301	0.478	0.186
mean of the dependent variable	0.627	0.710	0.755
number of observations	17516	18081	17070

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Panel A: controls variables = wave dummies, age and age squared, gender. Panel B: control variables = wave dummies, age and age squared, gender, mother tongue, parents' origins. Panel C: control variables = wave dummies, age and age squared, gender, mother tongue, parents' origins, living with the partner, number of children.

Table B5: Sensitivity to Time Controls

	(1)	(2)	(3)	(4)	(5)
Panel A: dependent variable = National Identity (number of observations = 17516)					
years of bilingual compulsory education	-0.005**	-0.005**	-0.005**	-0.005**	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.219	0.218	0.217	0.217	0.217
Panel B: dependent variable = Linguistic Identity (number of observations = 18081)					
years of bilingual compulsory education	-0.001	-0.001	-0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.610	0.609	0.609	0.609	0.609
Panel C: dependent variable = Independence (number of observations = 17070)					
years of bilingual compulsory education	-0.003	-0.003	-0.003*	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
adjusted R-squared	0.222	0.221	0.221	0.221	0.221
time control	wave dummies (baseline)	year dummies	grouped years dummies	linear trend	quadratic trend

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Control variables = wave dummies, gender, mother tongue, parents' origins. Grouped year dummies: 2014-2016, 2017-2018, 2019-2020.

Table B6: Baseline Results from other Data Sources

Dependent variable:	(1) National Identity	(2) Linguistic Identity	(3) Independence
ECP data			
years of bilingual compulsory education	-0.007*	0.008**	0.002
	(0.004)	(0.003)	(0.004)
R-squared	0.229	0.563	0.223
number of observations	4285	4417	4231
OGC data			
years of bilingual compulsory education	-0.007*	-0.001	
	(0.004)	(0.002)	
R-squared	0.200	0.563	
number of observations	7660	7954	
EULP data			
years of bilingual compulsory education		0.003	
		(0.002)	
R-squared		0.602	
number of observations		10189	

OLS estimations. *** significant at 1%; ** significant at 5%; * significant at 10%. Clustered standard errors by year of birth in parenthesis. Controls variables = wave dummies, age and age squared, gender, mother tongue, parents' origins (not available in OGC data).

UBIREA

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

WEBSITE: www.ub.edu/irea • **CONTACT:** irea@ub.edu

AQR

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona